Multiple applications, competing mechanisms, and market power

Albrecht, James; Cai, Xiaoming; Gautier, Pieter; Vroman, Susan

published in
Journal of Economic Theory
2020

DOI (link to publisher)
10.1016/j.jet.2020.105121

Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 190 (2020) 105121
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

Multiple applications, competing mechanisms, and market power ☆

James Albrecht [a,b], Xiaoming Cai [c,*], Pieter Gautier [d,b,e], Susan Vroman [a,b]

[a] *Georgetown University, United States of America*
[b] *IZA, Germany*
[c] *Peking University HSBC Business School, China*
[d] *Vrije Universiteit Amsterdam, Tinbergen Institute, Netherlands*
[e] *CEPR, United Kingdom of Great Britain and Northern Ireland*

Received 28 July 2019; final version received 23 September 2020; accepted 28 September 2020
Available online 5 October 2020

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 the firm's candidate. When the contract space includes application fees, there exists a continuum of symmetric equilibria of which only one is efficient, and it has a posted wage equal to match output. In the inefficient equilibria, the wage is below match output, and the value of a worker's application depends on whether he or she receives another offer. This allows individual firms to free ride on one another and gives firms market power. When we endogenize the number of applications and allow for general mechanisms, only the efficient equilibrium survives. By allowing for general mechanisms, we are able to examine the sources of inefficiency in the multiple applications literature.

JEL classification: C78; D44; D83

☆ We thank an Associate Editor and three referees as well as Philipp Kircher for valuable comments. Cai thanks the NSFC Grant 71803143 for support.
* Corresponding author.
E-mail addresses: albrecht@georgetown.edu (J. Albrecht), xmingcai@gmail.com (X. Cai), p.a.gautier@vu.nl (P. Gautier), susan.vroman@georgetown.edu (S. Vroman).

https://doi.org/10.1016/j.jet.2020.105121
0022-0531/

Keywords: Multiple applications; Directed search; Competing mechanisms; Efficiency; Market power

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, as in Moen (1997), is that the market equilibrium is efficient. Albrecht et al. (2006) and Galenianos and Kircher (2009) challenge this result and show 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, we explore the efficiency of competitive search equilibrium[1] in a labor market in which workers make multiple applications. For simplicity, we consider a static model. Firms can only consider a limited number of candidates - one in our model - even though a firm's chosen applicant may have a competing offer or offers. That is, if a firm's chosen candidate rejects its offer, the firm is unable to consider a second applicant, etc.[2] Our contribution is to allow firms to choose from a set of quite 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 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 promises to pay on the number of offers that its chosen candidate has, we show that competition implies that a worker with multiple offers receives a wage equal to 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. Allowing firms to condition wages on the number of applications received turns out to play no role in our analysis. This is in contrast to Coles and Eeckhout (2003) who show that in a small market in which each buyer approaches only one seller, the possibility of conditioning prices on the number of buyers creates a continuum of equilibria that are not payoff equivalent. Our model also generates a continuum of non-payoff-equivalent equilibria, and the driving force behind the existence of multiple equilbria in our paper is similar to the one in

[1] Different papers in the literature mean different things by "competitive search." What we mean is simply that (i) search is directed and (ii) the market is large enough that firms take market utility as given.

[2] Similarly, in Albrecht et al. (2006) and Galenianos and Kircher (2009), firms can consider only one candidate, although in an extension Albrecht et al. (2006) also consider "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.

Coles and Eeckhout (2003). Firms have multiple instruments to provide value to workers while workers only care about the joint value of the instruments, and because of this, each firm has a continuum of best responses. However, in Coles and Eeckhout (2003), payoff-relevant multiplicity is a small-market phenomenon. Each seller's posted mechanism affects buyers' market utility, but in a large market this effect disappears so that all equilibria become payoff equivalent. In contrast, in our model, payoff-relevant multiplicity is a large-market phenomenon. Equilibria that are characterized by different wage-fee pairs are not payoff equivalent in a large market because in all but one of the equilibria, firms have joint monopsony power (the source of which is discussed below).

In all equilibria, we find that firms charge positive application fees. Although fees deter 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 revenue 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 and fee ensures that workers receive their expected contribution to match surplus.

In actual labor markets, we rarely observe application fees. 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 (see Cheron and Decreuse (2017) and Albrecht et al. (2019)). 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 not efficient.

When workers apply to only one job, efficiency arises because (i) firms must offer workers the market utility, (ii) in a large market, firms take the market utility as given, and (iii) firms are residual claimants to the surplus so it is in their interest to offer efficient mechanisms.[3] When workers send out multiple applications, firms no longer can offer each applicant the market utility. Instead, firms offer workers the possibility of creating optimal application portfolios. The worker's payoff depends on his or her entire portfolio so a firm's expected payoff also depends on what happens to the applications that a worker sends to other firms. This is not something the firm can directly price. If workers can only apply to one job, the problem reduces to one of firms buying queues in a competitive market.

Bertrand competition for workers with multiple offers is not enough for efficiency. In our model, firms extract rents by lowering the wage when their chosen candidate has no other offers (in conjunction with application fees). That is, firms have market power even though search is directed and there is Bertrand competition, because the probability that the firm's candidate has another offer is less than one. Firms have joint monopsony power because worker value comes from the entire application portfolio. If other firms offer a bad deal (a low wage when their

[3] See, for example, Albrecht et al. (2014), Cai et al. (2018), Eeckhout and Kircher (2010a), and Peters (2013) for a literature review.

candidate has no other offers), this enhances an individual firm's market power. This externality also explains why we have a continuum of non-payoff-equivalent equilibria.

Efficiency under multiple applications requires that (i) the worker who is hired receives a wage equal to the match output regardless of whether this worker has other offers (otherwise, by the argument above, firms have market power) and implies that (ii) the worker receives his or her marginal contribution to the match. The only way that (i) and (ii) can be satisfied simultaneously is with a positive application fee. When workers send one application, condition (i) is not necessary so in that case, efficiency can be obtained without fees.

The above discussion suggests a simple scenario in which firms' market power disappears: the worker who is hired receives a wage equal to match output regardless of whether this worker has other offers. In this case, firm value is realized through application fees. Fees play a different role in our model than in the previous literature where fees arise because workers are risk-averse, there is incomplete information, or there is congestion in the meeting process. For example in Jacquet and Tan (2012) and Julien and Roger (2019), the application fee can be thought of as an insurance premium that the worker buys in order to get compensated when he or she is not hired. In Albrecht and Jovanovic (1986), fees arise due to incomplete information in pairwise matching markets. Fees or two-part tariffs have been studied extensively in the industrial organization literature; e.g., Armstrong and Vickers (2001), Rochet and Stole (2002), Yin (2004), and Thanassoulis (2007) show that firms offer a single two-part tariff with a marginal price equal to marginal cost in equilibrium. Armstrong and Vickers (2010) extends the above framework to allow consumers to buy different goods from different suppliers and find that firms offer a menu of efficient two-part tariffs with a discount for bundling that encourages buyers to do one-stop shopping. However, in this literature, firms don't face capacity constraints and a consumer's payoff at one firm does not depend on what he or she must pay at other firms. Finally, in Lester et al. (2015) and Cai et al. (2017, 2018), fees are critical for efficiency because they can price congestion externalities in the meeting process.

In our inefficient equilibria, workers with one offer receive a wage that is less than the full surplus. The value for the worker of applying to multiple jobs is the potential to generate multiple offers. This creates congestion without increasing output and is thus a pure rent-seeking activity. 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.

We also consider two extensions to our model. First, we introduce an application cost and endogenize the number of worker applications. For simplicity, we assume that the number of applications is a continuous variable. We find that a unique equilibrium exists and show that both firm entry and the number of applications per worker are efficient. In our second extension, we consider restrictions on the set of feasible wage mechanisms. We then relate our results to Albrecht et al. (2006) and Galenianos and Kircher (2009). Albrecht et al. (2006) allow firms to condition their wage on the number of 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 There is a continuum of identical unemployed workers (measure u) and a continuum of identical firms each with one vacancy (measure v). Both firms and workers are risk neutral. Initially, we take v as exogenous. We later endogenize v by allowing for free entry of vacancies. The productivity of a match is normalized to 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, where a is a positive integer. We assume for now that a is exogenous. 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).

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.[4] 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 worker it has selected). 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. If workers receive multiple offers they select the highest and in case of a tie, they simply randomize. 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 such as

4 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.

offer-beating strategies.[5] In Section 6, we look at the effect of various restrictions on the set of possible wage mechanisms, e.g., ruling out fees.

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.

Meeting technologies Consider a firm with expected queue length λ; i.e., λ is the expected number of workers applying to the firm. The meeting process is constant-returns-to-scale and characterized by a sequence of probability functions $\{p_n(\lambda) : n = 0, 1, 2, \dots\}$, where $p_n(\lambda)$ is the probability that a firm receives n applications. We assume that the expected number of applications equals the expected queue length. That is, $\sum_{n=1}^{\infty} n p_n(\lambda) = \lambda$ 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.[6] 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. Note that *if* all firms face the same expected queue length, then that common expected queue length is $\lambda = au/v$ since each worker sends out a applications. In general, the expected queue length depends on the wage mechanism and may vary from firm to firm.

If a firm receives n applications (an n-to-1 meeting), then the firm chooses a particular worker with probability $1/n$, so all workers who apply have the same probability to receive an offer from that firm. By an accounting identity, the probability that an application ends up in an n-to-1 meeting is $q_n(\lambda) = np_n(\lambda)/\lambda$, where $n \geq 1$, and the probability that an application leads to an offer from that firm is $\sum_{n=1}^{\infty} q_n(\lambda)/n$, which is simply $m(\lambda)/\lambda$. Finally, since we assume that meetings are nonrival, applications always arrive at firms, i.e. $\sum_{n=1}^{\infty} q_n(\lambda) = 1$. 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.[7] Keeping $m(\lambda)$ general allows us to see how properties of the equilibrium wage mechanism depend on the meeting technology.

For notational convenience, let $h(\lambda)$ denote the probability that an application fails to lead to an offer:

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

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

6 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 nonrival meeting technology), at one of the firms that received applicants in the first stage.

7 Note that for the urn-ball, $p_n(\lambda) = e^{-\lambda} \frac{\lambda^n}{n!}$ and for the geometric, $p_n(\lambda) = \frac{1}{1+\lambda} \left(\frac{\lambda}{1+\lambda}\right)^n$, where $n = 0, 1, 2, \dots$.

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

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

As $\lambda \to \infty$, $\frac{\lambda h'(\lambda)}{h(\lambda)} \to 0$.

2.2. Payoffs and equilibrium

We look for a symmetric pure-strategy equilibrium in which all firms choose the same wage mechanism (f, w_{ij}), where $i = 1, 2, \ldots, a$ and $j = 1, \ldots, \infty$. If all firms post the same wage mechanism (f, w_{ij}), workers will send their a applications randomly and independently. Recall that $p_j(\lambda)$ is the probability that a firm receives j applications, and $q_j(\lambda)$ is the probability that an application faces competition from $j-1$ other applications. The expected payoff or the *market utility* of workers is given by

$$U = -af + \sum_{i=1}^{a} \sum_{j_1, j_2, \ldots, j_i \geq 1} \binom{a}{i} h(\lambda)^{a-i} \frac{q_{j_1}(\lambda)}{j_1} \frac{q_{j_2}(\lambda)}{j_2} \cdots \frac{q_{j_i}(\lambda)}{j_i} \max(w_{ij_1}, w_{ij_2}, \ldots, w_{ij_i}), \tag{3}$$

where the first term gives the sum of application fees, the summand in the second term reflects the scenario in which the worker receives i offers out of a applications, and the first offer is from a firm who has received j_1 applications. Recall that w_{ij_k} is the promised wage at a firm that has received j_k applications when the selected worker has i offers in total. A worker with i offers, selects the highest one and in case of a tie, the worker randomizes among all firms that offer the highest wage. Recall that $h(\lambda)$ is the probability that an application fails to lead to an offer, and $q_j(\lambda)/j$ is the probability that an application results in an offer when the firm has $j-1$ other candidates.

The firm's expected payoff is given by,

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

The first term inside the large parentheses on the right-hand side is the fee income the firm receives when it has j applicants. The second term equals the additional profit in case its selected applicant has no other offers, which happens with probability $h(\lambda)^{a-1}$. Finally, $\mathcal{P}_{ij}$ in the last term denotes the probability that the firm has j applicants and the chosen applicant has $i > 1$ offers in total. $\mathcal{P}_{ij}$ is a complicated object which is given by

$$\mathcal{P}_{ij} = \sum_{j_1, \ldots, j_{i-1} \geq 1} \binom{a-1}{i-1} h(\lambda)^{a-i} \frac{q_{j_1}(\lambda)}{j_1} \cdots \frac{q_{j_{i-1}}(\lambda)}{j_{i-1}} \mathcal{C}(w_{ij}, w_{ij_1}, \ldots, w_{ij_{i-1}}) \tag{5}$$

where the summand reflects the scenario in which the selected worker has $i-1$ offers from his or her other $a-1$ applications, and the first offer of the $i-1$ offers is from a firm that has received j_1 applications so that the wage is w_{ij_1}, and so on. When the wage from the firm under consideration is not the highest of the i offers that the worker received, i.e., $w_{ij} < \max(w_{ij_1}, \ldots, w_{ij_{i-1}})$, then function $\mathcal{C}$, or the firm's winning probability, is zero. When w_{ij} is strictly larger than the other $i-$

1 offers, then $\mathcal{C}$ equals one. When there are n highest wage offers and w_{ij} is one of the n offers, $\mathcal{C}$ equals $1/n$ since workers randomize over the n highest wage offers with equal probability. Thus, the function $\mathcal{C}$ is a counting function, and it is formally defined by $\mathcal{C}(w_{i,j}, w_{ij_1}, \dots, w_{ij_{i-1}}) = \Pi_{k=1}^{i-1}\mathbb{I}(w_{ij} \geq w_{ij_k})(\sum_{k=1}^{i-1}\mathbb{I}(w_{ij} = w_{ij_k}))^{-1}$, where $\mathbb{I}$ is the standard indicator function.

Now consider the problem of a potential deviant firm. If the deviant firm posts a wage mechanism $(\widetilde{f}, \widetilde{w}_{ij})$, then it anticipates an expected number of applications, $\widetilde{\lambda}$. Workers apply to the deviant firm until they are indifferent between sending all a applications to non-deviants versus sending one application to the deviant firm and the other $a-1$ applications to non-deviant firms. This approach to a subgame perfect equilibrium is known as the *market-utility approach with single-firm deviation* and is also adopted by Albrecht et al. (2006) and Kircher (2009).[8]

An alternative approach in an equilibrium with a continuum of agents is to consider a joint deviation by an ε measure of firms and then let $\varepsilon \to 0$. In models in which workers can only send one application, the two approaches yield the same outcome. However, when workers can send multiple applications, the difference matters. When a single firm deviates, as in our model, workers can at most send one application to the deviant firm. When a measure of firms deviate, workers can potentially send all their applications to deviant firms. In Appendix A.10 we consider this alternative approach. We show that only the constrained efficient equilibrium survives when we allow for collective deviations. However, we believe that the market utility approach with single-firm deviation is more appropriate in our setting. This is the approach used in previous papers with multiple applications. The alternative approach is more appropriate for models in which market makers or platforms offer wage and market tightness pairs and workers and firms then choose which submarket to visit, as in Moen (1997).

Suppose that the deviant firm posts a wage mechanism $(\widetilde{f}, \widetilde{w}_{ij})$. Then its expected queue length $\widetilde{\lambda}$ is determined by the following indifference condition.

$$\begin{aligned} U = &-(a-1)f - \widetilde{f} \\ &+ h(\widetilde{\lambda})\sum_{i=1}^{a-1}\sum_{j_1,j_2,\dots,j_i\geq 1}\binom{a-1}{i}h(\lambda)^{a-1-i}\frac{q_{j_1}(\lambda)}{j_1}\cdots\frac{q_{j_i}(\lambda)}{j_i}\max(w_{ij_1},\dots,w_{ij_i}) \\ &+ \sum_{i=1}^{a}\sum_{j_0,j_1,\dots,j_i\geq 1}\binom{a-1}{i-1}h(\lambda)^{a-i}\frac{q_{j_0}(\widetilde{\lambda})}{j_0}\frac{q_{j_1}(\lambda)}{j_1}\cdots\frac{q_{j_{i-1}}(\lambda)}{j_{i-1}}\max(\widetilde{w}_{ij_0}, w_{ij_1},\dots,w_{ij_{i-1}}). \end{aligned} \tag{6}$$

The right-hand side is the expected payoff of a worker who sends one application to the deviant firm and the other $a-1$ applications to non-deviant firms. The first line gives the total amount of fees that the worker must pay (we do not restrict this term to be negative), the second line corresponds to the case in which the application to the deviant firm fails, which happens with probability $h(\widetilde{\lambda})$, but the worker receives i offers from the other $a-1$ applications, and the third line corresponds to the case where the worker receives i offers in total, and the first offer is from the deviant firm which received j_0 applications in total, and the second offer, if it exists, is from a non-deviant firm which receives j_1 applications, and so on.

[8] A microfoundation for this approach is to start with a finite economy and consider how workers respond to a deviation by a single firm and then take the limit by letting the number of workers and firms go to infinity while keeping their ratio constant. In models where workers can only send one application, Peters (1991, 1997, 2000) and Burdett et al. (2001) show that the equilibrium identified by the market utility approach is exactly the limit of the finite economy. However, as mentioned by Kircher (2009), such equivalence has not yet been proven for multiple-application models.

The expected payoff of the deviant firm is

$$\widetilde{\pi} = \sum_{j=1}^{\infty} p_j(\widetilde{\lambda}) \left(j\widetilde{f} + (1 - \widetilde{w}_{1j}) h(\lambda)^{a-1} + \sum_{i=2}^{a} (1 - \widetilde{w}_{ij}) \widetilde{\mathcal{P}}_{ij} \right). \tag{7}$$

The interpretation of (7) is the same as that of equation (4), where $\widetilde{\mathcal{P}}_{ij}$ is now given by

$$\widetilde{\mathcal{P}}_{ij} = \binom{a-1}{i-1} h(\lambda)^{a-i} \sum_{j_1,\ldots,j_{i-1}\geq 1} \frac{q_{j_1}(\lambda)}{j_1} \cdots \frac{q_{j_{i-1}}(\lambda)}{j_{i-1}} \mathcal{C}(\widetilde{w}_{ij}, w_{ij_1}, \ldots, w_{ij_{i-1}}), \tag{8}$$

which is the probability that the applicant selected by the deviant firm has $i-1$ other offers, and the first of these $i-1$ offers is from a firm that receives j_1 applications etc.

In a symmetric pure-strategy equilibrium, (i) given that firms choose optimal strategies, workers send out their applications optimally, and (ii) given workers' optimal application behavior, no individual firm has a profitable deviation. Formally,

Definition 1. In a symmetric pure-strategy equilibrium, firms choose (f, w_{ij}), where $i = 1, 2, \ldots, a$ and $j = 1, \ldots, \infty$, and workers receive their market utility U such that the following conditions are satisfied.

1. **Worker optimality.** Market utility U is given by equation (3), on and off the equilibrium path. For workers who apply to the deviant firm, the indifference condition, (6), holds.
2. **Firm optimality.** There exists no profitable deviation for firms. That is, the expected payoff from a deviation, which is given by equation (7), is no greater than the equilibrium payoff given by equation (4).

3. Characterization of equilibrium

3.1. The firm's problem

The firm's problem can be substantially simplified by recognizing that in equilibrium, the posted wage must be one when workers have multiple offers, i.e., $w_{ij} = 1$ if $i \geq 2$. As in Burdett and Judd (1983), if all firms post $w_{ij} < 1$ for some $i \geq 2$ and some j, then a deviant firm can post a slightly higher wage $\widetilde{w}_{ij} = w_{ij} + \varepsilon$, which leads to a discrete jump in the winning probability no matter how small ε is and hence a strictly higher expected profit.[9] This is formalized in Lemma 1.

Lemma 1. *In any equilibrium, if $p_j(\lambda) > 0$, then $w_{ij} = 1$ for all $i \geq 2$.*

Proof. See Appendix A.1. □

When $p_j(\lambda) = 0$, the choice of w_{ij} is irrelevant for both the worker and firm payoff. For completeness, we will set $w_{ij} = 1$ when $i \geq 2$ even when $p_j(\lambda)$ equals zero.

[9] This Bertrand outcome is the result of ex ante competition in mechanisms in which wages can be conditioned on the number of other offers that an applicant receives ex post. In particular, we are not motivating Bertrand competition by assuming that a firm can abandon its ex ante commitment to pay w_{ij} if faced with ex post competition for a worker's services.

The expression for the firm's expected payoff can now be substantially simplified. Consider equation (4). 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, Lemma 1 states that $w_{ij} = 1$ for $i > 1$; i.e., the third term in equation (4) equals zero. In short, the firm's expected payoff in equation (4) can be rewritten as

$$\pi = m(\lambda)(1 - w_1) h(\lambda)^{a-1} + \lambda f. \tag{9}$$

Now consider the worker side. Recalling that $q_j(\lambda) = j p_j(\lambda)/\lambda$, which implies that $\sum_{j=1}^{\infty} q_j(\lambda)/j = m(\lambda)/\lambda = 1 - h(\lambda)$, the market utility of the worker given in equation (3) can be rewritten as

$$U = \left(1 - h(\lambda)^a\right) - a(1 - h(\lambda)) h(\lambda)^{a-1} (1 - w_1) - af. \tag{10}$$

The first term is the probability that the worker receives at least one offer, the second term is the expected transfer from the worker to the firm if the worker has only one offer, and the third term is the sum of application fees. Note that the worker receives the total surplus minus the profit that the firm can appropriate, which is equal to the fees it receives plus a term that is zero except when the selected worker has exactly one offer.

Lemma 1 also greatly simplifies the problem of a deviant firm which will set $\widetilde{w}_{ij} = 1$ for $i \geq 2$. 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

$$\widetilde{\pi} = m(\widetilde{\lambda}) h(\lambda)^{a-1} (1 - \widetilde{w}_1) + \widetilde{\lambda} \widetilde{f}, \tag{11}$$

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.

The indifference condition for workers, equation (6), now becomes

$$\begin{aligned} U = {} & \left(1 - h(\widetilde{\lambda}) h(\lambda)^{a-1}\right) - (a-1) h(\widetilde{\lambda}) h(\lambda)^{a-2} (1 - h(\lambda))(1 - w_1) \\ & - (1 - h(\widetilde{\lambda})) h(\lambda)^{a-1} (1 - \widetilde{w}_1) - \widetilde{f} - (a-1) f, \end{aligned} \tag{12}$$

where U is given by equation (10). The right-hand side of equation (12) 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 one offer. The second term describes the case in which the worker receives exactly one offer from one of the non-deviant firms and receives no offer from the deviant firm, while the third term describes the case in which the worker has one offer from the deviant and the other $a - 1$ applications were rejected. Finally, the worker pays $\widetilde{f} + (a - 1) f$ in application fees.

The right-hand side of equation (12) 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 (12).

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 (11) where $\widetilde{\lambda}$ is implicitly determined by equation (12).

3.2. Equilibrium

We can now solve the deviant firm's problem. We start by defining the expected profit *per worker* in the deviant's queue.

$$\widetilde{\tau} = (1 - h(\widetilde{\lambda}))h(\lambda)^{a-1}(1 - \widetilde{w}_1) + \widetilde{f}. \tag{13}$$

The expected profit of the deviant firm, equation (11), can then be rewritten as

$$\widetilde{\pi} = \widetilde{\lambda}\widetilde{\tau}. \tag{14}$$

When each worker can only send one application, the expected profit per worker depends only on the mechanism posted by the deviant firm. However, with multiple applications, the expected profit per worker takes a more complicated form.

Keeping $\widetilde{\lambda}$ fixed, equations (13) and (14) imply that the deviant firm's expected profit depends on $\widetilde{w}_1$ and $\widetilde{f}$ through $\widetilde{\tau}$. A similar logic applies to workers who apply to the deviant firm. From equation (12), we can see that keeping $\widetilde{\lambda}$ fixed, the worker's payoff depends on $\widetilde{w}_1$ and $\widetilde{f}$ only through $\widetilde{\tau}$. Thus for a given $\widetilde{\lambda}$, the marginal rate of substitution (MRS) of utility for workers between $\widetilde{w}_1$ and $\widetilde{f}$ equals the MRS of the deviant firm's profit between $\widetilde{w}_1$ and $\widetilde{f}$. This means that keeping $\widetilde{\lambda}$ fixed, the deviant firm is indifferent between delivering U in terms of $\widetilde{w}_1$ and $\widetilde{f}$.

Based on the above discussion, we can rewrite the expected payoff of a worker who applies to the deviant firm. Substituting equation (13) into equation (12) gives

$$U = (1 - h(\widetilde{\lambda})h(\lambda)^{a-1}) - (a-1)h(\widetilde{\lambda})h(\lambda)^{a-2}(1 - h(\lambda))(1 - w_1) - (a-1)f - \widetilde{\tau}. \tag{15}$$

As in equation (12), the first term on the right-hand side gives the probability that the worker receives at least one offer, the second term gives the expected transfer (from the worker to the non-deviant firms) in terms of output and the third term gives the transfer to non-deviant firms in terms of fees. The last term is the expected transfer from the worker to the deviant firm.

The problem of a deviant firm reduces to choosing two variables, $\widetilde{\lambda}$ and $\widetilde{\tau}$, subject to one constraint given by equation (15). Combining equations (10) and (15) gives $\widetilde{\tau}$ as a function of $\widetilde{\lambda}$,

$$\widetilde{\tau} = \left(h(\lambda) - h(\widetilde{\lambda})\right) h(\lambda)^{a-1} + \left(ah(\lambda) - (a-1)h(\widetilde{\lambda})\right) h(\lambda)^{a-2}(1 - h(\lambda))(1 - w_1) + f \tag{16}$$

which implies that

$$\frac{d\widetilde{\tau}}{d\widetilde{\lambda}} = -h'(\widetilde{\lambda}) \left(h(\lambda)^{a-1} + (a-1)h(\lambda)^{a-2}(1 - h(\lambda))(1 - w_1)\right). \tag{17}$$

Since $h(\lambda)$ is strictly increasing, the right-hand side is strictly negative. Hence, if a firm wants to attract more workers, it must accept a lower profit per worker. By equation (14), the deviant firm's first-order condition with respect to $\widetilde{\lambda}$ is simply $0 = \widetilde{\tau} + \widetilde{\lambda} \cdot d\widetilde{\tau}/d\widetilde{\lambda}$, or

$$\widetilde{\tau} = \widetilde{\lambda}h'(\widetilde{\lambda}) \left(h(\lambda)^{a-1} + (a-1)h(\lambda)^{a-2}(1 - h(\lambda))(1 - w_1)\right) \tag{18}$$

Note that the deviant firm's expected profit is strictly concave in $\widetilde{\lambda}$.[10] So the above first-order condition is both necessary and sufficient and implies a unique solution for the deviant firm's optimization problem. From the above equation, it is clear that the optimal $\widetilde{\lambda}$ and $\widetilde{\tau}$ depend on the value of w_1 that the other firms choose. After choosing the optimal $\widetilde{\lambda}^*$ and $\widetilde{\tau}^*$, the deviant firm is indifferent between various combinations of $\widetilde{w}_1$ and $\widetilde{f}$ as long as they deliver $\widetilde{\tau}^*$ (see equation (13)). Note that the workers who applied to the deviant firm also only care about their promised utility and not the combination of $\widetilde{w}_1$ and $\widetilde{f}$ that delivers the utility.

Equating the right-hand sides of equations (16) and (18) yields the optimal $\widetilde{\lambda}^*$. For there to be no profitable deviations, the optimal $\widetilde{\lambda}^*$ must be λ. Thus equilibrium requires

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

Rewriting the above equation using $\frac{\lambda h'(\lambda)}{1-h(\lambda)} = 1-\frac{\lambda m'(\lambda)}{m(\lambda)}$ yields,[11]

$$f = f^e(w_1) \equiv h(\lambda)^{a-1}(1-h(\lambda))\left(w_1 - \frac{\lambda m'(\lambda)}{m(\lambda)} + (1-w_1)(a-1)\frac{\lambda h'(\lambda)}{h(\lambda)}\right) \tag{19}$$

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 urn ball and the geometric, the elasticity of $h(\lambda)$, i.e., $\lambda h'(\lambda)/h(\lambda)$ as defined in (2), is strictly decreasing in λ with supremum equal to 1 and 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 decreasing in w_1 for small λ. The above observations are illustrated in Fig. 1, where the meeting technology is given by the urn ball. 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 $\lambda = 3$ but decreasing in w_1 when $\lambda = 1$ and 2 (the dashed lines in the figure refer to the planner's solution and are discussed in the next section).

To summarize, we have shown that if all firms except a potential deviant choose f and w_1 satisfying equation (19), then a potential deviant's best response is,

$$\begin{aligned} &\widetilde{\lambda}^* = \lambda \\ &\widetilde{\tau}^* = h(\lambda)^{a-1}(1-h(\lambda))(1-w_1)+f \end{aligned}$$

Thus an individual firm has a continuum of best responses (as long as the expected profit per worker equals $\widetilde{\tau}^*$), which includes the equilibrium (f, w_1) as a special case.

[10] To show that $\widetilde{\pi}$ is concave note that $\frac{d^2\widetilde{\pi}}{d\widetilde{\lambda}^2} = 2\frac{d\widetilde{\tau}}{d\widetilde{\lambda}} + \widetilde{\lambda}\frac{d^2\widetilde{\tau}}{d\widetilde{\lambda}^2}$. Since $d\widetilde{\tau}/d\widetilde{\lambda}$ is given by equation (17), we have

$$\frac{d^2\widetilde{\tau}}{d\widetilde{\lambda}^2} = -h''(\widetilde{\lambda})(h(\lambda)^{a-1} + (a-1)h(\lambda)^{a-2}(1-h(\lambda))(1-w_1),$$

so the sign of $d^2\widetilde{\pi}/d\widetilde{\lambda}^2$ is that of $-2h'(\lambda) - \lambda h''(\lambda) = m''(\lambda) < 0$.

[11] In equation (19), we have chosen to express f using both $m(\lambda)$ and $h(\lambda)$ because this makes the comparison with the planner's solution easier. This comment also applies to the expected payoff of firms in equation (20). Equation (2) gives the relationship between the derivatives of $m(\lambda)$ and $h(\lambda)$.

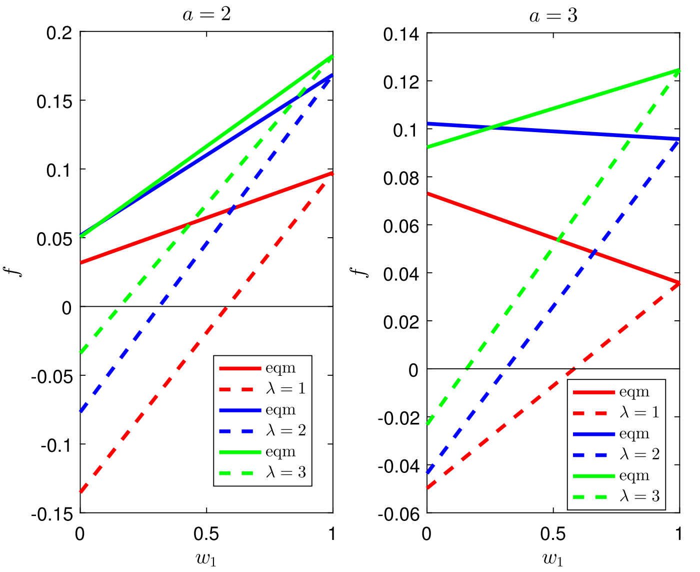

Fig. 1. Equilibrium and efficient combinations of (w_1, f) with $m(\lambda) = 1 - e^{-\lambda}$ and $\lambda = 1, 2, 3$. Equilibrium loci are solid and efficient loci are dashed. $f < 0$ implies an application subsidy.

Focusing on *symmetric* equilibria in which all firms choose the same f and w_1, there exists a continuum of equilibria. For each $w_1 \in [0, 1]$, there exists a corresponding f that satisfies equation (19), giving a continuum of equilibria that we can index by w_1. These equilibria are not payoff equivalent as we now show.

Substituting equation (19) into equation (9) gives the expected equilibrium payoff of firms,

$$\pi^e(w_1, \lambda, a) = \big(m(\lambda) - \lambda m'(\lambda)\big)\, h(\lambda)^{a-1} \left(1 + (1 - w_1)(a - 1)\frac{1 - h(\lambda)}{h(\lambda)}\right) \tag{20}$$

From the above equation we can see that firm payoff is linear in w_1 and that it is *strictly decreasing* in w_1. Thus different equilibria are not payoff equivalent. Since for fixed λ, total surplus is constant across equilibria, the worker payoff is strictly increasing in w_1 because the total surplus is shared by the worker and the firm.

Next we give an explicit expression for the market utility of workers by substituting equation (19) into equation (10). We then have

$$U^e(w_1, \lambda, a) = 1 - h(\lambda)^a - a\lambda h'(\lambda)h(\lambda)^{a-1} - a(a-1)\lambda h'(\lambda)(1 - h(\lambda))h(\lambda)^{a-2}(1 - w_1). \tag{21}$$

The above equation is linear and increasing in w_1 unless $a = 1$. 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 3, is strictly positive.[12] However, in the equilibrium with $w_1 = 0$,

[12] 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 (21), $U^e(1, \lambda, a) = 1 - h(\lambda)^a - a\lambda h'(\lambda)h(\lambda)^{a-1}$. 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.

$U^e(0, \lambda, a)$ 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. *(i) There exists a continuum of equilibria indexed by $w_1 \in [\underline{w}, 1]$ where f is given by equation (19). 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 $w_1 = \underline{w}$. (ii) 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.*

Since a deviant firm has a continuum of best responses, a continuum of equilibria occur naturally. When workers make multiple applications, these equilibria are not payoff equivalent. In this case, a firm can provide value to a worker in two ways. First, if the firm hires the worker when the worker has no other offers, the firm gives the worker a share w_1 of the match surplus. Second, if the firm selects the worker but the worker accepts a different offer, the firm's offer, even though it isn't accepted, allows the worker to capture the entire surplus at the other firm. In equilibria with high values of w_1, the firm payoff comes mainly from application fees, and workers care mostly about the probability of receiving at least one offer. However, in equilibria with low values of w_1, most of the value in a worker's application comes from its potential to help the worker extract full surplus from another firm. In this case, firms are attracting applications not primarily by promising a share of their own match surplus but rather by helping the worker extract surplus from other firms. In this sense, firms are free-riding on each other. In equilibria with low values of w_1, the worker payoff is low. In free-riding on one another, firms are exercising market power.

We can now contrast our result with a competitive search model in which workers can only send one application. The first part of Proposition 1 continues to hold, i.e., there is a continuum of equilibria indexed by $w_1 \in [0, 1]$ where f is given by equation (19). When $a = 1$, equation (19) 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, when workers can only send one application, all these equilibria are payoff equivalent. When $a = 1$, equation (20) continues to hold. It becomes $\pi^e(w_1, \lambda, 1) = m(\lambda) - \lambda m'(\lambda)$. That is, firm payoff is independent of w_1 so all equilibria are payoff equivalent.

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 $(f^e(w_1) < 0)$. 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 $\lambda > 0$,*

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq 1 - h(\lambda). \tag{22}$$

To understand this condition, note that 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.

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

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq \frac{\lambda m'(\lambda)}{m(\lambda)}, \tag{23}$$

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

Proof. See Appendix A.2. □

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 (19) is linear in w_1 and note that

$$f^e(0) = h(\lambda)^{a-1}(1-h(\lambda))\left(-\frac{\lambda m'(\lambda)}{m(\lambda)} + (a-1)\frac{\lambda h'(\lambda)}{h(\lambda)}\right), \tag{24}$$

$$f^e(1) = h(\lambda)^{a-1}(1-h(\lambda))\left(1-\frac{\lambda m'(\lambda)}{m(\lambda)}\right) = \lambda h(\lambda)^{a-1}h'(\lambda). \tag{25}$$

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. With multiple applications, the last term on the right-hand side of equation (24) 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 2, 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 (19), $f^e(w_1)$ 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.

We first present the total number of matches, or equivalently total surplus. Since workers send their a applications independently, the total number of matches is given by

$$M(u, v, a) = u\left(1 - h\left(\frac{au}{v}\right)^{a}\right). \tag{26}$$

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

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

$$S(u, v, a) = M(u, v, a) - c_v v. \tag{27}$$

The first-order condition for this problem is $c_v = \partial M(u, v, a)/\partial v$. In Lemma 3 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. This contribution is

$$\frac{\partial M(u, v, a)}{\partial v} = \left(m(\lambda) - \lambda m'(\lambda)\right) h(\lambda)^{a-1}. \tag{28}$$

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, $m(\lambda) - \lambda m'(\lambda)$ 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.[13]

Similarly, the marginal contribution of a worker is

$$\frac{\partial M(u, v, a)}{\partial u} = am'(\lambda)h(\lambda)^{a-1} + \left(1 - h(\lambda)^{a} - ah(\lambda)^{a-1}(1 - h(\lambda))\right). \tag{29}$$

When $a = 1$, $m'(\lambda)$ is the worker's marginal contribution. For the urn-ball meeting technology, this term 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.

[13] 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.

Lemma 3. *$M(u, v, a)$, which was defined in equation (26), is strictly concave in v and in u.*

Proof. See Appendix A.3. □

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 (20). What complicates the problem is that there is a continuum of equilibria, which are not payoff equivalent. Comparing equations (20) and (28) shows that in the equilibrium with $w_1 = 1$, the expected firm value is equal to the firm's marginal contribution to net surplus for all a. Hence, the equilibrium with $w_1 = 1$ is 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 (19). 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 social surplus.*

Proof. See Appendix A.4. □

Note that by free entry firms receive $c_v v$ so by equation (27), the expected payoff of a worker is net surplus per capita. Thus, again workers prefer an equilibrium with a higher w_1. Note also that the expected equilibrium payoff of a worker can never be negative. That is, at $w_1 = \underline{w}$, the expected payoff of workers is zero.

Understanding the efficiency result When $w_1 = 1$, workers always receive the full match surplus. Firms do not impose externalities on each other in the matching stage because their share of the match surplus is zero. Firms compete for workers' applications through application fees. Competition among firms then implies that in equilibrium, the application fee is efficient. Alternatively, we can understand this efficiency result 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 when $w_1 < 1$ With multiple applications, when $w_1 < 1$, a firm's payoff depends on the choices of other firms. When a firm's selected candidate has no other offers, the firm's payoff is $1 - w_1$ plus the income from fees, otherwise its payoff only consists of fee income. Excessive entry occurs because as we noted in the last section, firms free ride on one another in the sense that part of the worker's expected payoff from applying to a firm is supplied by other firms. Alternatively, we can understand excessive entry by calculating the fee that a social planner would set, $f^p(w_1)$, which is found by equating the expected payoff of a firm to its marginal contribution. That is,

$$f^p(w_1) = h(\lambda)^{a-1}(1 - h(\lambda))\left(w_1 - \frac{\lambda m'(\lambda)}{m(\lambda)}\right) \tag{30}$$

which is derived by setting (9) equal to equation (28).

Comparing the above equation with equation (19), which gives the equilibrium $f^e(w_1)$, shows that

$$f^p(w_1) < f^e(w_1) \quad \text{when} \quad w_1 < 1 \tag{31}$$

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 Fig. 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 (30) 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 social surplus. Proposition 3 confirms that this is indeed the case.

Because $M(u, v, a)$ is concave in v, net social surplus $S(u, v, a)$ defined in equation (27) is also concave in v. We already know that the number of vacancies 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 (28) 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 (25), leads to the efficient outcome. In this section we investigate whether this (in)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 the other workers. The effect of multiple applications on the firm side is ambiguous. On the one hand, the probability that a firm receives 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.[14]

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. We simplify the analysis in two ways. First, we ignore the constraint that a must be an integer and treat it as a continuous variable in both the planner's problem and the worker's problem in the decentralized equilibrium. Second, we take the number of firms as given.

5.1. The social planner's problem

The social planner's problem is to choose a to maximize $M(u, v, a) - uac_a$, total matches minus the cost of applications. The planner's first-order condition is,[15]

$$c_a = \frac{d}{da}\left(\frac{M(u, v, a)}{u}\right) = h(\lambda)^a \left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}\right). \tag{32}$$

The effect of a on the number of matches depends on the difference between two positive terms. The term $\frac{\lambda h'(\lambda)}{h(\lambda)}$ reflects a congestion effect. With a higher a, workers create more congestion for other workers by increasing λ, which reduces the probability of being matched per application. The term $-\log h(\lambda)$ reflects a sampling effect. More applications by an individual worker increase that worker's matching rate. The relative size of these effects depends on λ. Since $\lim_{\lambda\to 0}(-\log h(\lambda)) = \infty$ and by a first-order Taylor series expansion, $\lim_{\lambda\to 0} \frac{\lambda h'(\lambda)}{h(\lambda)} = 1$, the sampling effect dominates the congestion effect when λ is small, and so a higher a always leads to a larger number of matches when λ is small. As λ increases, the congestion effect becomes larger and the sampling effect gets smaller 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 (32) as a function of a is illustrated by the red solid line in Fig. 2. The congestion effect dominates when $a \cdot u/v = \lambda > \lambda^* = 2.66$, i.e., the right-hand side of equation (32) 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 (32) is always positive.[16]

We showed above that as $\lambda \to 0$, the right-hand side of equation (32) goes to infinity. We can also show that as $\lambda \to \infty$, the right-hand side of equation (32) goes to zero. Thus, there is at least one solution to the first-order condition. To show that the first-order condition has a unique solution, we impose the following condition on the meeting technology, which ensures that $M(u, v, a)$ is concave in a when a is below some critical value.

---

[14] Albrecht et al. (2006) showed this formally, see also equation (32).

[15] We use the following in deriving equation (32). For $f(x) > 0$,

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

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

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

When the equation in Condition 2 has no root, we set $\lambda^r = \infty$.

To see the significance of Condition 2, note that by equation (32), the sign of $\frac{\partial}{\partial a}(M(u,v,a)/u)$ depends only on $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$, which is a function of λ alone, and that

$$\frac{d}{d\lambda}\left(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}\right) = \frac{h'(\lambda)}{h(\lambda)}\left(\frac{\lambda h'(\lambda)}{h(\lambda)} - \frac{\lambda h''(\lambda)}{h'(\lambda)} - 2\right). \tag{33}$$

As we noted above, $\lim_{\lambda\to 0}\frac{\lambda h'(\lambda)}{h(\lambda)} = 1$. Using a Taylor series expansion, we also have $\lim_{\lambda\to 0}\frac{\lambda h''(\lambda)}{h'(\lambda)} = 0$. 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., $\lambda^r = \infty$, which is the case for the geometric meeting technology. For other meeting technologies such as the urn ball, the function crosses the horizontal axis exactly once.

Therefore, by Condition 2 when $\lambda < \lambda^r$, we have that $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is strictly decreasing, and when $\lambda > \lambda^r$, it is strictly increasing. We distinguish two scenarios. First, consider the case $\lambda^r < \infty$. Then $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is first decreasing and reaches a minimum at $\lambda = \lambda^r$. For $\lambda > \lambda^r$, it is strictly increasing and approaches its limit value of zero. The 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 $\lambda^m < \lambda^r$. Next, consider the case $\lambda^r = \infty$. In this case, $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is monotonically decreasing for $\lambda > 0$. Our discussion above implies that $(-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)})$ is also decreasing in a when a is small ($a \cdot u/v \le \lambda^r$). 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* $\lambda^m \le \lambda^r$ *such that* $M(u,v,a)$ *is strictly increasing in a when* $\lambda = au/v \le \lambda^m$ *and strictly decreasing in a when* $au/v \ge \lambda^m$. *Furthermore,* $M(u,v,a)$ *is strictly concave in a when* $au/v \le \lambda^r$.

Proof. See Appendix A.5. □

Since the planner's optimal choice of a must be such that $au/v \le \lambda^m$ and since $M(u,v,a)$ is concave in a in that region, the planner's first-order condition is both necessary and sufficient. The solution to the planner's problem is illustrated in Fig. 2 with $m(\lambda) = 1 - e^{-\lambda}$, which gives $\lambda^m = 2.66$ and $\lambda^r = 4.20$. From the figure, we can see that when $\lambda < \lambda^r$, 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 also have that when $\lambda^r = \infty$, which implies that $\lambda^m = \infty$, $M(u,v,a)$ is always increasing and concave in a.

5.2. Market equilibrium when a is endogenous

Next, we consider the choice of a by workers in the decentralized equilibrium. We again look for a symmetric, pure-strategy equilibrium in which all firms choose the same (f, w_{ij}); however, now a is endogenous and depends on the wage mechanisms that firms post. Our starting point is again Lemma 1, which continues to hold.

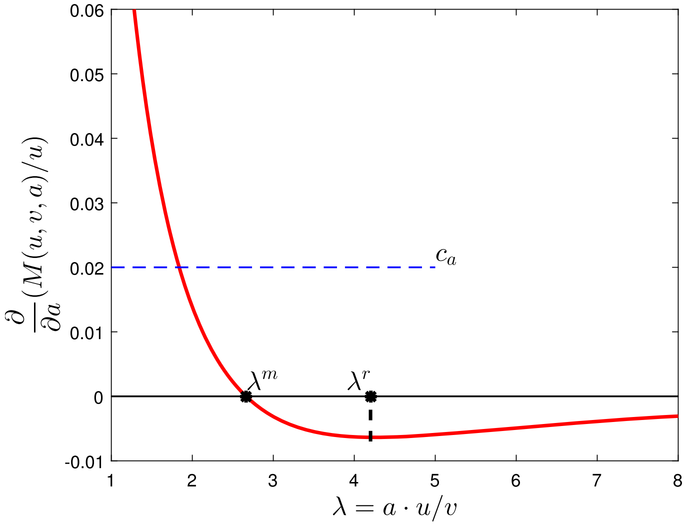

Fig. 2. Illustration of Proposition 4 where $m(\lambda) = 1 - e^{-\lambda}$, $u/v = 1$ and $c_a = 0.02$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

The expected payoffs of workers and firms in equilibrium are calculated as before, except that now a is a choice variable. Suppose that in equilibrium firms post f and w_1. By equation (10), the worker's problem is given by,

$$U = \max_{a'} 1 - h(\lambda)^{a'} - a'(1-h(\lambda))h(\lambda)^{a'-1}(1-w_1) - a'f - a'c_a. \tag{34}$$

Note that workers choose a' optimally while taking λ as given. Similarly, a firm's equilibrium payoff is given by equation (9), where a is the equilibrium number of applications that workers send and λ is equilibrium market tightness. Lemma 4 below implies that the optimal a' is unique.

The problem of a deviant firm becomes more complicated when workers can adjust a. As before, suppose that a deviant firm posts $\widetilde{f}$ and $\widetilde{w}_1$ and attracts a queue with length $\widetilde{\lambda}$, and that workers who apply to the deviant firm send out $\widetilde{a}$ applications. Then as in equation (11), the deviant firm's expected profit is given by

$$\widetilde{\pi}(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f}) = m(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-1}(1-\widetilde{w}_1) + \widetilde{\lambda}\widetilde{f}. \tag{35}$$

Note that for given $\widetilde{\lambda}$, $\widetilde{w}_1$ and $\widetilde{f}$, the deviant firm's expected payoff is higher if $\widetilde{a}$ is lower because the deviant firm then faces less competition. Similarly, by equation (12), the problem of workers who apply to the deviant firm is given by

$$\begin{aligned} V_1(\widetilde{\lambda}, \widetilde{w}_1, \widetilde{f}) \equiv \max_{\widetilde{a}} V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f}) &\equiv 1 - h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-1} \\ &- (\widetilde{a}-1)h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-2}(1-h(\lambda))(1-w_1) \\ &- (1-h(\widetilde{\lambda}))h(\lambda)^{\widetilde{a}-1}(1-\widetilde{w}_1) - \widetilde{f} - (\widetilde{a}-1)f - \widetilde{a}c_a. \end{aligned} \tag{36}$$

Note that the optimal choice of $\widetilde{a}$ for workers who apply to the deviant firm does not depend on $\widetilde{f}$, which means that we can denote the solution to (36) by $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$. Once a worker decides to apply to the deviant firm, the application fee is sunk and does not affect the number of additional applications that the worker sends. The choice of $\widetilde{a}$ does, however, depend on $\widetilde{w}_1$ because the higher $\widetilde{w}_1$ is, the lower is the marginal value of another offer and consequently the lower is the marginal benefit of sending an additional application.

Finally, the queue length $\widetilde{\lambda}$ faced by the deviant firm is again determined by the worker indifference condition.

$$U = V_1(\widetilde{\lambda}, \widetilde{w}_1, \widetilde{f}) = V_0(\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1), \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f}). \quad (37)$$

That is, the queue length $\widetilde{\lambda}$ is such that workers are indifferent between sending a applications to the non-deviants and sending one application to the deviant and $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1) - 1$ applications to the non-deviants. Before we proceed further, we present the following lemma which contains three useful technical results.

Lemma 4. *i) $V_1(\widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})$ is strictly decreasing in $\widetilde{\lambda}$. ii) the solution to problem (36) is unique so that $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$ is well-defined. iii) For given $\widetilde{\lambda}$, $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$ is strictly decreasing in $\widetilde{w}_1$.*

Proof. See Appendix A.6. □

The first part of the above lemma shows that equation (37) admits a unique solution $\widetilde{\lambda}$. The second part states that the optimal $\widetilde{a}$ in problem (36) is unique. Note that in the special case in which no firm deviates ($\widetilde{w}_1 = w_1$ and $\widetilde{\lambda} = \lambda$), this part implies that workers' equilibrium number of applications, the optimal a' in problem (34), is also unique. The third part shows that the optimal number of applications for workers who apply to the deviant firm is strictly decreasing in $\widetilde{w}_1$.

When a is exogenous, a deviant firm has a continuum of best responses. Both workers and the deviant firm only care about the expected profit per worker as defined by equation (13). No firm has an incentive to deviate from a candidate equilibrium, $(f^e(w_1), w_1)$. However, this logic breaks down when the number of applications is endogenous. To see this, consider a candidate equilibrium in which $w_1 < 1$ and $a = a^*$. The deviant firm can set a wage of $\widetilde{w}_1 = 1$ and adjust its fee to offer the same expected utility to a worker who sends a^* applications. This leaves the deviant firm with the same expected profit. However, when workers can adjust their applications, the workers who apply to the deviant firm will send fewer applications to other firms. This saves them application fees, so they are better off. The deviant firm can capture some of this gain by adjusting its fee. This makes offering $\widetilde{w}_1 = 1$ a profitable deviation. The existence of a profitable upward deviation holds for any $w_1 < 1$, so the unique equilibrium is the one with $w_1 = 1$ and the corresponding fee.

The above discussion shows that on and off the equilibrium path, workers always receive the full surplus from the job after paying the application fees. We can again use the analogy of competitive markets to understand this. Firms essentially supply lotteries in a competitive market and workers can choose the number of lottery tickets they buy. Thus both firm entry and the workers' applications are constrained efficient. The following proposition formally summarizes this.

Proposition 5. *When a is endogenous and Condition 2 holds, there exists a unique symmetric, pure-strategy equilibrium in which all firms choose $w_1 = 1$, f is given by equation (25) and workers' equilibrium number of applications is uniquely determined by equation (32). The decentralized equilibrium is efficient.*

Proof. See Appendix A.7. □

Note that workers have two reasons to send more than one application. First, this increases the probability of getting an offer; second, it increases the probability of getting multiple offers and accordingly to receive 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. Conventional competition among firms is not enough for efficiency. As we showed in Section 3, there are multiple equilibria and only one is efficient. Ultimately, it is the addition of an extensive margin, the worker's choice of a, that eliminates the multiplicity of equilibria and leads to a unique efficient equilibrium.

For simplicity we have treated a as a continuous variable while working with expressions derived under the assumption that a is discrete, for example, equations (34), (35), and (36). Alternatively, one could restrict a to be an integer. In that case, (i) there might not exist a symmetric, pure-strategy equilibrium, and (ii) any equilibrium would be characterized by inequalities, e.g., sending out a applications is better than sending out $a - 1$ or $a + 1$ applications. Despite the complexity, one can see that a deviant firm has an incentive to increase $\widetilde{w}_1$ while adjusting $\widetilde{f}$ so that the workers who apply to the deviant can make fewer applications elsewhere. However, when $\widetilde{a}$ is discrete, then if $\widetilde{w}_1$ is sufficiently close to 1, $\widetilde{a}$ will no longer respond to a higher $\widetilde{w}_1$, so a deviant firm again has a continuum of best responses implying that the discreteness in applications can be a (limited) source of market power. Nonetheless, even with a discrete, endogenizing the number of worker applications pushes the equilibrium towards efficiency.

6. Contract frictions

So far, we have allowed firms to compete with general wage mechanisms that include fees and wages conditioned on the number of other offers that the worker has. 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 the monopsony wage, which in their model is $w_1 = 0$. When we allow application fees, an equilibrium with $w_1 = 0$ also exists. Since there is excessive vacancy creation in Albrecht et al. (2006), allowing application fees makes the equilibrium with $w_1 = 0$ even less efficient. Finally, we note that our result generalizes the equilibrium of Albrecht et al. (2006) to a wide class of meeting technologies so 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.8, 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,

$$\lambda_H = 2\lambda_L. \tag{38}$$

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.8).[17] 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 (25).

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.

[17] This conclusion holds for both our definition of total surplus (equation (26)) and for that of Galenianos and Kircher (2009), which is $u[(1 - h(\lambda_H)) + h(\lambda_H)(1 - h(\lambda_L))]$.

Proposition 6. *When firms can charge an application fee but 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 $w = 0$. The equilibrium is not efficient.*

Proof. See Appendix A.9. □

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.9 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

In this paper, we investigate the efficiency of competitive search equilibrium in a labor market in which workers make multiple applications. We allow for a general nonrival meeting technology and for quite general wage mechanisms. Specifically, we allow each firm to condition its wage offer on (i) the number of offers its selected applicant receives and (ii) the number of applications the firm receives. We also allow firms to post an application fee or offer an application subsidy.

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 the firm promises to pay a worker who has no other offers, w_1, and (iii) the wage that the firm offers to a worker who has multiple offers. Competition in posted wage mechanisms drives this latter wage to 1, the value of a worker's output. When the number of applications a worker can make is given exogenously, we show that there exists a continuum of symmetric equilibria, i.e., equilibria in which all firms post the same (f, w_1) pair. 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 efficiency. All other equilibria lead to excessive vacancy creation.

Why are the equilibria that we derive generically inefficient? When a worker sends an application to a firm and pays the fee, the potential payoff to that application has two components. First, if the firm offers the worker a job and that is the worker's only offer, then the worker gets the wage the firm promised to pay when its selected applicant has no other offers. Second, if the firm offers the worker a job and the worker has another offer, then the worker gets the full value of the match output. When it comes to attracting worker applications, the relative importance of these two components depends on the mechanisms that other firms are posting. If other firms post a low value of w_1, then workers are primarily concerned with getting multiple offers, and an individual firm can post a low value of w_1 without losing many applications. A deviation to a higher w_1 is not attractive because it only leads to a small increase in fee income. However, if other firms post a high value of w_1, then generating multiple offers isn't as important for workers and the firm's payoff mainly comes from the fees. In this case, posting a lower w_1 substantially reduces fee income so the best an individual firm can do is to also post a high value of w_1. If other firms post a low value of w_1, 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 firms have. When $w_1 = 1$, this monopsony power vanishes. In this case, the level of vacancy creation only depends on f, and the application fee adjusts so that the payoff to the marginal firm equals the contribution its entry makes to net output.

We derived our main results taking the number of applications, a, per worker as exogenous. In an extension, we assume an application cost and endogenize a. When we do this, only the efficient equilibrium survives, and the worker choice of a is efficient. Why does endogenizing a eliminate the inefficient equilibria with $w_1 < 1$? A firm can deviate from w_1 by offering a wage of 1 while adjusting f to offer the same expected payoff to its applicants. With a fixed, this deviation also leaves the firm's expected payoff unchanged. However, workers who apply to the deviant firm now have less incentive to apply to other firms, so these workers can save on application costs. By deviating to a wage of 1, the firm can adjust its application fee to capture part of this savings.

The conventional wisdom is that competitive search equilibrium delivers vacancy creation at the socially optimal level, but this result is based on the unrealistic assumption that workers apply for only one job at a time. Building on Albrecht et al. (2006) and Galenianos and Kircher (2009), we analyze the efficiency of competitive search equilibrium in a much more general setting than was previously considered.

Appendix A. Proofs

A.1. *Proof of Lemma 1*

Suppose that $p_j(\lambda) > 0$, and $w_{ij} < 1$ for some $i \geq 2$ and $j \geq 1$. Consider a deviant firm that posts $\widetilde{w}_{ij} = w_{ij} + \varepsilon$ and $\widetilde{f} = f + \Delta f$. The deviant firm can choose Δf such that equation (6) holds at $\widetilde{\lambda} = \lambda$, i.e., the deviant firm adjusts its application fee so that the expected number of workers that it attracts stays the same. Note that $\Delta f \leq \varepsilon$ since the extra wage ε is only received by a worker if offered the job which happens with some probability, while the extra application fee Δf is paid by the worker for sure. Let $\varepsilon \searrow 0$ (hence the corresponding $\Delta f \searrow 0$), so the gain (resp. loss) from the higher application fee (resp. higher wage) is negligible; however, the deviant firm still enjoys a discrete jump in the winning probability. Specifically, $\widetilde{\mathcal{P}}_{ij}$ in equations (7) and (8) increases by *at least* $\binom{a-1}{i-1} h(\lambda)^{a-i} (q_j(\lambda)/j)^{i-1}$, which corresponds to the scenario in which the deviant firm receives j applications and its chosen worker has $i \geq 2$ offers in total and all other $i-1$ competitors receive exactly j applications (note that in equation (7), $\widetilde{\mathcal{P}}_{ij}$ is multiplied by $p_j(\widetilde{\lambda})$ and by the construction of Δf, $\widetilde{\lambda} = \lambda$). Finally, note that the deviant firm can alternatively choose to reduce w_{1j} instead of increasing f, and the above argument continues to hold. □

A.2. *Proof of Lemma 2*

First consider (22) ⇒ (23). Equation (22) implies

$$\frac{\lambda h'(\lambda)}{h(\lambda)} \geq 1 - \frac{\lambda h'(\lambda)}{1 - h(\lambda)} \tag{39}$$

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 (22) ⇒ (23).

Next, consider (23) $\Rightarrow$ (22). 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 (23) in terms of $h(\lambda)$ only, which is exactly (39). Multiplying both sides of (39) by $h(\lambda)(1 - h(\lambda))$ gives $\lambda h'(\lambda) \geq h(\lambda)(1 - h(\lambda))$, which is exactly (22).

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 (23) holds. □

A.3. *Proof of Lemma 3*

Recall that from equation (28), we have $\partial M(u, v, a)/\partial v = \left(m(\lambda) - \lambda m'(\lambda)\right) h(\lambda)^{a-1}$. 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.4. *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 (20). To start, we ignore the 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 (20),

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

The first two terms are positive; the third term is negative. Since the first term is positive, it suffices to show that the sum of the second term and third terms is positive. That is, we need to verify that

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

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 (20) yields $\pi^e(1, \lambda, a) = h(\lambda)^{a-1}\left(m(\lambda) - \lambda m'(\lambda)\right)$, which is exactly the right-hand side of equation (28). 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 payoff 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 (20) 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 3, 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 the worker expected payoff in equilibrium must 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. Suppose that the worker expected payoff would be negative were $w_1 = 0$. Then, 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.5. *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 (33), when $\lambda < \lambda^r$, $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ is strictly decreasing, and when $\lambda > \lambda^r$, 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 $\lambda^r = \infty$, then $-\log h(\lambda) - \frac{\lambda h'(\lambda)}{h(\lambda)}$ is strictly decreasing in λ, ranging from ∞ when $\lambda = 0$ to zero when $\lambda = \infty$, and (ii) when $\lambda^r < \infty$, it is first decreasing until its minimum value is reached at $\lambda = \lambda^r$ 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 $\lambda^r < \infty$, there must exist a unique λ^m 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

$$\frac{\partial^2 M(u,v,a)}{\partial a^2} = -h(\lambda)^a\left(\left(\log h(\lambda) + \frac{\lambda h'(\lambda)}{h(\lambda)}\right)^2 + \frac{1}{a}\frac{\lambda h'(\lambda)}{h(\lambda)}\left(2 - \frac{\lambda h'(\lambda)}{h(\lambda)} + \frac{\lambda h''(\lambda)}{h'(\lambda)}\right)\right) \tag{40}$$

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

A.6. Proof of Lemma 4

i) From equation (36) we can see that $V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})$ is linear in $h(\widetilde{\lambda})$ with a negative coefficient. Since $h(\widetilde{\lambda})$ is strictly increasing in $\widetilde{\lambda}$, $V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})$ is strictly decreasing in $\widetilde{\lambda}$. Given $\widetilde{w}_1$ and $\widetilde{f}$, consider two queues with $\widetilde{\lambda}' < \widetilde{\lambda}''$. Suppose that $V_0(\widetilde{a}, \widetilde{\lambda}'', \widetilde{w}_1, \widetilde{f})$ is maximized at $\widetilde{a}''$. Then $V_1(\widetilde{\lambda}', \widetilde{w}_1, \widetilde{f}) \geq V_0(\widetilde{a}'', \widetilde{\lambda}', \widetilde{w}_1, \widetilde{f}) > V_0(\widetilde{a}'', \widetilde{\lambda}'', \widetilde{w}_1, \widetilde{f}) = V_1(\widetilde{\lambda}'', \widetilde{w}_1, \widetilde{f})$, where the first inequality comes from the definition of V_1 and the second inequality is because V_0 is strictly decreasing in $\widetilde{\lambda}$.

ii) From equation (36) we have

$$\frac{\partial V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})}{\partial \widetilde{a}} = \frac{\partial}{\partial \widetilde{a}}\Big(1 - h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-1} - (\widetilde{a}-1)h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-2}(1-h(\lambda))(1-w_1) - (\widetilde{a}-1)f - \widetilde{a}c_a\Big) - (1-h(\widetilde{\lambda}))\,(1-\widetilde{w}_1)\,h(\lambda)^{\widetilde{a}-1}\log h(\lambda). \tag{41}$$

$$\frac{\partial^2 V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})}{\partial \widetilde{a}^2} = h(\lambda)^{\widetilde{a}-2}\log h(\lambda)\Big[-\log h(\lambda)\cdot(1-h(\lambda))h(\widetilde{\lambda})(1-w_1)(\widetilde{a}-1) - \log h(\lambda)h(\lambda)\big(1-(1-h(\widetilde{\lambda}))\widetilde{w}_1\big) - 2(1-h(\lambda))h(\widetilde{\lambda})(1-w_1)\Big] \tag{42}$$

Equation (42) implies that since $h(\lambda) < 1$, the sign of $\partial^2 V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})/\partial\widetilde{a}^2$ is opposite to the term in between square brackets, which is linear and increasing in $\widetilde{a}$. Thus as a function of $\widetilde{a}$, $\partial^2 V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})/\partial\widetilde{a}^2$ crosses the x-axis at most once and from above, i.e., $V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})$ is first possibly convex and then concave. Thus the optimal $\widetilde{a}$ must be unique (assuming that the corner solution $\widetilde{a} = 0$ is never optimal so that workers are willing to apply to the deviant firm).

iii) Note by equation (41), $\partial^2 V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})/\partial\widetilde{a}\partial\widetilde{w}_1 = (1-h(\widetilde{\lambda}))h(\lambda)^{\widetilde{a}-1}\log h(\lambda) < 0$, $V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})$ is strictly submodular in $(\widetilde{a}, \widetilde{w}_1)$. Therefore, $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$ is strictly decreasing in $\widetilde{w}_1$. □

A.7. Proof of Proposition 5

Our first step is, as in Section 3.2, to represent a deviant firm's expected profit in the form of equation (14), i.e., $\widetilde{\pi} = \widetilde{\lambda}\widetilde{\tau}$ where $\widetilde{\tau}$ is defined by (13) as

$$\widetilde{\tau} = (1-h(\widetilde{\lambda}))h(\lambda)^{a-1}(1-\widetilde{w}_1) + \widetilde{f}.$$

Combining equations (36) and (37) yields

$$\widetilde{\pi} = \widetilde{\lambda}\Big[1 - h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-1} - (\widetilde{a}-1)h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-2}(1-h(\lambda))(1-w_1) - (\widetilde{a}-1)f - \widetilde{a}c_a - U\Big], \tag{43}$$

where $\widetilde{a} = \widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$, and the term in brackets is the expected profit per worker. As before, $\widetilde{w}_1$ and $\widetilde{f}$ dropped out from the above equation. The above equation suggests that instead of $\widetilde{w}_1$ and $\widetilde{f}$, we can treat the deviant firm's choice variables as $\widetilde{a}$ and $\widetilde{\lambda}$. Given the optimal $\widetilde{a}$ and $\widetilde{\lambda}$, the optimal $\widetilde{w}_1$ can be backed out uniquely because $\widetilde{a}^*(\widetilde{\lambda}, \widetilde{w}_1)$ is strictly decreasing in $\widetilde{w}_1$. Next, given $\widetilde{a}$, $\widetilde{\lambda}$, and $\widetilde{w}_1$, the optimal $\widetilde{f}$ can be backed out by equation (36).

The first-order condition for the maximization of $\widetilde{\pi}$ with respect to $\widetilde{a}$ for the deviant firm is

$$\frac{\partial\widetilde{\pi}}{\partial\widetilde{a}} = \widetilde{\lambda}\frac{\partial}{\partial\widetilde{a}}\Big(1 - h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-1} - (\widetilde{a}-1)h(\widetilde{\lambda})h(\lambda)^{\widetilde{a}-2}(1-h(\lambda))(1-w_1)$$

$$- (\widetilde{a} - 1) f - \widetilde{a} c_a \Big)$$

Recall that the first-order condition of the workers who apply to the deviant firm is $0 = \partial V_0(\widetilde{a}, \widetilde{\lambda}, \widetilde{w}_1, \widetilde{f})/\partial \widetilde{a}$. Thus by equation (41), the equation above can be rewritten as

$$\frac{\partial \widetilde{\pi}}{\partial \widetilde{a}} = \widetilde{\lambda}(1 - h(\widetilde{\lambda}))\,(1 - \widetilde{w}_1)\,h(\lambda)^{\widetilde{a}-1} \log h(\lambda)$$

As long as $\widetilde{w}_1 < 1$, reducing $\widetilde{a}$ strictly increases the expected profit of the deviant firm. Therefore, the deviant firm should always strictly prefer to increase $\widetilde{w}_1$ no matter what value of (f, w_1) other firms post. This in turn implies that in any possible equilibrium, we must have $w_1 = 1$, and when considering deviations, we can set $\widetilde{w}_1 = 1$. Next, we calculate the equilibrium application fee f.

Along the equilibrium path, the workers' problem defined in equation (34) now reduces to

$$U = \max_{a'} 1 - h(\lambda)^{a'} - a' f - a' c_a$$

The above problem is concave, so the following first-order condition is both necessary and sufficient.

$$-h(\lambda)^a \log h(\lambda) = f + c_a \tag{44}$$

The problem of workers who apply to the deviant firm reduces to

$$\max_{\widetilde{a}} 1 - h(\widetilde{\lambda}) h(\lambda)^{\widetilde{a}-1} - \widetilde{f} - (\widetilde{a} - 1) f - \widetilde{a} c_a.$$

The above problem is concave in $\widetilde{a}$ so that the first-order condition below is both necessary and sufficient.

$$\frac{\partial}{\partial \widetilde{a}} \left(1 - h(\widetilde{\lambda}) h(\lambda)^{\widetilde{a}-1}\right) = f + c_a \tag{45}$$

The above equation implicitly determines the optimal $\widetilde{a}$ as a function of $\widetilde{\lambda}$, i.e., $\widetilde{a}^*(\widetilde{\lambda}, 1)$.

Next, consider the problem of a deviant firm. By equation (43) its profit is given by

$$\widetilde{\pi} = \widetilde{\lambda} \left[1 - h(\widetilde{\lambda}) h(\lambda)^{\widetilde{a}-1} - (\widetilde{a} - 1) f - \widetilde{a} c_a - U\right]$$

In the above equation, $\widetilde{a}$ is a function of $\widetilde{\lambda}$, $\widetilde{a}^*(\widetilde{\lambda}, 1)$. The first-order condition of the deviant firm with respect to $\widetilde{\lambda}$ is

$$0 = \frac{\partial \widetilde{\pi}}{\partial \widetilde{\lambda}} + \frac{\partial \widetilde{\pi}}{\partial \widetilde{a}} \frac{\partial \widetilde{a}^*(\widetilde{\lambda}, 1)}{\partial \widetilde{\lambda}} = \left[1 - h(\widetilde{\lambda}) h(\lambda)^{\widetilde{a}-1} - (\widetilde{a} - 1) f - \widetilde{a} c_a - U\right] - \widetilde{\lambda} h'(\widetilde{\lambda}) h(\lambda)^{\widetilde{a}-1}$$

where we used the fact that $0 = \frac{\partial \widetilde{\pi}}{\partial \widetilde{a}}$ (see equation (45)). Evaluating the above first-order condition at $\widetilde{\lambda} = \lambda$ and $\widetilde{a} = a$ together with equation (34) implies that

$$f = \lambda h'(\lambda) h(\lambda)^{a-1}$$

which is exactly equation (25) except that in the above equation a is endogenous. Combining the above equation with (44) yields the equilibrium number of applications:

$$-h(\lambda)^a \log h(\lambda) = \lambda h'(\lambda) h(\lambda)^{a-1} + c_a.$$

This is exactly the planner's first-order condition for the socially optimal a. Therefore, the equilibrium number of applications is also constrained efficient. □

A.8. Galenianos and Kircher (2009) with general meeting technologies

In this Appendix, we extend Galenianos and Kircher (2009) with $a = 2$ to general meeting technologies. The environment is the same as our model in Section 2 except 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*

$$h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) = m(\lambda_H) - \lambda_H m'(\lambda_L) \tag{46}$$

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

$$\frac{1}{\lambda_H} + \frac{1}{\lambda_L} = \frac{2}{\lambda}, \tag{47}$$

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

$$U_{GK} = \left(2 - \frac{m(\lambda_H)}{\lambda_H}\right) m'(\lambda_L) \tag{48}$$

$$\pi_{GK} = h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) \tag{49}$$

Proof. This 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 of 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

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

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, 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 the right-hand side of equation (46). The left-hand side of equation (46) 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 (46) as $\mathcal{D}(\lambda_L, \lambda_H)$. That is, for $\lambda_H \ge \lambda_L$, 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 $\lambda_H \ge \lambda_L$), so the above equation admits at least one solution. Furthermore, $\frac{\partial}{\partial \lambda_L}\mathcal{D}(\lambda_L, \lambda_H) = -m''(\lambda_L)(h(\lambda_H)\lambda_L - \lambda_H) < 0$ since $h(\lambda_H)\lambda_L < \lambda_L < \lambda_H$. Similarly, $\frac{\partial}{\partial \lambda_H}\mathcal{D}(\lambda_L, \lambda_H) = h'(\lambda_H)(m(\lambda_L) - \lambda_L m'(\lambda_L)) + (m'(\lambda_L) - m'(\lambda_H)) > 0$. Therefore, along the curve $\mathcal{D}(\lambda_L, \lambda_H) = 0$ we have $\frac{d\lambda_H}{d\lambda_L} > 0$. Since the constraint $1/\lambda_L + 1/\lambda_H = v/u$ 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 (46) simplifies to $\lambda_H = 2\lambda_L$. By equation (47), we then have

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

By equation (48), the expected payoff of workers is given by $U_{GK} = 32(1+3\lambda)/(2+3\lambda)(4+3\lambda)^2$. Similarly, by equation (49), 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: $M_{GK}(u,v) = u(1 - h(\lambda_H) + h(\lambda_H)(1 - h(\lambda_L))) = u((8+18\lambda)/(2+3\lambda)(4+3\lambda))$. The workers' marginal contribution to surplus is $\partial M_{GK}(u,v)/\partial u = 32/(4+3\lambda)^2 - 4/(2+3\lambda)^2$. Similarly, one can calculate the workers' marginal contribution to surplus when the surplus function is $M(u,v)$ in equation (26), 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.9. 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,

$$f_L = -m'(\lambda_L)h(\lambda_H) \tag{50}$$

$$f_H = \frac{1}{\lambda_H}(m(\lambda_H) - \lambda_H m'(\lambda_H)). \tag{51}$$

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

$$h(\lambda_H)\left(m(\lambda_L) - \lambda_L m'(\lambda_L)\right) = m(\lambda_H) - \lambda_H m'(\lambda_H). \tag{52}$$

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 (47), we also have the firm adding-up constraint: $1/\lambda_H + 1/\lambda_L = 2/\lambda$. Note that equation (52) is different from equation (46), which implies that even though 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 (52), 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 there must be longer expected queues at the $w = 0$ firms, so a majority of the firms offer $w = 1$.

Claim 1: A worker's 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

$$U = (1 - h(\lambda_H)) - f_H - f_L. \tag{53}$$

With probability $1 - h(\lambda_H)$, a worker matches with a firm that offers $w = 1$, and with probability $h(\lambda_H)(1 - h(\lambda_L))$, a worker matches 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 $m'(\lambda) \le (\frac{m(\lambda)}{\lambda})^2 = (1 - h(\lambda))^2$ 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 (52) 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 length $\widetilde{\lambda}$.

Now suppose that workers who apply to the deviant firm send the other applications to firms with $w = 0$. Then the deviant firm's expected profit is

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

Here we assume that even if the deviant firm posts $\widetilde{w} = 0$,[18] 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 substitute 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 (53). By equation (51), 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 their other applications to firms with $w = 1$. The deviant firm's expected profit is

$$\widetilde{\pi} = h(\lambda_H)m(\widetilde{\lambda})(1 - \widetilde{w}) + \widetilde{\lambda}\widetilde{f}. \tag{54}$$

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

$$U = (1 - h(\lambda_H)) + h(\lambda_H)(1 - h(\widetilde{\lambda}))\widetilde{w} - f_H - \widetilde{f} \tag{55}$$

When the deviant firm posts $\widetilde{w} < 1$, equation (54) states that the deviant firm hires if and only if its chosen candidate fails to get an offer from the other firm. Equation (54) also holds when $\widetilde{w} = 1$ because the deviant firm's payoff comes only from application fees. Equation (55) also continues to hold. From equation (55), we can solve for $\widetilde{f}$ and then substitute it into the deviant firm's expected profit, which gives

[18] Posting $\widetilde{w} = 0$ is a deviation if $\widetilde{f} \neq f_L$.

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

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

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

where in the second equality we used equation (53). By equation (50), the optimal $\widetilde{\lambda}$ is simply λ_L. Hence in this case the deviant cannot do better than posting $w = 0$ and f_L. 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 (52) and the firm adding-up constraint we can solve for the equilibrium λ_L and λ_H, which are given below:

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

We can then substitute the above equation into equations (52) and (53) to get the equilibrium expected payoff of firms, π^*_{GK}, and the expected payoff of workers, U^*_{GK}. These are complicated objects, but since all are functions of λ only, one can 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.8. □

A.10. Collective firm deviations

Consider the benchmark model where the number of applications, a, is fixed. So far we have analyzed equilibrium using the standard directed-search type of equilibrium analysis, i.e., considering a deviation by a single firm. In this Appendix, we allow for deviations by an arbitrarily small measure of firms, where by "arbitrarily small" we mean that the mechanisms these firms post do not affect market utility.[19] The key difference between allowing for single-firm deviations versus collective deviations is that with a single-firm deviation, workers who apply to the deviant firm must send their other $a-1$ applications to non-deviant firms. With collective deviations, workers who apply to a deviant firm can choose how to allocate their remaining $a-1$ applications between deviant and non-deviant firms. In directed search models with $a = 1$, the two approaches give the same answer. Interestingly, with multiple applications, they don't. With single-firm deviations, as we have shown in the main text, there exists a continuum of symmetric pure-strategy equilibria; with collective deviations, the unique symmetric pure-strategy equilibrium is the efficient one with $w_1 = 1$. The fundamental reason for this difference, as we explain below, is that if workers make multiple applications, the expected payoff associated with any one of a worker's applications depends on the outcome of that worker's other applications.

With collective deviations, we now prove that the unique symmetric pure-strategy equilibrium is the efficient one with $w_1 = 1$. The strategy of our proof is as follows. Consider a candidate equilibrium (f, w_1). For this to be an equilibrium, there must be no profitable (collective) deviation. We start by considering "small" deviations from (f, w_1) and show that the only candidate equilibrium that survives all possible small deviations is the one with $w_1 = 1$. Then, having isolated a unique candidate (pure-strategy, symmetric) equilibrium, we verify that this candidate equilibrium also survives "large" deviations. The details are as follows.

[19] To be precise, we are assuming that both the single firm and the firms that deviate collectively are of measure zero.

Consider a candidate equilibrium (f, w_1). Suppose an arbitrarily small measure of firms deviates by posting $(\widetilde{f}, \widetilde{w}_1)$ and that the expected queue that results at a deviating firm is $\widetilde{\lambda}$. How many applications can a deviant firm expect to receive? This is determined by a worker indifference condition. Let U_k be the expected payoff of workers who send k applications to deviant firms and $a-k$ to non-deviant firms. The market utility of workers is simply U_0, which is given by equation (10),

$$U_0 = \left(1 - h(\lambda)^a\right) - a(1-h(\lambda))h(\lambda)^{a-1}(1-w_1) - af.$$

More generally, U_k is given by

$$\begin{aligned} U_k = &\left(1 - h(\widetilde{\lambda})^k h(\lambda)^{a-k}\right) - (a-k)h(\widetilde{\lambda})^k h(\lambda)^{a-k-1}(1-h(\lambda))(1-w_1) \\ &- k(1-h(\widetilde{\lambda}))h(\widetilde{\lambda})^{k-1}h(\lambda)^{a-k}(1-\widetilde{w}_1) - k\widetilde{f} - (a-k)f. \end{aligned} \tag{56}$$

With probability $(a-k)h(\widetilde{\lambda})^k h(\lambda)^{a-k-1}(1-h(\lambda))$, the worker receives exactly one offer from a non-deviant firm and none from deviant firms and with probability $k(1-h(\widetilde{\lambda}))h(\widetilde{\lambda})^{k-1}h(\lambda)^{a-k}$, the worker receives exactly one offer from a deviant firm but none from non-deviant firms.

We start by considering the special case in which the deviation is small, i.e., $\widetilde{w}_1$ and $\widetilde{f}$ are close to w_1 and f respectively. We can then use a Taylor series expansion to solve the worker problem: workers who decide to visit the deviant firms need to choose the optimal k to maximize their expected payoff U_k given by equation (56). To simplify the exposition below, we assume that the deviant firms choose $\widetilde{w}_1$ and $\widetilde{\lambda}$ instead of $\widetilde{w}_1$ and $\widetilde{f}$, which is without loss of generality because given $\widetilde{w}_1$ and $\widetilde{\lambda}$, $\widetilde{f}$ follows from the worker indifference condition. The solution of the worker problem is given by the following lemma.

Lemma 5. *Suppose the firm deviation is small, i.e., $\widetilde{\lambda}$ and $\widetilde{w}_1$ are close to λ and w_1 respectively. Set $\Delta\lambda = \widetilde{\lambda} - \lambda$ and $\Delta w_1 = \widetilde{w}_1 - w_1$. If*

$$\frac{h'(\lambda)}{2(1-h(\lambda))h(\lambda)}\left((1-a(1-w_1))h(\lambda) + (a-2)(1-w_1)\right) > \frac{\Delta w}{\Delta\lambda} \tag{57}$$

then for workers who decide to visit deviant firms, the optimal $k = 1$, i.e., $U_0 = U_1 > U_j$ for all $j \geq 2$. If the reverse inequality, $<$, holds, then the optimal $k = a$, i.e., $U_0 = U_a > U_j$ for all $j = 1, \ldots, a-1$.

Proof. Consider the following difference for $j, k \geq 1$:

$$U_k - \left(\frac{j-k}{j}U_0 + \frac{k}{j}U_j\right).$$

Substituting equation (56) into the above expression shows that it is strictly positive if and only if

$$\begin{aligned} &1 - h(\widetilde{\lambda})^k h(\lambda)^{a-k} - (a-k)h(\widetilde{\lambda})^k h(\lambda)^{a-k-1}(1-h(\lambda))(1-w_1) \\ &- k(1-h(\widetilde{\lambda}))h(\widetilde{\lambda})^{k-1}h(\lambda)^{a-k}(1-\widetilde{w}_1) \\ &> \frac{j-k}{j}\left[1 - h(\lambda)^a - ah(\lambda)^{a-1}(1-h(\lambda))(1-w_1)\right] + \frac{k}{j}\Big[1 - h(\widetilde{\lambda})^j h(\lambda)^{a-j} \\ &- (a-j)h(\widetilde{\lambda})^j h(\lambda)^{a-j-1}(1-h(\lambda))(1-w_1) - j(1-h(\widetilde{\lambda}))h(\widetilde{\lambda})^{j-1}h(\lambda)^{a-j}(1-\widetilde{w}_1)\Big] \end{aligned} \tag{58}$$

The above inequality is complicated, but it does not involve f and $\widetilde{f}$. To simplify it, we treat (58) as a function of $\widetilde{w}_1$ and $\widetilde{\lambda}$ and evaluate it using a second-order Taylor expansion around $\widetilde{w}_1 = w_1$ and $\widetilde{\lambda} = \lambda$, which yields,[20]

$$(j-k)\Delta\lambda^2 \left(\frac{h'(\lambda)}{2(1-h(\lambda))h(\lambda)} ((1-a(1-w_1))h(\lambda) + (a-2)(1-w_1)) - \frac{\Delta w}{\Delta \lambda} \right) > 0 \tag{59}$$

Note that the term in the large parentheses is independent of j and k, and is equivalent to (57). Suppose that it is strictly positive, i.e., (57) holds, then it must be that $j > k$ for all $j \geq 2$ and therefore the optimal $k = 1$. Similarly, if (57) holds with $<$, then the optimal $k = a$. □

Suppose the deviant firms choose Δw_1 and $\Delta\lambda$ such that the optimal $k = 1$. Then we are back in the case considered in the text: the deviant firm's expected payoff depends only on $\widetilde{\lambda}$ (see equations (14) and (15)), and the deviant firm's first-order condition implies that for w_1 and f to be part of an equilibrium, a necessary condition is that equation (19) must hold. That is, if we were to only consider small deviations that lead workers to choose $k = 1$, then all of the equilibria identified in Proposition 1 would be potential equilibria with collective deviations.

However, we also need to consider the case in which the reverse inequality holds in (57) and the optimal $k = a$. In this case, the worker indifference condition is $U_0 = U_a$; i.e.,

$$U = U_0 = 1 - h(\widetilde{\lambda})^a - a(1-h(\widetilde{\lambda}))h(\widetilde{\lambda})^{a-1}(1-\widetilde{w}_1) - a\widetilde{f}$$

and the expected payoff of a deviant firm is

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

A deviant firm receives at least one application with probability $m(\widetilde{\lambda})$ and with probability $h(\widetilde{\lambda})^{a-1}$ its chosen candidate has no offers from the other $a-1$ applications, which were also submitted to deviant firms. Combining the above two equations yields

$$\widetilde{\pi} = \frac{1}{a}\widetilde{\lambda}\left(1 - h(\widetilde{\lambda})^a - U\right).$$

This equation has a clear interpretation. Each worker sends out a applications, which generates an expected output of $1 - h(\widetilde{\lambda})^a$, and receives an expected payoff of U. Since the worker applies to a firms, each firm can only expect a share $1/a$ of the expected output net of the worker payoff. The above equation shows that the expected profit of deviant firms only depends on $\widetilde{\lambda}$, a feature also present when workers send exactly one application to deviant firms (see equations (13) and (15)). For the deviant firms, the first-order condition with respect to $\widetilde{\lambda}$ is

$$U = 1 - h(\widetilde{\lambda})^a - a\widetilde{\lambda}h(\widetilde{\lambda})^{a-1}h'(\widetilde{\lambda})$$

For there to be no profitable deviations, the above equation must hold at $\widetilde{\lambda} = \lambda$, in which case the right-hand side is exactly the worker's marginal contribution to surplus. We know from Proposition 1 that in the candidate equilibria indexed by w_1, the worker's equilibrium payoff is always strictly smaller than the marginal contribution to surplus unless $w_1 = 1$. Thus only the efficient

[20] To derive inequality (58), we replace $\widetilde{w}_1$ with $w_1 + \Delta w$ and $h(\widetilde{\lambda})^m$ with its second-order Taylor series expansion: $h(\lambda)^m + mh(\lambda)^{m-1}h'(\lambda)\Delta\lambda + (m/2)h(\lambda)^{m-2}((m-1)h'(\lambda)^2 + h(\lambda)h''(\lambda))\Delta\lambda^2$.

equilibrium in which $w_1 = 1$ can survive small deviations where the reverse inequality holds in (57).

Thus we are left with one candidate equilibrium, the one with $w_1 = 1$ and f given by equation (25). We next verify that this candidate equilibrium can survive any deviation, not just small ones with $\widetilde{\lambda}$ and $\widetilde{w}_1$ close to λ and w_1, respectively. Thus, consider an arbitrary (possibly large) deviation from the candidate equilibrium with $w_1 = 1$. Suppose workers who send applications to deviant firms find it optimal to send k applications to the deviant firms and $a - k$ applications to the non-deviant firms. Then, the worker indifference condition is

$$\begin{aligned}1 - h(\lambda)^a - af = U = 1 - h(\widetilde{\lambda})^k h(\lambda)^{a-k} - k(1 - h(\widetilde{\lambda}))h(\widetilde{\lambda})^{k-1}h(\lambda)^{a-k}\,(1 - \widetilde{w}_1)\\ - k\widetilde{f} - (a-k)f,\end{aligned}$$

and the expected payoff of the deviant firms is

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

Combining the above two equations yields

$$\widetilde{\pi} = \frac{1}{k}\widetilde{\lambda}\left(1 - h(\widetilde{\lambda})^k h(\lambda)^{a-k} - (a-k)f - U\right)$$

This equation is concave in $\widetilde{\lambda}$, so the first-order condition is both necessary and sufficient for the maximum, and it is given by

$$0 = 1 - h(\widetilde{\lambda})^k h(\lambda)^{a-k} - (a-k)f - U - \widetilde{\lambda}kh(\widetilde{\lambda})^{k-1}h'(\widetilde{\lambda})h(\lambda)^{a-k}$$

Since $U = 1 - h(\lambda)^a - af$, the above equation holds at $\widetilde{\lambda} = \lambda$, which implies that there are no profitable deviations. We have thus shown that the only equilibrium is the constrained efficient one where $w_1 = 1$. The following proposition summarizes this.

Proposition 8. *Suppose that an arbitrarily small measure of firms can deviate collectively and that the total number of applications per worker is exogenous. Then, there exists a unique symmetric, pure-strategy equilibrium in which all firms post $w_1 = 1$ and f is given by equation* (25).

The intuition is that at any candidate equilibrium with $w_1 < 1$, an arbitrarily small measure of firms could offer a slightly higher w_1 and they would expect an increase in both their queue and payoff. This logic does not work if a single firm deviates because the value of what a firm offers to a worker depends on what other firms offer. If no other firm offers a high w_1, most of the value that the firm offers comes from the chance of Bertrand competition in case of multiple offers. In that case, workers are more keen on visiting the non-deviant firms which have shorter queues.

References

Albrecht, J.W., Decreuse, B., Vroman, S.B., 2019. Directed search and phantom vacancies. Mimeo.

Albrecht, J.W., Gautier, P.A., Vroman, S.B., 2006. Equilibrium directed search with multiple applications. Rev. Econ. Stud. 73, 869–891.

Albrecht, J.W., Gautier, P.A., Vroman, S.B., 2014. Efficient entry with competing auctions. Am. Econ. Rev. 104 (10), 3288–3296.

Albrecht, J.W., Jovanovic, B., 1986. The efficiency of search under competition and monopsony. J. Polit. Econ. 94 (6), 1246–1257.

Armstrong, M., Vickers, J., 2001. Competitive price discrimination. Rand J. Econ. 32 (4), 579–605.

Armstrong, M., Vickers, J., 2010. Competitive non-linear pricing and bundling. Rev. Econ. Stud. 77 (1), 30–60.

Burdett, K., Judd, K.L., 1983. Equilibrium price dispersion. Econometrica 51, 955–969.
Burdett, K., Shi, S., Wright, R., 2001. Pricing and matching with frictions. J. Polit. Econ. 109, 1060–1085.
Cai, X., Gautier, P., Wolthoff, R., 2017. Search frictions, competing mechanisms and optimal market segmentation. J. Econ. Theory 169, 453–473.
Cai, X., Gautier, P., Wolthoff, R., 2018. Meetings and mechanisms. Mimeo.
Chade, H., Smith, L., 2006. Simultaneous search. Econometrica 74 (5), 1293–1307.
Cheron, A., Decreuse, B., 2017. Matching with phantoms. Rev. Econ. Stud. 84 (3), 1041–1070.
Coles, M.G., Eeckhout, J., 2003. Indeterminacy and directed search. J. Econ. Theory 111, 265–276.
Eeckhout, J., Kircher, P., 2010a. Sorting and decentralized price competition. Econometrica 78, 539–574.
Eeckhout, J., Kircher, P., 2010b. Sorting vs screening - search frictions and competing mechanisms. J. Econ. Theory 145, 1354–1385.
Galenianos, M., Kircher, P., 2009. Directed search with multiple job applications. J. Econ. Theory 114, 445–471.
Gautier, P.A., Holzner, C.L., 2017. Simultaneous search and efficiency of entry and search intensity. Am. Econ. J. Microecon. 9 (3), 245–282.
Hosios, A.J., 1990. On the efficiency of matching and related models of search and unemployment. Rev. Econ. Stud. 57, 279–298.
Jacquet, N.L., Tan, S., 2012. Wage-vacancy contracts and coordination frictions. J. Econ. Theory 147 (3), 1064–1104.
Julien, B., Roger, G., 2019. Incentive contracts and efficiency in a frictional market. Mimeo.
Kircher, P., 2009. Efficiency of simultaneous search. J. Polit. Econ. 117, 861–913.
Lester, B., Visschers, L., Wolthoff, R., 2015. Meeting technologies and optimal trading mechanisms in competitive search markets. J. Econ. Theory 155, 1–15.
Moen, E.R., 1997. Competitive search equilibrium. J. Polit. Econ. 105, 385–411.
Peters, M., 1991. Ex ante price offers in matching games: non-steady states. Econometrica 59 (5), 1425–1454.
Peters, M., 1997. A competitive distribution of auctions. Rev. Econ. Stud. 64 (1), 97–123.
Peters, M., 2000. Limits of exact equilibria for capacity constrained sellers with costly search. J. Econ. Theory 95, 139–168.
Peters, M., 2013. Competing mechanisms. In: Vulkan, N., Roth, A.E., Neeman, Z. (Eds.), The Handbook of Market Design. Oxford University Press, pp. 609–628.
Pissarides, C.A., 2000. Equilibrium Unemployment Theory. Oxford University Press.
Rochet, J.-C., Stole, L.A., 2002. Nonlinear pricing with random participation. Rev. Econ. Stud. 69 (1), 277–311.
Shimer, R., 2005. The assignment of workers to jobs in an economy with coordination frictions. J. Polit. Econ. 113 (5), 996–1025.
Thanassoulis, J., 2007. Competitive mixed bundling and consumer surplus. J. Econ. Manag. Strategy 16 (2), 437–467.
Yin, X., 2004. Two-part tariff competition in duopoly. Int. J. Ind. Organ. 22 (6), 799–820.