Efficiency of Simultaneous Search*

Philipp Kircher†

January 27, 2007

Abstract

We analyze a directed search labor market. Firms compete for labor through public wage postings. Then, workers choose their search intensity by deciding how often and where to apply for a job. In the last stage of the interaction, firms and applicants are matched according to a stable assignment. This can be interpreted as sequential offers by firms to their applicants. In contrast to models where only a single offer is made by each firm, the equilibrium is constrained efficient along the three operative margins: entry of firms, number of matches and number of applications. Surprisingly, wage dispersion is necessary for the market to achieve constrained efficiency despite homogeneity of workers and firms. For vanishing application costs the equilibrium outcome converges to the unconstrained efficient competitive outcome.

*This paper circulated under the title "Efficiency of of Simultaneous Directed Search with Recall". The idea for this paper grew out of earlier fruitful collaboration with Manolis Galenianos. I also thank Ken Burdett, Jan Eeckhout, Stephan Lauermann, Wolfram Merzyn, Georg Nöldeke, Andrew Postlewaite and Randy Wright for many helpful discussions. This paper benefitted from the input of seminar participants at the University of Bonn, University of Michigan, University of Pennsylvania, the NBER and the Conference on Matching and Two-Sided Markets.

†University of Pennsylvania. e-mail: kircher@econ.upenn.edu.

1 Introduction

The efficiency of labor markets rests on the presence of competitive forces. The classical model predicts efficient outcomes, but is at odds with important labor market phenomena such as the simultaneous occurrence of open vacancies and unemployment. Some form of frictions seems necessary to explain this. The idea that some notion of efficiency will prevail even in the presence of frictions has been put forth at least since the discussion of the natural rate of unemployment by Friedman (1968) and Phelps (1967).

Recent models of directed search have tried to formalize this by allowing firms to publicly post wage announcements in order to attract workers. Frictions are captured through a restriction on workers' strategies that leads to randomization in the way they apply for jobs and, thus, to miscoordination: Some jobs have more than one applicant and some applications remain unemployed, while other jobs might not receive any applicant and remain vacant. Nevertheless, a higher wage directs more applicants to apply there, and the probability of remaining vacant for the firm decreases. For this class of models it has been shown in a series of contributions that the competitive forces of the public wage announcement lead to a constrained efficient allocation, as shown in Hosios (1990), Montgomery (1991), Moen (1997), Shimer (1996, 2005), Shi (2001, 2002) and Mortensen and Wright (2002).[1]

Models in this spirit typically incorporate a strong restriction on the number of firms a worker can apply to: a worker is assumed to apply only for a single job per period, which is obviously quite unrealistic. Two recent exceptions are Albrecht, Gautier and Vroman (2006) and Galenianos and Kircher (2006) who formulate models in which workers send $N > 1$ applications. They show that equilibrium is inefficient, implying a role for labor market interventions.[2] However, these papers, while relaxing the restriction to $N = 1$ applications, impose another strong assumption. They restrict firms to only contact one of their applicants and to remain vacant if this applicant decides to work for another firm, even when many other workers have applied for the job. This raises the question about the source of the inefficiency. Do multiple applications constitute a new source of inefficiency that directed search cannot internalize? Or does the failure to achieve constrained efficiency stem from details of the process by which firms

[1] Earlier models in which workers meet firms randomly rejected the hypothesis that frictional markets yield efficiency (Diamond 1982; Mortensen 1982a, 1982b; and Pissarides 1984, 1985). The inability of the wage to direct the allocation of labor seems to be the important driver of these results. Hosios (1990) characterizes generic inefficiencies of entry in random search models with bargaining, and his extension is the first to show that constrained efficiency is recovered in a directed search setting.

[2] A minimum wage improves efficiency in Albrecht, Gautier and Vroman (2006). In Galenianos and Kircher (2006) wage dispersion is associated with inefficiencies that could be alleviated by a uniform government-mandated wage.

contact their applicants?

This paper relaxes the assumption that a firm only communicates with one of its applicants. Rather, we assume that communication with all applicants results in a matching that is stable on the network that arises in the application process. That is, we assume that no firm remains vacant while one of its applicants is matched at a lower wage. If this would happen, both the firm and the worker would be better off by deviating and forming a match amongst themselves. This stability notion has a natural interpretation in the labor market: Firms make offers sequentially to their applicants, and workers always accept the best offer they have obtained up to that point. If a worker rejects a firm's offer, the firm "recalls" whether it has other applicants and contacts the next one. Stability arises because applicants are free to change their mind if a better offer comes along. Therefore, a vacant firm with an applicant whose best offer is at a lower wage can make him an offer and ceases to be vacant. This offer-process is assumed to happen in a short period of time prior to the start of the actual employment relationship. The stability notion and the interpretation as a sequential (deferred acceptance) process, developed by Gale and Shapley (1962), has been successfully applied in many areas; the novelty of our approach is that workers first have to apply to the firms (i.e. make themselves known) in order to be available for a match.[3,4]

We prove that, once the stringent assumption in the prior literature is relaxed in this way, efficiency is recovered. In particular, equilibrium is constrained efficient along the three operative margins. *Application efficiency*: for a given number of applications and a given number of firms the number of matches is constrained optimal. *Entry efficiency*: the constrained optimal number of firms enter. *Efficiency of Search Intensity*: the number of applications that workers send is constrained efficient. Thus, even in the presence of strategic search intensity does the market achieve a constrained efficient allocation if firms can communicate with all their workers. Firms' wage announcements allow them to price applications in a way that internalizes all externalities.

The analysis of the efficiency properties of our model reveals two additional insights: Firstly, we identify an efficiency role of wage dispersion despite the fact that firms and workers are homogeneous. Wage dispersion is not just a sign of frictions; it is the optimal response to those

[3] The sequential process described here leads to stability in a finite economy. For our continuum economy we view the sequential process as an interpretation. (Kircher (2006) shows some convergence properties for such a process in a large economy.) Our assumption that the wages are fixed when determining stability has some precedence in, e.g., Bulow and Levin (2006). However, they assume that all agents know each other, while our focus is on the application process by which workers make themselves known.

[4] Our assumption that workers observe the wages and have to apply for jobs prior to a stable matching fits some of the features of the market for medical interns in the United States, in which hospitals post job descriptions and meet with applicants prior to a centralized matching procedure. See Section 6 for details.

frictions. It allows the market to achieve different hiring probabilities at different firms, which is important for search efficiency because it reduces instances in which one worker could obtain two jobs and another none.[5] Secondly, the paper also adds to the literature on asymptotic efficiency of search markets. We show that for vanishing application costs the equilibrium converges to the unconstrained efficient outcome of a frictionless Walrasian economy. Such convergence results are novel for environments with simultaneous (job) search.[6]

To our knowledge this is the first attempt to integrate the two-sided strategic considerations of a search environment with stability concepts used in matching markets.[7] The paper draws on three strands of literature. We use insights from the directed search literature (e.g. Burdett, Shi and Wright (2001)) to model the frictions and information flows in the market. With multiple applications workers face a simultaneous portfolio choice. For this type of problem Chade and Smith (2006) consider an individual agent's choice and Galenianos and Kircher (2006) derive implications in an equilibrium framework. For the final matching we apply stability concepts pioneered by Gale and Shapley (1962) to the network that formed in the search process.

In the following, we first present the model. Section 3 characterizes the equilibrium. Section 4 analyzes efficiency and investigates the role of wage dispersion in achieving it. Notation and exposition remain much more tractable when we consider at most two applications per worker, therefore Sections 2 to 4 are restricted to this case. Section 5 lifts this restriction and, additionally, discusses convergence for vanishing application costs. Section 6 discusses and justifies our main modeling assumptions. It also expands on related literature. In particular, it highlights the cause of failure of efficiency in models without recall. Omitted proofs are gathered in the appendix.

[5]For homogenous workers and firms such an efficiency role is novel. Models that entail dispersion but do not entail such an efficiency role include Acemoğlu and Shimer (2000); Albrecht, Gautier and Vroman (2006); Burdett and Judd (1983); Burdett and Mortensen (1998); Butters (1977); Delacroix and Shi (2005); Galenianos and Kircher (2006); Gautier and Moraga-González (2005).

[6]Asymptotic efficiency has been established in sequential search e.g. in Gale (1987). For an overview and quite general specifications see Lauermann (2006). In simultaneous search, Acemoğlu and Shimer (2000) and Albrecht, Gautier and Vroman (2006) present limit results, yet converge to some constrained efficient outcome where frictions are still present on the other market side.

[7]Gautier and Moraga-González (2005) present a three player example with a similar concept in an environment where wages are unobservable. Their main analysis of a large market assumes no recall, and also exhibits inefficiencies.

2 The Model

2.1 Environment and Strategies

There is a measure 1 of workers and a large measure V of potential firms. The measure v of active firms is determined by free entry. Each active firm is capacity-constrained and can only employ a single worker, each worker can only work for a single firm. A vacant firm has a productivity normalized to zero, a firm that employs a worker has a productivity normalized to one but has to pay the wage bill. All agents are risk neutral. Firms maximize expected profits. Workers maximize expected wage payments.

The market interaction has three stages. First, potential entrants can become active by paying a setup cost $K < 1$, and active firms publicly post a wage commitment. Next, workers observe all posted wages. Each worker decides on the number $i \in \{0, 1, 2\}$ of applications he wants to send at cost $c(i)$, where $c(0) = 0$ and the marginal costs $c_i = c(i) - c(i-1)$ are assumed to be weakly increasing. The worker also decides on the i active firms to which he applies. The applications form the links in a network between workers and firms. In the final stage workers and firms are matched, the posted wage is paid and matched pairs start production.

We assume a stable matching on the network, i.e. matches form such that no firm remains vacant but has an applicant that starts to work for some other firm at a lower wage. This assumption is based on the idea that otherwise both could deviate and be jointly better off. Our matching specification in the next subsection embeds this concept. We will sometimes interpret this as an outcome of a sequential process in which firms make offers and workers accept or reject, but workers can change their mind if they get a better offer, as outlined in the introduction. Since rejected firms recall that they had other applicants and offer the job to them, we use the term "recall" to refer to such a sequential interpretation. Another interpretation is that offers are made top-down, with the higher wage firms moving first to make offers and the lower wage firms waiting to see which of their applicants have not been hired at higher wages before making offers themselves.

The notion of a large, anonymous market is captured by the assumption that agents' equilibrium strategies are symmetric and anonymous. This standard assumption of the directed search literature implies that all firms use the same entry, posting and hiring strategy, and do not condition their strategy on the identity of the worker. All workers use the same application strategies, and do not condition on the firms' identities.[8] The symmetry of the workers' appli-

[8] These assumptions are discussed in Section 6. We should note that we could purify the firms' posting strategy, and anonymity of the workers' strategies merely simplifies the notation.

cation strategies usually requires mixed strategies which create the market frictions: sometimes multiple workers apply for the same job, and sometimes none apply at all.

A pure strategy for a firm is its entry decision $e \in \{Enter, Out\}$ and a wage offer $w \in [0,1]$. A mixed strategy for a firm is a probability ϕ of playing $Enter$ and a cumulative distribution function F on $[0,1]$. Throughout the paper we adopt the law of large numbers convention, which for instance implies that F is also the realized distribution of wage offers and $v = \phi V$ is the realized measure of active firms.[9] A worker observes the distribution of posted wages and decides on the number of applications $i \in \{0,1,2\}$ that he wants to send. If $i = 1$ he also decides on a wage $w \in [0,1]$ to which he applies; if $i = 2$ he decides on a wage tuple $(w_1, w_2) \in [0,1]^2$. This fully characterizes his strategy given the anonymity assumption, which implies that he randomizes equally over the firms that offer the same wage. A mixed strategy for a worker is a tuple $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$, where γ_i is the probability of sending i applications, and a tuple $\boldsymbol{G} = (G^1, G^2)$, where G^i is a cumulative distribution function over $[0,1]^i$ which describes the way the worker sends his i applications. If $i = 2$ we will throughout assume that $w_1 \leq w_2$, and will denote by $G^2_j(\tilde{w})$ the marginal distribution over w_j. By $G^2_j(\tilde{w}|w)$ we denote the conditional distribution over w_j, conditional on the other application w_i, $i \neq j$, being sent to wage w.

2.2 Expected Payoffs and Equilibrium Definition

To describe the expected payoffs under mixed strategies, let $\eta(w)$ denote the probability that a firm that posts wage w hires a worker. Let $p(w)$ be the probability that an application to wage w yields an offer sometime during the matching stage. These are endogenous objects, yet once they are defined, the profit of a firm posting wage w - omitting entry costs - is

$$\pi(w) = \eta(w)(1-w). \tag{1}$$

The profits comprise the margin $1-w$ if a worker is hired, multiplied by the probability $\eta(w)$ of hiring.

The utility of a worker who sends no applications is $U_0 = 0$. A worker who applies with one application to wage w obtains utility $U_1(w) = p(w)w - c(1)$, i.e. the expected profit minus the

[9] It will be convenient to assume that each combination in the support of the randomization of workers and firms is chosen by a continuum of agents in order to apply the law of large numbers convention per wage level. We assume the set of agents to be sufficiently large.

cost of the application. A worker who applies to wages (w_1, w_2) with $w_1 \leq w_2$ obtains utility

$$U_2(w_1, w_2) = p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1 - c(2). \tag{2}$$

The worker's utility is given by the wage w_2 if he is made an offer at that wage, which happens with probability $p(w_2)$. With the complementary probability $1 - p(w_2)$ he does not receive an offer at the high wage and his utility is w_1 if he gets an offer for his low wage application, which happens with probability $p(w_1)$. He always incurs the cost for the two applications.

When agents randomize over wages, their payoff is determined by appropriately averaging the payoffs at individual wages.

We will now relate $\eta(\cdot)$ and $p(\cdot)$ to the strategies (ϕ, F) and $(\boldsymbol{\gamma}, \boldsymbol{G})$. We will first determine the payoffs under the assumption that all firms and workers follow this strategy profile. We will then consider individual deviations from this profile. In the following we will talk about the offer set $\mathcal{V}$, which refers to the support of the wage offer distribution F, and about the application set $\mathcal{W}$, which refers to the support of $G^1(\cdot)$ (if $\gamma_1 > 0$) joint with the union of the support of $G_1^2(\cdot)$ and $G_2^2(\cdot)$ (if $\gamma_2 > 0$).

We first consider wages that are in the offer and in the application set. Let $\lambda(w)$ denote the ratio of applications to firms at wage w. We refer to $\lambda(w)$ as the gross queue length, because it describes the average queue of applications per job at wage w. It is characterized by the following mass balance:[10]

$$\gamma_1 G^1(w) + \gamma_2 G_1^2(w) + \gamma_2 G_2^2(w) = v \int_0^w \lambda(\tilde{w})\, dF(\tilde{w}) \quad \forall w \in [0, 1]. \tag{3}$$

The left hand side denotes the expected mass of applications that are sent to wages up to w. It is given by the probability that workers who send one application send it below w, and the probability that workers who send two applications send either their low or their high application below w. It is the inflow of applications to wages up to w. These are dispersed over the firms that offer wages up to wage w. This outflow is specified on the right hand side. It is given by the ratio of applications per firm multiplied with the number of firms, aggregated over all relevant wages.

The crucial observation is that not all applications are "effective" in the sense that the firm can hire the applicant. The applicant cannot be hired if he receives a strictly better offer, or

[10] We define $\lambda(\cdot)$ on $\mathbb{R}_+ \cup \{\infty\}$ to account for the case that a negligible fraction of firms might (non-optimally) receive a mass of applications.

has already received a weakly better offer. Denote by $\psi(w)$ the fraction of applications that are unavailable for hiring because the applicant gets matched to a higher wage firm, which we will define more precisely below. Then the ratio of effective applications per firm is given by

$$\mu(w) = (1 - \psi(w))\lambda(w). \tag{4}$$

We call $\mu(w)$ the *effective* queue length at w.

The probability that a firm with wage w has at least one effective application is given by $1 - e^{-\mu(w)}$. This is due to the anonymity of the workers' strategy, which leads to random assignment of applications to firms at a given wage. In a finite economy this implies that the number of effective applications is binomially distributed; for a large economy this is approximated by the Poisson distribution under which the probability that a firm receives no effective application is $e^{-\mu(w)}$. If the firm receives at least one effective application, i.e. an application by somebody who has not gotten a better or equally good offer elsewhere, it is able to fill its vacancy by our assumption of stability. Therefore the hiring probability for a firm is

$$\eta(w) = 1 - e^{-\mu(w)}. \tag{5}$$

Now consider the probability of getting a job at wage w for a worker who wants to obtain a job at that wage, i.e. who is effective in the sense that he did not get a better job offer. The worker only cares about his rival applicants that are effective, and not about those applicants that get hired at better wages. Given that there are $1 - e^{-\mu(w)}$ matches per firm and $\mu(w)$ effective applications per firm, the probability of an effective application to yield a match is given by[11]

$$p(w) = \frac{1 - e^{-\mu(w)}}{\mu(w)}, \tag{6}$$

with the convention that $p(w) = 1$ if $\mu(w) = 0$, which describes the limit as the effective queue length converges to zero.

Finally, consider the probability $\psi(w)$ that an offer does not lead to a match because the sender receives and accepts a different offer. $\psi(\cdot)$ is trivially zero if workers send only one application, i.e. if $\gamma_2 = 0$. Otherwise, consider some application sent to wage w, and let $\hat{G}(\tilde{w}|w)$ denote the probability that the sender had a second application and sent it to a wage

[11]For a careful but intuitive derivation of (5) and (6) as the limit for a finite but large economy see Burdett, Shi and Wright (2001).

weakly lower then $\tilde{w}$.[12] Similarly, let $\hat{g}(w|w)$ denote the probability that the sender sent a second application to w, i.e. $\hat{g}(w|w) = \hat{G}(w|w) - \lim_{\tilde{w}\nearrow w} \hat{G}(\tilde{w}|w)$. Then $\psi(w)$ is given by

$$\psi(w) = \int_{\tilde{w}>w}^{1} p(\tilde{w}) d\hat{G}(\tilde{w}|w) + \frac{p(w)}{2}\hat{g}(w|w). \tag{7}$$

That is, $\psi(w)$ is the average probability that the applicant sent two applications and the other application was strictly higher and successful. If the other application was sent to the same wage, the probability that it was successful is $p(w)$, and in this case the unconditional probability for each firm to make the offer first is 1/2.[13] The system defined by (4), (6) and (7) is recursive: At the highest offered wage the probabilities $p(w)$, $\eta(w)$ and $\psi(w)$ can be determined, and are then used to evaluate the corresponding terms at lower wages. Since the specification allows firms to hire any worker that is not matched at higher or equally high wages, the final matching is stable in the sense that no firm is vacant while one of its applicants is matched at a strictly lower wage. Due to the indifference of the firm about which worker to hire, the stable matching is clearly not unique. Our specification is based on the standard anonymity assumption that firms hire one of their effective applicants at random.

To round off the specification, briefly consider the case of a wage in the offer set that is not in the application set. In this case nobody applies, and we specify $\lambda(w) = \mu(w) = 0$. We interpret the polar case of wages in the application but not in the offer set as free disposal of applications without the chance of receiving an offer. This covers all possibilities in the support of the workers and/or firms randomization, and their average ex-ante payoffs can be calculated.

Now consider deviations. For a firm, any wage in the offer set can be evaluated as described above as it arises as a possible realization of F. Yet for a deviation to wages $w \notin \mathcal{V}$ the firm has to form a belief about the workers reaction. This problem is common in the directed search literature, and the most common approach is to assume that the queue length at the deviant is exactly such that workers are indifferent between applying and not applying. If the effective queue length would be higher than the level at which workers are indifferent, workers should adjust their applications and apply there less. If it were lower, applying to this wage would yield a strict gain compared to applying elsewhere and workers should apply more. We will

[12] There are $\gamma_1 dG^1(w)$ single applications, $\gamma_2 dG_1^2(w)$ low applications and $\gamma_2 dG_2^2(w)$ high applications at w, adding to a total measure $T(w) = \gamma_1 dG^1(w) + \gamma_2 dG_1^2(w) + \gamma_2 dG_2^2(w)$. Then $\hat{G}(\tilde{w}|w) = \sum_{j=1}^{2}[\gamma_2 dG_j^2(w)/T(w)]G_{-j}^2(\tilde{w}|w)$, where $-j \in \{1,2\}/\{j\}$.

[13] One can think about workers that apply twice to the same wage as randomizing in advance about which offer they would prefer to accept in case they get both offers. In this case an applicant is not available to a firm if the other firm is preferred and makes this applicant an offer.

call the highest utility a worker can obtain at wages in the offer set as the Market Utility. Let $U_i^* = \sup_{\boldsymbol{w} \in \mathcal{V}^i} U_i(\boldsymbol{w})$ denote the highest utility a worker can get by sending i applications. Then we can define the Market Utility as $U^* = \max\{U_0, U_1^*, U_2^*\}$.

Now consider some wage $w \notin \mathcal{V}$ and assume the queue length at this wage were $\mu(w) \in [0, \infty]$, which defines $p(w)$ as in (6). This queue length determines the attractiveness of applying to this wage for a worker. That is, a worker who applies there with one application can at best get $U_1(w)$. A worker who sends two applications and applies there with his low application can at best get $\hat{U}_{1,l}(w) = \sup_{\tilde{w} \in \mathcal{V}} U(w, \tilde{w})$. Applying with the high application he can get $\hat{U}_{2,h}(w) = \sup_{\tilde{w} \in \mathcal{V}} U(\tilde{w}, w)$. Let $\hat{U}(w) = \max\{U_1(w), \hat{U}_{2,l}(w), \hat{U}_{2,h}(w)\}$. We will assume

Definition 1 (Market Utility Assumption)
For any $w \notin \mathcal{V}$, $\mu(w) > 0$ if and only if $U^* = \hat{U}(w)$.

It states that workers are indifferent between obtaining the Market Utility and applying to the not-offered wage (possibly combined with the most attractive offered wage).[14] Only if the wage is too low it is not possible to adjust the effective queue length to obtain indifference, in which case the effective queue length is set to zero because nobody would apply there. While the arguments presented here have only intuitively appealed to "reasonable" responses of workers to deviating wage offers, papers by Burdett, Shi and Wright (2001) and Peters (1997, 2000) rigorously establish equivalence of the Market Utility Assumption and the subgame perfect response in (a limit of) finite economies in which workers send one application.

Firms also have to evaluate the profit of entering if $\phi = 0$, i.e. no other firm enters. Assume some firm enters and offers a wage. In the same spirit of subgame perfection, workers either do not apply at all because the wage is too low, or they drive the queue length down to a level where they are indifferent between applying or not applying. Anticipating this response, the firm will enter if it can find a profitable wage offer. Formally, we assume

Definition 2 (Entry Assumption)
$\phi > 0$ if there exist $w \in [0, 1]$ and $\mu(w) \in \mathbb{R}_+$ such that $U_1(w) = c_1$ and $\pi(w) > K$.

Finally, consider the deviation of an individual worker. All other workers still use their mixed strategy. The competition for each job is therefore unchanged, and he still obtains payoff $U_i(\boldsymbol{w})$ when he applies to wages in the application set. Wages $w \in \mathcal{V} \backslash \mathcal{W}$ yield an offer for sure. Other wages cannot provide profitable deviation even if they are offered by some (possibly

[14] Inspection of the worker's problem in Section 3.1 reveals that even if both wages are not in the offer set it is impossible to obtain a utility above U^*, given this assumption.

deviating) firm when the worker's belief about the other workers' behavior corresponds to the belief summarized in the Market Utility Assumption.

We define an equilibrium as follows.

Definition 3 (Equilibrium) *An equilibrium is a tuple* $\{\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ *of strategies for the agents such that there exists* π^**,* U^* *and* $\mu(\cdot)$ *satisfying*

1. *(a)* $\pi(w) = \pi^* \geq \pi(w')$ *for all* $w \in \mathcal{V}$ *and* $w' \in [0,1]$ *if* $\phi > 0$*.*

   (b) $\pi^* = K$ *if* $\phi > 0$*.*

2. *(a)* $U_i(\boldsymbol{w}) \geq U_i(\boldsymbol{w'})$ *for all* $\boldsymbol{w} \in \mathrm{supp}G^i$ *and* $\boldsymbol{w'} \in [0,1]^i$*, if* $\gamma_i > 0$*.*

   (b) $U_i^* = U^*$ *if* $\gamma_i > 0$*.*

3. $\mu(\cdot)$ *conforms to (3) - (7) and the Market Utility and Entry Assumptions hold.*

Condition 1 a) and b) specify profit maximization and free entry.[15] Condition 2 a) implies that workers who send i applications send them optimally given the wage offer distribution and the behavior of other workers. Condition 2 b) ensures that workers send out the optimal number of applications. Condition 3 reiterates the determination of the effective queue length. The distinction between a) and b) allows the discussion of an exogenous number of applications and/or exogenous number of firms using the appropriate subset of conditions. While the exposition uses the terminology of a game, the definition resembles a competitive equilibrium with a somewhat non-standard feasibility constraint that embeds the frictions.

3 Equilibrium Characterization

In this section we characterize the equilibrium properties of the model and show

Summary 1 *An equilibrium exists. Generically the following holds: The equilibrium is unique; all workers send the same number of applications; the number of offered wages equals the number of applications; each worker applies with one application to each wage.*

We will proceed in three subsections: First we analyze the workers' search behavior given a distribution of wages and given the number of applications. Then we analyze the distribution of wages that firms will optimally set. Finally, we characterize equilibrium play.

[15] To ensure that entry implies zero profits it is sufficient to assume that $V > 1/K$.

3.1 Workers' Search Decision

To analyze the workers' search decisions, first consider a single worker who observes all wages and - given the strategy of the other workers - knows the probability of success at each wage. That is, he knows all pairs $(w, p(w))$. Equilibrium condition $2a)$ implies that each wage to which workers apply has to be optimal. For a worker with $i = 1$ the application choice is trivial. An application to w' is optimal if and only if $p(w')w' = u_1 \equiv \max_{w \in [0,1]} p(w)w$, i.e. he chooses a wage with the highest expected return u_1. For a worker with $i = 2$ the analysis is slightly more involved. Let $\bar{w}$ be the highest wage out of all wages that deliver u_1, i.e. $\bar{w} = \sup\{w \in [0,1] | p(w)w = u_1\}$.

Lemma 1 *Assume that an optimal choice for a worker with $i = 2$ exists. The optimal choice involves sending one application to a wage weakly below $\bar{w}$ and one application to a wage weakly above $\bar{w}$.*

Proof: The worker who sends $i = 2$ applications maximizes

$$\max_{(w_1,w_2)\in[0,1]^2} p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1, \tag{8}$$

where for now we omit the costs $c(2)$ of sending the applications, since they are fixed for a worker who is determined to send two applications. Note that we have set up problem (8) without the restriction that $w_1 \leq w_2$. Nevertheless it is immediate that a worker who has the choice between two wages will always accept the higher over the lower. Therefore any solution to (8) has $w_1 \leq w_2$.[16]

Next, note that w_1 is only exercised if w_2 failed. (8) immediately implies that for w_1 only the expected return $p(w)w$ is important, and his optimal decision resembles that of workers with a single application. I.e. he chooses w_1 such that

$$p(w_1)w_1 = u_1. \tag{9}$$

Taking this into account, any high wage w_2 is optimal if it fulfills

$$p(w_2)w_2 + (1 - p(w_2))u_1 = u_2, \tag{10}$$

[16] Assume a worker would choose (w_1, w_2). By (8) he gets $U(w_1, w_2) = p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1$. Now assume he reversed the order to get $U(w_2, w_1) = p(w_1)w_1 + (1 - p(w_1))p(w_2)w_2$. $U(w_1, w_2) \geq U(w_2, w_1)$ if and only if $w_2 \geq w_1$.

where $u_2 \equiv \sup_{w\in[0,1]} p(w)w + (1-p(w))u_1$. Clearly any combination of w_1 and w_2 that satisfies (9) and (10) solves the maximization problem (8). Since we know that any solution to the latter problem has $w_2 \geq w_1$, it has to hold that the highest low wage associated with (9) has to be weakly lower than the lowest high wage associated with (10). The highest low wage is given by $\bar{w}$. *Q.E.D.*

At high wages the worker takes into account the possibility of obtaining a low wage offer. He is willing to accept a lower expected pay ($p(w_2)w_2 < u_1$) in return for a high upside potential if he gets a job (high w_2), because if he does not get the good job the low wage application acts as a form of insurance.[17] Since all workers face the same maximization problem we obtain the following

Proposition 1 (Workers' Application Behavior) *In any equilibrium in which some workers send applications, i.e. $\gamma_1 + \gamma_2 > 0$, they apply in such a way that the following conditions for the hiring probability $p(w)$ [and for the effective queue length $\mu(w)$ because of (6)] hold:*

$$
\begin{aligned}
p(w) &= 1 \;\; \forall\, w \in [0, u_1] && (11)\\
p(w)w &= u_1 \;\forall\, w \in [u_1, \max\{\bar{w}, 1\}] && (12)\\
p(w)w + (1-p(w))u_1 &= u_2 \;\forall\, w \in [\bar{w}, 1], && (13)
\end{aligned}
$$

for some tuple $(u_1, u_2, \bar{w})$. For the tuple it holds that

i) u_1 is the highest utility obtainable by a single application (excluding application costs), i.e. $u_1 = U_1^* + c(1) = \max_{w\in\mathcal{V}} p(w)w$.

ii) if a positive fraction $\gamma_2 > 0$ of workers sends two applications, then u_2 is the highest utility obtainable with two applications (excluding application costs), i.e. $u_2 = U_2^* + c(2) = \max_{w\in\mathcal{V}} p(w)w + (1-p(w))u_1$. The cutoff wage is $\bar{w} = u_1^2/(2u_1 - u_2)$.

iii) if workers do not send two applications ($\gamma_2 = 0$), then u_2 is given by an indifference between applying and bearing the costs, i.e. $u_2 = u_1 + c_2$, and $\bar{w} = u_1^2/(u_1 + c_2)$. If $\bar{w} = u_1^2/(u_1 + c_2) > 1$ the set $[\bar{w}, 1]$ is empty.

Low wages do not receive applications, and therefore probability of getting a job for somebody who wants to apply is one. An applicant could get the job for sure but still no applicants

[17] See Chade and Smith (2006) and Galenianos and Kircher (2006) for a longer discussion of the tradeoffs under simultaneous search.

want to apply. Wages in the intermediate range receive the low applications that workers are only willing to send if (9) is fulfilled, and high wages receive high applications under condition (10). We should note that even if workers do not send high wage applications, i.e. $\gamma_2 = 0$, the queue lengths at high wages might be determined by (10). If a deviant posts a high wage, the Market Utility Assumption specifies that workers are indifferent. If the second application is quite costly, indifference implies that the workers are indifferent between sending their single application to the deviant or to the offered wage. Yet if the second application is not very costly, it might be optimal to continue to send the single application to some offered wage but to send an additional application to the deviant. Indifference then implies that the marginal benefit of the additional application is zero, i.e. $u_2 - u_1 - c_2 = 0$, which then determines u_2 and governs the queue length at high wages.

3.2 Firms' Wage Setting

This subsection focuses on the nature of equilibrium wage dispersion. We first show that wage dispersion is a necessary feature of any equilibrium with $\gamma_2 > 0$. We then show that this leads to exactly two wages being offered in equilibrium.

Consider the case where $\gamma_2 > 0$, which implies that $v > 0$ as otherwise there is no reason to apply. Before we proceed, we will briefly rewrite firms' profits in a convenient way. Consider some (candidate) equilibrium characterized by u_1, u_2 and $\bar{w}$, which by Proposition 1 characterizes the workers application behavior. We will call firms that end up offering wages below $\bar{w}$ as low wage firms, and those offering wage above $\bar{w}$ as high wage firms. The problem for each individual firm is to maximize $(1 - e^{-\mu(w)})(1 - w)$ under the constraint that $\mu(w)$ is given by (11), (12) and (13). This is a standard maximization problem. Writing the profit function as $\pi(w) = \mu(w)p(w)(1 - w)$, we can use the constraint to substitute out the wage and write profits as a function of the effective queue length

$$
\begin{aligned}
\pi(\mu) &= 1 - e^{-\mu} - \mu u_1 && \forall\ \mu \in [\underline{\mu}, \bar{\mu}], && (14)\\
\pi(\mu) &= (1 - e^{-\mu})(1 - u_1) - \mu(u_2 - u_1) && \forall\ \mu \in [\bar{\mu}, \mu(1)], && (15)
\end{aligned}
$$

where $\bar{\mu} = \mu(\bar{w})$ and $\underline{\mu} = 0 = \mu(u_1)$. We can interpret this as follows. Firms that offer a wage below u_1 "buy" a queue length of zero and make zero profits. Low wage firms that offer wages in $[u_1, \bar{w}]$ "buy" a queue length according to (12) and obtain profits as in (14), while high wage firms with wages in $[\bar{w}, 1]$ "buy" a queue length as in (13) and obtain profits as in (15). Since at wages above u_1 there is a one-to-one relation between the wage and the effective queue length, we can view the individual firm's problem as simply a choice regarding the preferred effective queue length.

The profit function is continuous, but has a kink at $\bar{w}$ (respectively $\bar{\mu}$). This is due to the fact that workers trade off the effective queue length against the wage differently for high and low applications. At high wages the queue length responds stronger to a wage change since workers are more "risky" due to the fallback option at low wages. This kink implies immediately that it cannot be optimal for a firm to offer wage $\bar{w}$, because raising the wage induces many additional effective applications while reducing the wage induces only a relatively small reduction in effective applications. Therefore either it is profitable to offer a higher wage, or if that is not profitable then it is profitable to offer a lower wage.

This rules out an equilibrium in which some workers send more than one application but all firms offer the same wage, because the offered wage would coincide with the cutoff wage and individual firms would want to deviate.

Proposition 2 (Wage Dispersion) *There does not exist an equilibrium with $\gamma_2 > 0$ in which only one wage is offered, i.e. in which $\mathcal{V}$ is a singleton.*

We should note that the argument that rules out one-wage equilibria is different from those in most other papers on wage dispersion with homogenous workers and firms. Usually there is an appeal to a discontinuity of the following kind: If all firms offer the same wage, there is a strictly positive probability that a firms' offer is rejected because the worker accepts some other equally good offer; so if a deviating firm offers a slightly higher wage, at least as many workers apply, and all applicants accept an offer for sure. This yields a jump in profits.[18] In this model there is no discontinuity despite the fact that workers accept an offer for sure at a slightly higher wage. This positive jump in profits is offset by the fact that *fewer* workers apply to the deviant.[19] Workers internalize that only effective applications imply competition. At the market wage, only a fraction of the (other) applications are effective, while at the deviant all applications are effective. If the deviants' wage is only slightly higher, less workers apply because otherwise the competition would make an application unattractive. As a consequence profits change continuously. Nevertheless, the kink in the profit function induced by a different "risk-return"-tradeoff of workers implies wage dispersion.

[18]This happens e.g. in Burdett and Judd (1983), Burdett and Mortensen (1998), Acemoğlu and Shimer (2000), Gautier and Moraga-González (2004) and Galenianos and Kircher (2006).

[19]In the introduction we explained "directedness" as the ability to attract more applications when offering higher wages. By this we mean more *effective* applications. As shown here, gross applications do not need to be higher.

Next we show that in an equilibrium in which some workers send two applications exactly two wages will be offered, one strictly below and one strictly above $\bar{w}$. This immediately implies that workers with one application send it to the low wage, and workers with two applications send one to each of the wages.

Proposition 3 (Discrete Set of Offered Wages) *In any equilibrium with $\gamma_2 > 0$, exactly two distinct wages will be offered, i.e. $\mathcal{V} = \{w_1^*, w_2^*\}$. It holds that $w_1^* < \bar{w} < w_2^*$.*

Proof: Since wage dispersion implies that not all wages are zero, $u_2 > u_1 > 0$. Individual firms take u_1, u_2 and $\bar{w}$ as given. For low wage firms we can write the profits as a function of the queue length as in (14). The function is strictly concave on $[0, \bar{\mu}]$. Therefore all low wage firms will offer the same wage. For high wage firms profits can be written as in (15), which is strictly concave on $[\bar{\mu}, \mu(1)]$. Therefore all high wage firms will offer the same wage. Finally, assume one group, say low wage firms, offered the wage $\bar{w}$. Since there is wage dispersion, high wage firms will offer $w_2^* > \bar{w}$. But since their problem is strictly concave on $[\bar{w}, 1]$, they make strictly higher profits than firms at $\bar{w}$, which yields the desired contradiction. A similar argument rules out that $\bar{w}$ is offered by high wage firms. *Q.E.D.*

Given that only two wages are offered in equilibrium, we will for notational simplicity index variables referring to low wage firms by *1* and those referring to high wage firms by *2*.[20] Let d_1 be the equilibrium fraction of firms offering the low wage, and d_2 the fraction offering the high wage. Then $v_i = vd_i$ denotes the measure of firms at the respective wage, and equation (3) implies gross queue lengths (i.e. measure of applications per firm) of $\lambda_2 = \gamma_2/v_2$ and $\lambda_1 = (\gamma_1 + \gamma_2)/v_1$. At high wages workers only apply strictly lower, so that every worker is available for hiring and $\mu_2 = \lambda_2$. At low wages, a fraction $\gamma_2/(\gamma_1 + \gamma_2)$ applies to the high wage, and so by (7) a fraction $\psi_2 = p_2\gamma_2/(\gamma_1 + \gamma_2)$ of applicants cannot be hired because they got better jobs, where $p_2 = (1 - e^{-\mu_2})/\mu_2$. Therefore the effective number of application per firm at the low wage is $\mu_1 = (1 - p_2\gamma_2/(\gamma_1 + \gamma_2))\lambda_1$. With these notational simplifications we establish

Corollary 1 *In an equilibrium with $\gamma_2 > 0$ profits and wages for high and low wage firms*

[20]That is, let π_i be the profit, w_i the wage, λ_i the gross queue length, μ_i the effective queue length, η_i the hiring probability, p_i the probability of getting an offer when applying at a type-i firm, and ψ_i the probability that a worker accepts another offer, where $i = 1$ when we refer to low wage firms and $i = 2$ when we refer to high wage firms.

respectively are given by

$$\pi_1 = 1 - e^{-\mu_1} - \mu_1 e^{-\mu_1}, \tag{16}$$

$$w_1^* = \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1}), \tag{17}$$

$$\pi_2 = (1 - e^{-\mu_2} - \mu_1 e^{-\mu_1})(1 - e^{-\mu_1}), \text{ and} \tag{18}$$

$$w_2^* = \mu_2 e^{-\mu_2}(1 - e^{-\mu_1})/(1 - e^{-\mu_2}) + e^{-\mu_1}. \tag{19}$$

Proof: We know that neither low wage nor high wage firms are constrained, because the equilibrium wages are different from $\bar{w}$, and it is easy to see that $w_1 > 0$ (otherwise workers would not apply) and $w_2 < 1$ (otherwise high wage firms would make less profits than low wage firms). Therefore wages are given by first order conditions. For low wage firms, the first order condition of (14) with respect to μ leads to

$$u_1 = e^{-\mu} = e^{-\mu_1}. \tag{20}$$

The second equality follows since in equilibrium all low wage firms will choose the same queue length, or rather the wage associated with it. When substituted into (14) this leads to the expression for the profits. By (12) we know that $w_1^* = u_1/p_1$ and we immediately get the corresponding wage. For high wage firms, the first order condition of (15) implies

$$u_2 - u_1 = e^{-\mu}(1 - u_1) = e^{-\mu_2}(1 - e^{-\mu_1}). \tag{21}$$

Substitution back into (15) yields the expression for the profits. By (13) we know that $w_2^* = (u_2 - u_1)/p_2 + u_1$, and substitution leads to the expression for the high wage. *Q.E.D.*

In the case of a single application ($\gamma_1 > 0$ and $\gamma_2 = 0$) the arguments above easily establish that only one wage is offered according to (17), yielding profits given by (16). Obviously in this case the probability that an offer leads to a hire is one, i.e. $\mu_1 = \lambda_1$. This is also the result obtained in Burdett, Shi and Wright (2001). The introduction of a second application essentially establishes two markets. The profits in each are given by $(1-e^{-\mu_i}-\mu_i e^{-\mu_i})(1-u_{i-1})$. In the low market u_0 is identical to the workers' true outside option of zero, but there is some connection to the high market induced by the strictly positive probability that an offer is rejected. In the high market the rejection probability is zero, but u_1 is greater than zero as it reflects the workers' endogenous outside option induced by the presence of the low market. Apart from these spillovers, each market operates essentially as a single one-application market.

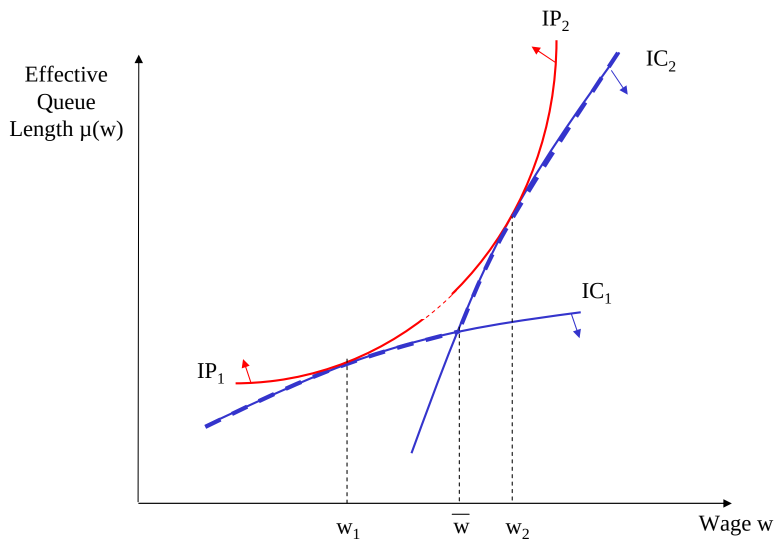

Figure 1: Equilibrium behavior. IC_1 and IC_2: Worker's indifference curve for the low and high wage, respectively. IP_1 and IP_2: Isoprofit curves for low and high wage firms.

The findings in the last sections are summarized in figure 1. The workers' indifference curve IC_1 for the low wage applications is given by (12). Low wage firms take this into account and offer a wage w_1 such that no individual firm wants to deviate, which means their isoprofit curve IP_1 is tangent to IC_1. The indifference curve IC_2 for the high application is by (13) steeper than for the low application because of the fallback option due to the low application. The actual queue length that firms expect is the dashed line. High wage firms take this into account and offer wage w_2 in such a way that no firm wants to deviate, i.e. such that their isoprofit curve IC_2 is tangent to IP_2. Note that a single wage $w_1 = w_2 = \bar{w}$ cannot be an equilibrium, because at the kink of the indifference curves it is impossible to place the isoprofit curve tangent, and therefore firms would want to deviate. The isoprofit curves for low and high wage firms have to coincide to provide equal profits for firms. In the next section we prove that this is possible.

3.3 Equilibrium Outcome

In this section we derive the equilibrium outcome. Before we turn to the full equilibrium, it will be a useful first step to exogenously fix the number of applications that workers send. We will show existence and uniqueness of an (appropriately adjusted) equilibrium with and without free entry.[21]

Lemma 2 *For given* $(\gamma_0, \gamma_1, \gamma_2)$ *with* $\gamma_1 + \gamma_2 > 0$ *it holds that*

1. *for given* $v > 0$ *there exists unique* $(F, \boldsymbol{G})$ *such that equilibrium conditions* 1*a*), 2*a*) *and* 3 *hold;*

2. *with free entry there exists unique* $(\phi, F, \boldsymbol{G})$ *such that equilibrium conditions* 1*a*), 1*b*), 2*a*) *and* 3 *hold.*

Proof: We will show part 2 here, part 1 is relegated to the appendix. The effective queue length μ_1 at the low market wage is according to (18) determined by

$$1 - e^{-\mu_1} - \mu_1 e^{-\mu_1} = K. \tag{22}$$

Since the left hand side of (22) is strictly increasing in μ_1 and is zero for $\mu_1 = 0$ and one for $\mu_1 \to \infty$, μ_1 is unique. The wage is given by (17).

If $\gamma_2 > 0$, the queue length at high wage firms is by (18) given by

$$(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})(1 - e^{-\mu_1}) = K. \tag{23}$$

Since μ_1 is unique, μ_2 is unique. The high wage is given by (19). Since $\mu_1^* = (\gamma_1 + \gamma_2 - \gamma_2 p_2)/v_1$ and $\mu_2^* = \gamma_2/v_2$, both v_1 and v_2 are uniquely determined, which characterizes the equilibrium entry and the randomization over the two wages.

Clearly firms are willing to offer these wages. The wages were determined by the appropriate first order conditions and therefore no other wage in either the high wage or the low wage region can offer a higher profit. Since (12) and (13) were used as constraints to construct the profits, they remain valid and workers are indeed willing to apply in the prescribed way. *Q.E.D.*

Knowing the equilibrium interaction for a fixed number of applications, we can turn to the analysis of the equilibrium interaction when the number of applications is endogenous. c_1 and c_2 may take any non-negative values as long as $c_1 \leq c_2$. In analogy to the free entry conditions (22) and (23) we will define the following four numbers μ_1^*, μ_2^*, u_1^* and u_2^* recursively as follows:

[21] Note that the equilibrium definition does not tie down F in case $\phi = 0$ and G_i in case $\gamma_i = 0$. For the discussion of uniqueness, assume that in these cases the respective distribution takes on some unique form. As a technical detail, note that here we fix $\boldsymbol{\gamma}$ but in the Market Utility Assumption the costs still show up. To ensure consistency, assume that when $\gamma_2 = 0$ only one application plays a role (e.g. $c_1 = 0$ and $c_2 = \infty$) while when $\gamma_2 > 0$ both play a role (e.g. $c_1 = c_2 = 0$).

$(1 - e^{-\mu_2^*} - \mu_2^* e^{-\mu_2^*})(1 - u_{i-1}^*) = K$, where $u_i^* - u_{i-1}^* = e^{-\mu_i^*}(1 - u_{i-1}^*)$ and $u_0^* = 0$. These numbers are uniquely determined by the exogenous parameter K. Moreover, (20) established that the marginal utility of the first application in equilibrium is always u_1^* whenever at least some workers send out applications, and by (21) the marginal utility of the second application is $u_2^* - u_1^*$ whenever at least some workers send two applications. This is independent of the exact structure of $\boldsymbol{\gamma}$. We will establish the following proposition, which is a stronger version of Summary 1 with which we started the section.

Proposition 4 (Equilibrium Outcomes) *An equilibrium exists. Furthermore*

1. *For $c_1 > u_1^*$ the unique equilibrium has $v = 0$ and $\boldsymbol{\gamma} = 0$.*

2. *For $c_1 < u_1^*$ and $c_2 > u_2^* - u_1^*$, in the unique equilibrium all workers send one application and one wage is offered.*

3. *For $c_1 < u_1^*$ and $c_2 < u_2^* - u_1^*$, in the unique equilibrium all workers send two applications, two wages are offered, and each worker applies to each wage.*

Proof: In case 1 the marginal utility is too low to induce any worker to send the first application. Therefore $v = 0$ and $\boldsymbol{\gamma} = 0$. This is consistent with the Entry Assumption.

In cases 2 and 3, the marginal utility is strictly higher than the marginal cost of the first application. By the Entry Assumption there will be positive entry: At queue length μ' such that $e^{-\mu'} = c$ and wage $w' = \mu' e^{-\mu'}/(1 - e^{-\mu'})$, it holds that $U_1(w') = c$ and profits $\pi' = 1 - e^{-\mu'} - \mu' e^{-\mu'} > K$. Given positive entry, clearly each worker will send at least one application.

Now we have to analyze if it can be the case that workers only apply once, and firms only offer one wage. Assume this is the case. For $c_2 < e^{-\mu_1^*}(\mu_1^* - 1 + e^{-\mu_1^*})/\mu_1^*$ it can be shown that we get a contradiction, because a worker strictly prefers to apply twice at the unique market wage than to send only one application. Yet even if c_2 is not that small, firms might not be willing to offer only one wage - despite the fact that the wage is determined by their first order condition. At high (not offered) wages the queue length might increase fast because workers would send their high application if these high wages were offered, which is reflected in $\bar{w} < 1$ in the part *iii*) of Proposition 1. Since the queue length is continuous, and the offered wage is strictly optimal on $[u_1, \bar{w}]$, a firm that is looking for a profitable deviation has to find the optimal wage in the interior of $[\bar{w}, 1]$. Since $u_2 = c_2 + u_1$ according to Proposition 1, we have by (15) the profit $\pi(\mu) = (1 - e^{-\mu})(1 - e^{-\mu_1^*}) - \mu c_2$ for a deviant that offers a wage in $(\bar{w}, 1)$. If there is a profitable deviation, it must be profitable to deviate to $\hat{\mu}$ given by the first order condition $e^{-\hat{\mu}}(1-e^{-\mu_1^*}) = c_2$, which implies $\hat{\mu} < \mu_2^*$ in case 2 and $\hat{\mu} > \mu_2^*$ in case 3. Substitution leads to an optimal deviation profit of

$$\pi(\hat{\mu}) = (1 - e^{-\hat{\mu}} - \hat{\mu}e^{-\hat{\mu}})(1 - e^{-\mu_1^*}). \tag{24}$$

Comparing (24) with (23) establishes that $\pi(\hat{\mu})$ is strictly smaller than K in case 2, making a deviation unprofitable, and strictly larger than K in case 3, yielding a strictly profitable deviation (the wage associated with $\hat{\mu}$ is indeed above $\bar{w}$ in case 3). Therefore an equilibrium with one wage is possible in case 2 and not in case 3.

Finally, it is immediate that in case 2 an equilibrium with two wages cannot exist because by $u_2^* - u_1^* < c_2$ the marginal utility of the second application is too low, while an equilibrium with two wages can exist in case 3 since $u_2^* - u_1^* > c_2$. Therefore in case 2 everyone sends one application to the unique wage, while in case 3 every worker sends two applications, one to each of the two wages. Uniqueness is then ensured by Lemma 2. *Q.E.D.*

For the case $c_2 < u_2^* - u_1^*$ it is worth emphasizing that two wages are offered because firms anticipate that workers will send an additional application when they offer a high wage (this is captured by the Market Utility Assumption). If we consider a candidate equilibrium in which all firms offer a single wage, it is this feature that leads to a high queue length for a deviant with a high wage and makes such a deviation profitable.

In the case where $c_1 = e^{-\mu_1^*}$ we have multiplicity of equilibria: for any $\gamma_1 \in [0, 1]$ and $\gamma_0 = 1 - \gamma_1$ an equilibrium exists, and workers are exactly indifferent between applying once and not applying. If $c_2 = e^{-\mu_2^*}(1 - e^{-\mu_1^*})$ an equilibrium exists in which workers randomize between one and two applications, i.e. it exists for any $\gamma_2 \in [0, 1]$ and $\gamma_1 = 1 - \gamma_2$.

4 Efficiency

To discuss the efficiency properties of the equilibria just characterized, we will follow Pissarides (2000) and others by using the following notion of constrained efficiency: An equilibrium is constrained efficient if it maximizes the output minus entry and application costs, given the frictions in the market. The frictions stem from the requirement that workers and firms use some symmetric strategies $(\boldsymbol{\gamma}, \boldsymbol{G})$ and (ϕ, F) and are then allocated according to the recall process. Denoting by Υ^i the set of cumulative distribution functions over $[0, 1]^i$ and by $\triangle_3$ the three-dimensional unit simplex, the strategy spaces of workers and firms are $\mathcal{G} = \triangle_3 \times \Upsilon^1 \times \Upsilon^2$ and $\mathcal{F} = [0, 1] \times \Upsilon^1$. Then

Definition 4 (Constrained Efficiency) *An equilibrium* $\{\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}\} \in \mathcal{F} \times \mathcal{G}$ *is constrained efficient if it maximizes*

$$\max_{(\phi',F')\in\mathcal{F},(\boldsymbol{\gamma}',\boldsymbol{G}')\in\mathcal{G}} M(\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}') - \phi' V K - \gamma_1' c(1) - \gamma_2' c(2), \tag{25}$$

where $M(\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}') = \phi' V \int_0^1 \eta(w) dF'$ *is the number of matches when* $\eta(w)$ *is determined by (3) - (7) using the relevant parameters* $\{\phi', F', \boldsymbol{\gamma}', \boldsymbol{G}'\}$.

As discussed in the introduction, efficiency might fail on any of the operative margins: (1) Application inefficiency: For a given vector of applications and a given number of firms, the application behavior might lead to a suboptimal number of matches. (2) Entry inefficiency: For a given vector of applications too many or too few firms enter. (3) Inefficiency of search intensity: Workers apply too much or too little given the costs. If such inefficiencies arise, government interventions such as minimum wage legislation may improve efficiency.

While inefficiencies arise if firms can only communicate with a single one of their applicants, we will show that in our setting directed search balances all these margins.[22] This implies that interventions aimed at improving the market outcome would have to target the decentralized elements of the application or assignment process. Simple interventions such as minimum wage legislation or taxation will not improve the market outcome.

Shimer (1996) explains the efficiency property for the one-application case roughly as follows: The workers' response to a change in the wage yields an implicit price for the desired queue length, therefore firms can price the queue length of applicants exactly at its marginal cost. In this model two variables matter: The gross queue length $\lambda(\cdot)$ and the retention probability $1 - \psi(\cdot)$. Both have to be adjusted by a single-dimensional wage. This is possible because for firms *and* for workers only the combination of both matters. The workers respond to a change in the wage by changing their applications such that the *effective* queue length $\mu(\cdot) = [1-\psi(\cdot)]\lambda(\cdot)$ rises to a new level of indifference. Therefore the same logic holds here. The effective queue length is priced at marginal cost.

Proposition 5 (Efficiency) *The equilibrium market outcome is constrained efficient.*

We will prove the proposition in the next three subsections that are dedicated to different margins. For a given number of firms and a given vector of applications, optimization according to equilibrium conditions $1a$), $2a$) and 3) leads to an optimal number of matches. We show that wage dispersion is crucial for the market to realize the efficient outcome. Adding free

[22] See the introduction and section 6.2 for a discussion of restricted recall.

entry condition 1b) yields also an optimal number of firms in the market. Adding additionally condition 2b) for the number of applications yields an optimal search intensity, and therefore constrained efficiency of the overall equilibrium interaction.

4.1 Application Efficiency

For a given vector of applications $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ and firms v, we will show that the search outcome as characterized in the first part of Lemma 2 is constrained efficient. As we will see, this depends on the ability of the market to generate different wages for low and high applications.

For $\gamma_2 > 0$ we start by analyzing a narrower concept that we will call 2-*group-efficiency*. We will assume that there are two groups of firms, one preferred over the other, and all workers who send at least one application send one at random to the non-preferred group, and workers who send two applications send the second one at random to the preferred group. This setup corresponds to the equilibrium outcome. Let $d \in [0,1]$ be the fraction of firms in the preferred group. Search is two-group-efficient if d is chosen optimally given the assumptions just made.

Lemma 3 *For a given $v > 0$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ with $\gamma_2 > 0$, the strategy combination implied by equilibrium conditions* 1a)*,* 2a) *and* 3 *yields two-group-efficient search.*

Proof: The optimization problem is given by

$$\max_{d \in [0,1]} M(d) = vd(1 - e^{-\mu_2}) + v(1-d)(1 - e^{-\mu_1}), \tag{26}$$

where $\mu_1 = (1 - \gamma_2 p_2/(\gamma_1 + \gamma_2))\lambda_1$, $\mu_2 = \lambda_2 = \gamma_2/(vd)$, $\lambda_1 = (\gamma_1 + \gamma_2)/(v(1-d))$, $\lambda_2 = \gamma_2/(vd)$, and $p_2 = (1 - e^{-\lambda_2})/\lambda_2$. The first derivative is given by

$$\begin{aligned} \frac{\partial M}{\partial d} = v \quad \Bigg[& \quad 1 - e^{-\lambda_2} - (1 - e^{-\mu_1}) + de^{-\lambda_2}\frac{\partial \lambda_2}{\partial d} \\ & + \quad \frac{e^{-\mu_1}}{v\lambda_1}\left[-\gamma_2\lambda_1 \frac{\partial p_2}{\partial d} + (\gamma_1 + \gamma_2 - \gamma_2 p_2)\frac{\partial \lambda_1}{\partial d} \right] \Bigg]. \end{aligned}$$

The last expression in the first line equals $-\lambda_2 e^{-\lambda_2}$ and the second line equals $e^{-\mu_1}[-(1 - e^{-\lambda_2} - \lambda_2 e^{-\lambda_2}) + \mu_1]$. This yields

$$\frac{\partial M}{\partial d}\frac{1}{v} = (1 - e^{-\lambda_2} - \lambda_2 e^{-\lambda_2})(1 - e^{-\mu_1}) - (1 - e^{-\mu_1} - \mu_1 e^{-\mu_1}) = 0, \tag{27}$$

where the last equality denotes the first order condition. (27) coincides with the equal profit condition between high and low wage firms. By the proof of Proposition 2 part 1 we know that this uniquely determines the measure of firms in the high and the low group. Boundary solutions cannot be optimal because one application would be waisted (global concavity follows mathematically from $\partial^2 M/\partial d^2 = -v[\lambda_2^2 e^{-\lambda_2}(1-e^{-\mu_1})/d + e^{-\mu_1}(1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2}-\mu_1)^2/(1-d)] < 0$). *Q.E.D.*

Next we show that the search outcome is constrained efficient. The proof relies on establishing that two groups are sufficient to obtain the optimal allocation.

Proposition 6 (Application Efficiency) *For given $v = \phi V$ and given $\boldsymbol{\gamma}$, the application process is constrained efficient, i.e. it holds that*

$$M(\phi, F, \boldsymbol{\gamma}, \boldsymbol{G}) = \max_{F' \in \Upsilon^1, \boldsymbol{G}' \in \Upsilon^1 \times \Upsilon^2} M(\phi, F', \boldsymbol{\gamma}, \boldsymbol{G}'),$$

where $\{\boldsymbol{G}, F\}$ conform to equilibrium conditions 1*a*)*,* 2*a*) *and* 3*.*

Finally, we show that two groups are indeed necessary to obtain the optimal allocation when a fraction of workers send two applications. One group of firms that all have equal hiring probability is not efficient. This result reveals that wage dispersion is necessary for a constrained efficient search process. If only a single wage were offered, i.e. $\mathcal{V} = \{w\}$ for some w, then then all firms would have the same hiring probability, which we show to be inefficient.

Proposition 7 (Inefficiency without Wage Dispersion) *Given $v = \phi V$ and $\boldsymbol{\gamma}$ with $\gamma_2 > 0$, identical hiring probabilities for all firms cannot be constrained efficient. That is, if $F \in \Upsilon^1$ and $\boldsymbol{G} \in \Upsilon^1 \times \Upsilon^2$ such that $\eta(w) = \bar{\eta}\ \forall\ w \in \mathcal{V}$ then $(\boldsymbol{G}, F) \notin \arg\max_{F' \in \Upsilon^1, \boldsymbol{G}' \in \Upsilon^1 \times \Upsilon^2} M(\phi, F', \boldsymbol{\gamma}, \boldsymbol{G}')$. In particular, a single offered wage $\{w\} = \mathcal{V}$ cannot achieve constrained efficiency if $\gamma_2 > 0$.*

In the proof we show that two groups can achieve the same hiring probabilities as a random process. But the non-preferred group is too small compared to the optimum. All workers would take a job from a firm in the preferred group. For those that end up taking jobs with firms in the non-preferred group this is their last chance to avoid unemployment. Increasing workers' matching probability in the non-preferred group at the cost of decreasing their matching probability in the preferred group improves matching for those workers for whom it is the last option to avoid unemployment at the expense of a lower matching probability for those who still might have another option.

This result is surprising because with one application different hiring probabilities for firms are only warranted when there are productivity differences.[23] Here the source for different hiring probabilities is a sorting externality. Figure 2 illustrates this. At a unique market wage representing a random application behavior, the indifference curve IC_1 for the low and IC_2 for the high applications cross at the same point as the firms isoprofit curve IC. Since the actual (dashed) indifference curve is kinked, it is not possible to achieve tangency with the isoprofit curve. Area A indicates mutual gains for workers and firms from sending the low applications to firms with different queue length and wage. Similar gains are indicated by area B for the high applications. The exact choice of w influences which gains are more prominent, but due to the kink it is never possible to eliminate both. Figure 1 shows that it is possible to achieve tangency with two wages.[24]

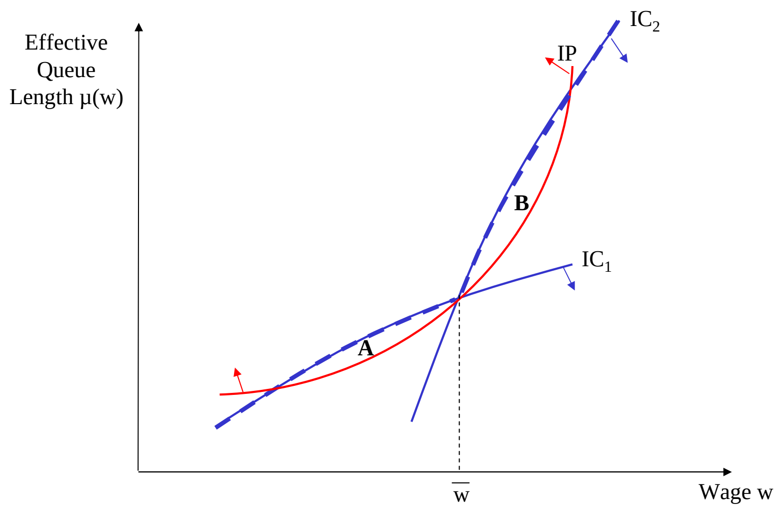

Figure 2: Inefficiencies at a unique market wage. Areas A and B indicate mutual benefits for both workers and firms.

Note that we have restricted the efficiency notion to strategies that are anonymous. In

[23]Shimer (2005) analyzes productivity differences when only one application is possible. If workers and firms are homogenous only one hiring probability would be efficient (similar to our case for $\gamma_1 > 0$ but $\gamma_2 = 0$), and only with heterogeneities of firms or workers different hiring probabilities are efficient.

[24] Neither the figure nor the intuition apply to the case without recall. Without recall, it is not possible to graph all relevant aspects in two dimensions, because next to the wage both the gross queue length $\lambda(\cdot)$ and the retention probability $\psi(\cdot)$ remain separately important. Regarding the intuition, without recall equal hiring probabilities for workers mean that the number of workers who have their last option in either group is the same. Only with recall the rejected positions in the low group become available again for yet unmatched workers, which induces a proportionally larger number of workers employed without another option in the non-preferred group, yielding a positive externality when placing relatively more firms in the non-preferred group.

specific, workers treat firms with the same wage identically. We noted that anonymity is merely for notational convenience (footnote 8 and 27), and that the real restriction that induces the frictions is the assumption that workers use symmetric strategies. It can be shown that symmetric strategies imply that workers will apply to firms in such a way that firms at the same wage have equal hiring probabilities (otherwise those with low hiring probabilities would be strictly more attractive). Therefore, a market in which all firms offer the same wage cannot be efficient even without the anonymity restriction, because still all firms would end up having the same hiring probability.

4.2 Entry efficiency

We now turn to the efficiency of the entry decision. Let $M^*(v, \boldsymbol{\gamma})$ be the number of matches when there are $\boldsymbol{\gamma}$ applications and v firms and application behavior is constrained efficient. For given $\boldsymbol{\gamma}$ we have determined in the second part of Proposition 2 the unique entry in the (appropriately adjusted) equilibrium. This entry is constrained efficient given the search intensity γ.

Proposition 8 (Entry Efficiency) *Given $\boldsymbol{\gamma}$, entry is constrained efficient. That is*

$$M^*(v, \boldsymbol{\gamma}) - vK = \max_{v' \in [0,\infty)} M^*(v', \boldsymbol{\gamma}) - v'K, \tag{28}$$

where v arises when equilibrium conditions 1*a*)*,* 1*b*)*,* 2*a*) *and* 3 *are fulfilled.*

Proof: The number of matches is given by $M^*(v, \boldsymbol{\gamma}) = 1 - \gamma_2 \prod_{i=1}^{2}(1 - p_i) - \gamma_1(1 - p_1)$, where p_1 and p_2 are the probabilities of getting a job at less preferred and the more preferred group under two-group-efficient search. If $\gamma_0 = 1$, then $v = 0$ arises and is clearly optimal. If $\gamma_0 < 1$, clearly $v = 0$ is not optimal given $K < 1$. $v \to \infty$ is also not optimal since the number of matches is bounded by the measure of workers but costs would grow unboundedly. The first order condition to the problem is

$$\begin{aligned} K &= \gamma_2[\partial p_1/\partial v](1 - p_2) + \gamma_2[\partial p_2/\partial v](1 - p_1) + \gamma_1[\partial p_1/\partial v] \\ &= [\partial p_1/\partial v](\gamma_1 + \gamma_2)(1 - \psi_1) + [\partial p_2/\partial v](1 - p_1). \end{aligned} \tag{29}$$

p_i depends on v directly since the measure $d_i v$ of firms in group i depends on v directly. It also depends on v indirectly since the two-group-efficient fraction d_i is a function of v. Yet by the envelop theorem the indirect effect is zero and we can neglect the effect on d_i. Consider the first term on the right hand side first. We can write $\partial p_1/\partial v = [\partial p_1/\partial \mu_1][\partial \mu_1/\partial v]$. One can show that $[\partial\mu_1/\partial v](\gamma_1+\gamma_2)(1-\psi_2) = -d_1\mu_1^2 - \gamma_2\mu_1[\partial p_2/\partial v]$. Noting that $[\partial p_1][\partial v] = -[(1-e^{-\mu_1}-\mu_1 e^{-\mu_1})/\mu_1^2][\partial\mu_1/\partial v] = -[(p_1-e^{-\mu_1})/\mu_1][\partial\mu_1/\partial v]$ we obtain

$$[\partial p_1/\partial v](1-p_2) = d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(p_1-e^{-\mu_1})[\partial p_2/\partial v].$$

Then (29) reduces to

$$\begin{aligned} K &= d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(1-e^{-\mu_1})[\partial p_2/\partial v] \\ &= d_1(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + d_2(1-e^{-\mu_1})(1-e^{-\mu_2}-\mu_2 e^{-\mu_2}) \\ &= 1-e^{-\mu_1}-\mu_1 e^{-\mu_1}, \end{aligned}$$

where the second line follows by taking the appropriate derivative and the last line follows as a consequence of two-group-efficient search (see (27)). The last line also denotes the profits of low wage firms in equilibrium. Applying (27) again yields a condition equal to the profits of high wage firms. The first order condition is unique by the same argument that established that a unique v implies zero profits, and the entry implied by equilibrium conditions 1*a*), 1*b*), 2*a*) and 3 coincides with the entry implied by the first order condition. Since the first order condition is unique and boundary solutions are not optimal, it describes the global optimum. *Q.E.D.*

4.3 Efficiency of Search Intensity

The number of applications that workers send in equilibrium is also constrained efficient. We will account for the associated entry of firms and the search outcome, and therefore also immediately establish the overall constrained efficiency of the equilibrium as in Proposition 5.

To gain intuition, consider the case of an individual worker who sends one application rather than none. By the above analysis his marginal benefit is $e^{-\mu_1^*} - c_1$. The benefit for society comprises the cost $-c_1$ and the additional production of one unit of output in the case that the firm did not have another effective applicant, the probability of which is $e^{-\mu_1^*}$. Therefore private and social benefits coincide. Similarly, if two wages are offered and a worker sends a second application rather than only a single one, his private marginal benefit is $e^{-\mu_2^*}(1-e^{-\mu_1^*}) - c_2$. Additional production arises only if the high firm does not have another effective applicant but the low firm does, which has a probability $e^{-\mu_2^*}(1-e^{-\mu_1^*})$. Again social and private benefits coincide. Note that the marginal benefit is independent of $\boldsymbol{\gamma}$, and therefore the decisions of other workers summarized in $\boldsymbol{\gamma}$ provide no externality on other workers. This is due to the fact that any positive externality on firms is dissipated in free entry, and the entry compensates any negative effects on other workers.[25]

Let $v(\boldsymbol{\gamma})$ be the entry for a given vector $\boldsymbol{\gamma}$ of applications as implied by equilibrium conditions 1a), 1b), 2a) and 3. Then $M^{**}(\boldsymbol{\gamma}) = M^*(v(\boldsymbol{\gamma}), \boldsymbol{\gamma})$ denotes optimal number of matches for a given vector $\boldsymbol{\gamma}$ of applications given optimal entry and optimal application decisions.

Proposition 9 (Efficiency of Search Intensity and Overall Efficiency)
The equilibrium vector $\boldsymbol{\gamma}$ of applications is constrained efficient, i.e. $\boldsymbol{\gamma} \in \arg\max_{\boldsymbol{\gamma}' \in \triangle_3} M^{**}(\boldsymbol{\gamma}') - Kv(\boldsymbol{\gamma}') - \gamma_1' c(1) - \gamma_2' c(2)$.

Proof: For a given $\boldsymbol{\gamma}$ we know that equilibrium conditions 1a), 1b), 2a) and 3 yield the optimal entry and the optimal number of matches. Moreover, under these conditions firms always receive zero profits and all surplus accrues to workers. Comparing different $\boldsymbol{\gamma}$, it is immediate that each worker always attains a marginal utility of $u_1^* - c_1$ for his first application, and $u_2^* - u_1^* - c_2$ for his second application. Clearly the equilibrium conditions in Proposition 4 specify the socially optimal entry. For the case where $c_1 = u_1^*$ (or $c_2 = u_2^* - u_1^*$) the privat and social benefits of the first (or second) application are zero, and therefore every equilibrium for this case is constrained efficient. *Q.E.D.*

5 Generalization to $N > 2$ applications

In this section consider the case where workers can send any number $i \in \mathbb{N}$ of applications, at a cost $c(i)$. We retain the assumption that $c(0) = 0$ and that marginal costs $c_i = c(i) - c(i-1)$ are weakly increasing. We also assume $c(i) > 0$ for some $i \in \mathbb{N}$. We will establish existence, (generically) uniqueness and constrained efficiency of the equilibrium. The analysis in the preceding sections is a special case for $c(i) = \infty$ for all $i > 2$. Since many arguments are straightforward generalizations of that special case, we focus mainly on the changes that are necessary to adapt the prior setup. At the end of this section we show convergence to the outcome of a competitive economy when the costs for applications vanish.

5.1 Extended Setup and Main Result

The extension mainly requires adaptations of the workers' setup, while it remains essentially unchanged for firms. Define N as the largest integer such that $c(N) \leq 1$. Clearly, it is neither

[25] This argument applies equilibrium conditions 1a), 1b), 2a) and 3 for different $\boldsymbol{\gamma}$, i.e. we compare partial equilibria for different $\boldsymbol{\gamma}$.

individually nor socially optimal to send more than N applications. Then the workers strategy is a tuple $(\boldsymbol{\gamma}, \boldsymbol{G})$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, ..., \gamma_N) \in \triangle_N$ and $\boldsymbol{G} = (G^1, G^2, ..., G^N) \in \times_{i=1}^N \Upsilon^i$. γ_i denotes the probability of sending i applications, and G^i denotes the cumulative distribution function over $[0,1]^i$ that describes to which wages the applications are sent. Let again $(w_1, ..., w_i)$ satisfy $w_1 \leq w_2 \leq ... \leq w_i$ and let G^i_j denote the marginal distribution of G^i over w_j. Then we can define $\mathcal{W}$ as the union of the support of all G^i_j with $\gamma_i > 0$. A worker who applies to $(w_1, ..., w_i)$ attains in analogy to (2) the utility

$$U_i(w_1, ..., w_i) = \sum_{j=1}^{i} \left[\prod_{k=j+1}^{i} (1 - p(w_k)) \right] p(w_j) w_j - c(i). \tag{2'}$$

A worker who applies nowhere attains $U_0 = 0$. Instead of (3) the relevant condition is now

$$\sum_{i=1}^{N} \left[\gamma_i \sum_{j=1}^{i} G^i_j(w) \right] = v \int_0^w \lambda(\tilde{w}) \; dF(\tilde{w}). \tag{3'}$$

To specify $\psi(w)$ in the extended setup, consider a firm at wage w that receives an application and let $\hat{G}(\tilde{\mathbf{w}}|w)$ denote probability that the sender applied with his other $N-1$ applications to wages weakly below $\tilde{\mathbf{w}}$. If the sender only sent $i < N-1$ other applications, then we code (only for this definition) the additional $N-1-i$ applications as going to wage -1. So $\tilde{\mathbf{w}} = (\tilde{w}_1, ..., \tilde{w}_{N-1}) \in ([0,1] \cup \{-1\})^{N-1}$. Let $h(\tilde{\mathbf{w}}|w)$ count the number of applications sent to wage w when the worker applies to $\tilde{\mathbf{w}}$ and w. Replacing (7) we now specify

$$\psi(w) = \int \left[1 - \frac{1 - (1 - p(w))^{h(\tilde{\mathbf{w}}|w)}}{p(w) h(\tilde{\mathbf{w}}|w)} \prod_{\tilde{w}_j > w} [1 - p(\tilde{w}_j)] \right] d\hat{G}(\tilde{\mathbf{w}}|w). \tag{7'}$$

The product $\prod_{\tilde{w}_j > w} [1 - p(\tilde{w}_j)]$ describes the probability that the applicant will not take a job at a strictly better wage. Its multiplier gives the probability that a worker will not turn down a job offer because of a job at another firm with the same wage, conditional on failing at higher wages (see the appendix for a derivation). Then the integrand gives the probability that the worker takes the job at a different firm, which is integrated over the relevant wages to which workers apply.

The definitions for all other variables, i.e. μ, p and η and π remain unchanged. The definition of the Market Utility Assumption now has to take into account the expanded possibilities of workers. Let $U^*_i = \sup_{\boldsymbol{w} \in \mathcal{V}^i} U_i(\boldsymbol{w})$. Then $U^* = \max_{i \in \{0,...,N\}} U^*_i$ denotes the Market Utility. Let $X_i(w) \subset [0,1]^i$ denote the set of i-tuples $(w_1, ..., w_i)$ with $w_j = w$ for some $j \in \{0, ..., i\}$ and

$w_k \in \mathcal{V}$ for all $k \neq j$. Then we can define $\hat{U}_i(w) = \sup_{\boldsymbol{w} \in X_i(w)} U_i(\boldsymbol{w})$ as the optimal utility if the worker applies to wage w and to $i-1$ other offered wages. The Market Utility assumption then states that for $w \notin \mathcal{V} \bigcup \mathcal{W}$ we have $\mu(w) > 0$ if and only if $U^* = \max_{i \in \{1,\ldots,N\}} \hat{U}_i(w)$. With these adjustments the equilibrium definition extends to this section.

We are now in the position to extend the result from the previous section. Again we recursively define μ_i^* and u_i^* as functions of the exogenous parameter K. Let $u_0^* = 0$. For all $i \in \mathbb{N}$ let $(1 - e^{-\mu_i^*} - \mu_i^* e^{-\mu_i^*})(1 - u_{i-1}^*) = K$ and $u_i^* = e^{-\mu_i^*}(1 - u_{i-1}^*) + u_{i-1}^*$. Note that $u_i^* - u_{i-1}^*$ is strictly decreasing in i, while c_i is weakly increasing. We will show

Proposition 10 (Generalized Equilibrium Properties) *An equilibrium exists. It is constrained efficient. Generically it is unique: if $c_{i^*} < u_{i^*}^* - u_{i^*-1}^*$ and $c_{i^*+1} > u_{i^*+1}^* - u_{i^*}^*$, every workers sends i^* applications, i^* wages will be offered, and every worker applies to each wage.*

The proof relies essentially on an induction of the arguments presented in Sections 3 and 4 to higher numbers of applications. The workers again partition the wages into intervals relevant to each of their applications. The equilibrium interaction in each interval corresponds to that in the one application case, again with the adjustment that the workers "outside option" incorporates the expected utility that can be obtained at lower wages, while the queue length incorporates the fact that some applicants are lost to higher wage firms. Efficiency obtains again for similar reasons, only that now i^* wages are necessary to obtain the optimal allocation in the search process.

5.2 Convergence to the Competitive Outcome

We will now show that the equilibrium allocation converges to the unconstrained efficient allocation of a competitive economy when application costs become small. For $K < 1$ the competitive outcome has an equal number of workers and active firms, i.e. $v = 1$, each active firm and each worker is matched, and the market wage is $1 - K$ and coincides with the utility of each worker.

We will consider a sequence of cost functions such that the marginal cost of the i'th application converges to zero for all $i \in \mathbb{N}$. Rather than looking at these functions directly, it will be convenient to simply consider the associated equilibrium number i^* of applications that each worker sends.[26] Vanishing costs amounts to $i^* \to \infty$. Let $v(i^*)$ denote the equilibrium measure of active firms, $\eta(i^*) = \int \eta(w) dF$ and $\varrho(i^*) = v(i^*)\eta(i^*)$ the average probability of being

[26] For the case of multiple equilibria, consider for simplicity the case where all workers send the same number of applications.

matched for a firm and a worker in the economy. Let $w(i^*)$ denote the average wage conditional on being matched and $U^*(i^*) = u^*_{i^*} - c^{i^*}(i^*)$ the equilibrium utility when i^* applications are sent, where $c^{i^*}(\cdot)$ denotes some cost function that supports an equilibrium with i^* applications per worker. We will show that

Proposition 11 (Convergence) *The equilibrium outcome converges to the competitive outcome, i.e.,* $\lim_{i^*\to\infty} v(i^*) = 1$*,* $\lim_{i^*\to\infty} \eta(i^*) = \lim_{i^*\to\infty} \varrho(i^*) = 1$ *and* $\lim_{i^*\to\infty} w(i^*) = \lim_{i^*\to\infty} U^*(i^*) = 1 - K$*.*

The structure of the proof uses the intuition for the competitive economy: For a given measure v of active firms the competitive economy implies that (only) the long side of the market gets rationed and the short side appropriates all surplus. We will show that for small frictions (i^* large) this still holds approximately. Then it trivially follows that $v(i^*) \to 1$ because otherwise the firms either generate to much or too little profits to cover entry. Since nearly all agents get matched, zero profits imply a wage of $1 - K$. Despite the fact that workers send more applications, the costs vanish faster than the increase in the number of applications and therefore their utility equals the wage of $1 - K$ in the limit.

6 Discussion and Outlook

6.1 Discussion of Main Assumptions

The paper builds on a micro-foundation of frictional markets based on coordination problems between agents. One underlying assumption is that firms treat similar workers alike and do not condition on applicants' names, which seems plausible in larger markets. The other assumption is that workers cannot coordinate to each apply to a different firm. This is modeled by the requirement that in equilibrium workers use symmetric strategies, which has the advantage that a worker does not need to know his "role" in the application process and can deploy the same strategy as everybody else. In particular, this does not require each individual to know each other player's individual strategy, but he just plays against the population-wide strategy of the other workers. That the induced coordination problems actually arise even in small groups has been shown experimentally by Ochs (1990) and Cason and Noussair (2003). A deeper discussion can be found e.g. in Shimer (2005).[27]

[27] We have also assumed anonymity of the workers application strategy. This is merely for notational convenience. It immediately implies identical queue lengths at firms that post the same wage. This can be derived as a result following from anonymous strategies by firms and symmetric strategies by workers. See e.g. Shimer

When workers can choose their search intensity by sending multiple applications, new modeling choices arise that are absent in one-application models. After the application stage every worker is "linked" to the multiple firms to which he applied. A firm might be "linked" to multiple workers who applied to it (other firms might only have one worker or no worker at all). Since each firm can only hire one worker, and each worker can only work for one firm, any simultaneous application model has to specify how matches between firms and workers are formed given those links. It also has to specify the division of surplus for a given match.

For the allocation the main novelty in this paper is to consider a stable assignment on the network. We have interpreted this as the outcome of sequential process in which firms contact additional applicants when their offer is rejected. One might think about firms calling up applicants sequentially on the phone. The important feature in this interpretation is that workers can reconsider their options when they receive a better offer in a later phone call. Otherwise they would have to be strategic in which offers to accept and would not necessarily wait for the best offer that comes along. If workers can switch when a better offer is made, a firm with a higher wage can always attract an applicant away from a lower wage firm by making him an offer, which lends justification to our stability notion that no firm remains vacant while one of its applicants starts to work for a lower wage firm.

For the division of the surplus this paper assumes wage commitments. That is, the wage paid in the final period is the posted wage, and firms cannot counter the proposals of other firms. This again has the main advantage of keeping the matching process tractable, as it leads to a simple stability notion in the assignment problem. In general it is not clear in which way wages would adjust if they could be negotiated even ex post. On the one hand firms loose the incentive to pay high wages once the workers have already applied, especially if many workers applied to the firm. On the other hand the worker might have other offers, in which case there is an incentive for higher wages. The inability to reduce the wage even when there are many applicants might have to do with reputation considerations. The inability to increase the wage when the worker has other options might be due to the fact that other offers are difficult to observe or verify. If a verifiable offer triggers a bidding war for the worker's services, every firm would try to make its own offer in a way that does not allow the worker to show it to some other firm (while it would like the other firms to make their offer verifiable so that it can counter them). Other reasons that preclude counteroffers might be company (or university) rules do not allow more than the allocated budget for hiring decisions.

A final case for the assumption of wage commitments arises in environments in which market

(2005) for a derivation in an infinite economy model. Burdett, Shi and Wright (2001) demonstrate this property nicely in a finite economy model.

rules entail binding job and wage descriptions ahead of the final matching. One example is the market for medical interns in the US. Roth (1984, p. 995-996, footnotes omitted) describes the procedure that governs this market as follows:

> Under this procedure, students and hospitals would continue to make contact and exchange information as before. (It is worth noting in this regard that the complete job description offered by a hospital program in a given year was customarily specified in advance, see, e.g., Stalnaker [1953]. Thus, the responsibilities, salary, ect., associated with a given internship, while they might be adjusted for year to year in response to a hospital's experience in the previous year's market, were not subject of negotiation with individual candidates.) Students would then rank in order of preference the hospital programs to which they had applied, hospitals would similarly rank their applicants, and all parties would submit these rankings to a central bureau, which would use this information to arrange a matching of students to hospitals and inform the parties of the result.

The algorithm used by the central bureau coincides with the Deferred Acceptance Algorithm by Gale and Shapley (1962) with hospitals as proposers, on which the allocation in this paper is based. Our model captures some of the elements of this procedure, mainly that interns and hospitals are only matched if the intern applied to the hospital and that the attractiveness of the job in terms of job description and salary is specified in advance.[28]

6.2 Relation to the Literature

The idea to combine competition though wage (or price) announcements with frictions induced by coordination failures in the application process goes back to Peters (1984, 1991). The idea that the competition induces constrained efficient market outcomes has already been noted by Hosios (1990) and Montgomery (1991); see also the introduction. An instructive presentation of the equilibrium consideration for a finite economy is presented in Burdett, Shi and Wright (2001). Our model relates to these single-application models, because even with multiple applications the equilibrium interaction within each segment induced by the workers' response resembles the interaction in the one application model, with some adjustments for the spillovers of one segment onto the other (see Section 3.2).

Other directed search models that allow for multiple applications are Albrecht, Gautier and Vroman (2006) and Galenianos and Kircher (2006). Both models consider a fixed number of applications per worker for their analytical derivations. The main difference arises in the allocation of workers to firms on the given network. Both models assume that each firm makes exactly one offer at random to one of their applicants, and if that offer is rejected the firm

[28]Extending our setup to allow for heterogeneity will be important to resemble this environment more closely.

remains vacant no matter how many other applications it received. This implies that search intensity can reduce efficiency because with more applications some workers attract multiple offers and preclude others from obtaining even the rejected jobs.[29]

Albrecht, Gautier and Vroman (2006) also differ in their assumption regarding the division of surplus. They assume that firms pay at least the posted wage, but engage in Bertrand competition if two of them offer a job to the same worker. They consider equilibria with a single posted wage, and show that the equilibrium wage offer equals the workers' outside option. Wages are zero in equilibrium because a higher posted wage does not yield any advantage if the worker gets two offers. While paid wages are in part higher due to the Betrand competition, the low offered wage nevertheless leads to excessive entry. This remains even in an extension in which firms can make two offers.

Galenianos and Kircher (2006) is closer to the model considered here in assuming commitment to the offered wage. Firms make a single offer, and workers accept the highest one. Despite the difference in recall, the equilibrium interaction also features wage dispersion with the number of wages corresponding to the number of applications. While the worker's problem is quite different in the way that strategies of other agents translate into the relevant hiring probabilities, the structural trade-offs for each worker are in fact similar, implying the separation property (Lemma 1) in both models.[30] This suggests a robustness of the equilibrium structure to the specifics of the recall process. The search process in Galenianos and Kircher (2006) is inefficient because wage dispersion leads to different hiring probabilities, but efficient search would equalize the hiring probabilities. That is, one wage would be optimal.[31]

The way workers and firms match after the applications are sent crucially affects the ability of the market to achieve constrained efficiency. As explained in Section 4, constrained efficiency arises in this paper because workers – as well as firms – care about the effective queue length, and therefore they adjust it in response to the wage announcements. Without recall, firms and workers care about different things. Firms are interested in effective applications. But workers now care about all applications, no matter if these are effective or non-effective: If another worker applies and gets an offer, the job is lost even if that worker takes a better position. If a

[29]This effect is present when workers apply to many firms, see Albrecht, Gautier and Vroman (2006) for a discussion of this effect. The number of matches would be improved if workers concentrated some of their application on a zero measure of firms, i.e. if they would dispose of them.

[30]Delacroix and Shi (2005) and Shi (2006) obtain similar equilibrium properties in continuous time directed search models with on-the-job search. In those models the current job acts as a strictly positive default for further applications in a similar way that the low wage application in the present model acts as a default for the high wage application. The efficiency properties of these on-the-job search models have not yet been investigated.

[31]See footnote 24 for the difference in terms of search efficiency.

firm raises the wage (keeping the other wages constant), workers adjust their applications such that the gross queue length reaches a new level of indifference. Therefore, the congestion by workers who end up with better jobs allows firms only "price" to price the gross queue length, but not the effective queue length that they really care about. A higher wage leads to more applications, but potentially from workers that apply to other firms where it is easier to get jobs. The queue of effective applications may therefore in fact decrease.

The model considered here shows that strategic search intensity in the form of multiple applications is not per se at odds with efficiency in directed search markets. If the communication of firms with all their workers leads to a stable assignment, the more efficient allocation on the given network translates into efficient market interaction in earlier stages that determine entry, the number of applications, and where workers apply. As discussed in Section 4, it is the fact that not only firms but also the workers care about effective applications that allows firms to price the applications efficiently, as it eliminates the wedge between gross and effective applications. The issue of gross vs. effective applications has not attracted attention in models in which workers can only apply to a single firm since workers never decline an offer as they have no alternative by assumption. Therefore, in single-application models every (gross) application is effective by construction, and efficiency obtains.

The specification of the matching outcome is based on the idea of stable assignments pioneered by Gale and Shapley (1962). In our setting we restrict stability to firms and workers that are linked to one another in the application process, and assume that wage postings are commitments. Bulow and Levin (2006) also study non-cooperative wage commitments prior to non-transferable utility matching. Like most papers on the marriage market, they allow for heterogeneities but do not consider limited "links". Exceptions are Roth and Perason (1999) and Immorlica and Mahdian (2005), who consider a random network of links. The present paper neglects heterogeneities other than through wages, but treats link formation as an active choice.

6.3 Conclusion

This paper incorporates a micro-foundation for search intensity into a directed search framework. Directed search can here be interpreted as strategic but frictional link formation between workers and firms. Search intensity can be viewed as a choice on the number of links that the worker wants to obtain. We consider a stable allocation on the network, resembling a process in which firms contact ("recall") additional applicants if their offer gets rejected. Firms' wage announcements price the network efficiently, given the workers' coordination problem. Equilibrium wage dispersion turns out to be the optimal response of the market to the presence of frictions, since it allows for a network structure that minimizes coordination failure. While other work has shown that multiple applications lead to inefficiencies in a directed search setting when recall is restricted, we show that constrained efficiency prevails in the presence of strategic search intensity when recall is allowed.

While we focused our attention on efficiency properties, the relative ease with which search intensity and recall can be incorporated suggests that the model can be applied to answer wider questions. In a first step we considered the connection between productivity and search intensity: Lower search costs (or equivalently higher productivity) imply more search and in the limit the equilibrium outcome approximates the unconstrained efficient competitive allocation. Additional interesting questions concern the interaction between simultaneous and sequential search in a repeated labor market interaction.[32] We expect simultaneous search to dominate in markets with long time-frames between applications and final hiring decisions. The introduction of heterogeneity is also left for future research. If only firms are heterogeneous in terms of productivity, it seems likely that higher productivity firms will offer higher wages, and we expect firms' wage offers to be clustered with the number of clusters equal to the number of applications (similar to the mass points in this analysis). For two-sided heterogeneity the matching process will have to be adapted to account for the firms preferences over different workers.

7 Appendix

Proof of Proposition 1:

For wages strictly below u_1 the result is immediate because the Market Utility cannot be obtained. At wage $w = u_1$, $\mu(w) > 0$ would imply that the Market Utility cannot be reached. Wages strictly above u_1 have $\mu(w) > 0$, as otherwise $p(w)w = w > u_1$ and workers would receive more than the Market Utility when applying there.

We have shown that it is optimal to send low applications to wages below $\bar{w}$, which implies that (9) has to hold for all wages in $(u_1, \bar{w}]$ in order to provide the Market Utility. $u_1 \equiv \sup_{w \in \mathcal{V}} p(w)w$ since the optimum has to be obtained at some wage that is actually offered.

For $\gamma_2 > 0$, it is optimal to apply with the high application to wages above $\bar{w}$, and the effective queue length is therefore governed by (10). Again the optimum is attained for wages that are actually offered. The effective queue length has to be continuous at $\bar{w}$, as otherwise the job finding probability $p(w)$ for workers would be discontinuous and some wage in the

[32] Using an approach similar to the extension in Galenianos and Kircher (2005) one can show that with exogenous separations the steady-state of a repeated interaction is similar to the one-shot interaction analyzed here. Only the workers' outside option of not being matched is then strictly positive.

neighborhood of $\bar{w}$ would offer a utility different from the Market Utility. Therefore $\bar{w}$ is determined as the wage where both (9) and (10) hold.

For $\gamma_2 = 0$, all wages above u_1 that are in the offer set $\mathcal{V}$ have to conform to (9) because they receive single applications. So $\bar{w} \geq \sup \mathcal{V}$. But if a higher (not offered) wage would be offered, workers might prefer to send a second application there rather than relocating their first one. Assume the queue length would be governed by (9) for all wages in $(u_1, 1]$. If $p(w)w + (1 - p(w))u_1 - c_2 \geq u_1$, workers would like to send a second application. Therefore $\bar{w}$ is the smallest wage for which that inequality holds (but at most 1). At higher wages the inequality would be strict, i.e. workers would get a utility higher than the Market Utility by sending a second application. To fulfill the Market Utility Assumption the additional utility of the second application has to equal its cost, so $u_2 - u_1 = c_2$ has to hold at high wages and the effective queue length is again governed by (10). *Q.E.D.*

Proof of Proposition 2:

Consider a (candidate) equilibrium in which all active firms offer wage $w^* \in (0, 1)$. Almost all applications are sent to w^* because of (3) and equilibrium condition $2a$) on worker optimality. $w^* > 0$ then implies $u_1 = p(w^*)w^* > 0$. Moreover $w^* = \bar{w}$. If not, i.e. $\bar{w} > w^*$ or $\bar{w} < w^*$, then a mass of applications would be sent strictly above or below the offered wage, yielding a contradiction. Then profits for wages above w^* are given by (15), for wages in $[p(w^*)w^*, w^*]$ by (14), and for wages below $p(w^*)w^*$ profits are zero.

The left derivative of the profits with respect to the queue length at $\bar{\mu} = \mu(w^*)$ is obtained by the differentiating (14) to get $\pi'_-(\bar{\mu}) = e^{-\bar{\mu}} - u_1$, and the right derivative by differentiating (15) which yields $\pi'_+(\bar{\mu}) = e^{-\bar{\mu}}(1-u_1)-(u_2-u_1)$. In equilibrium it needs to hold that firms will neither want to increase their wage nor decrease their wage. This leads to $\pi'_+(\bar{\mu}) \leq 0 \leq \pi'_-(\bar{\mu})$. But $\pi'_+(\bar{\mu}) \leq \pi'_-(\bar{\mu})$ implies

$$-e^{-\bar{\mu}}u_1 - (u_2 - u_1) \leq -u_1. \tag{30}$$

For a single market wage it holds that $u_2 = u_1 + (1 - \bar{p})u_1$ with $\bar{p} = \frac{1-e^{-\bar{\mu}}}{\bar{\mu}}$. We can therefore write $u_2 - u_1 = (1 - \bar{p})u_1$. Then (30) reduces to

$$(1 - e^{-\bar{\mu}} - \bar{\mu}e^{-\bar{\mu}})u_1 \leq 0. \tag{31}$$

We know that $u_1 > 0$. It is easily shown that the term in brackets is strictly positive for any $\bar{\mu} > 0$, yielding the desired contradiction.

For the extremes, consider $w^* = 1$ first. At $w^* = 1$ firms make zero profits. Since the effective queue length at wages close to 1 is positive by (12), wages below one provide profitable deviations. Now consider $w^* = 0$. Equilibrium profits are strictly smaller than one because not all firms get matched. (13) implies that at wages $w' > 0$ firms can hire for sure, i.e. the effective queue length at wages above zero is infinity. Therefore, small increases in the wage are profitable. *Q.E.D.*

Proof of Lemma 2, part 1:

Instead of equations (22) and (23), we now have

$$1-e^{-\mu_1}-\mu_1 e^{-\mu_1} \quad = \quad \hat{\pi}, \tag{32}$$

$$(1-e^{-\mu_2}-\mu_2 e^{-\mu_2})(1-e^{-\mu_1}) \quad = \quad \hat{\pi}, \tag{33}$$

for some endogenous profit $\hat{\pi}$. Consider $\hat{\pi}$ as a free parameter. For a given $\hat{\pi}$ (32) and (33) uniquely determine the measure $\hat{v}_1$ and $\hat{v}_2$ of firms in the low and high group. That is, $\hat{\pi}$ is supported by a unique measure $\hat{v} = \hat{v}_1 + \hat{v}_2$ of firms. We want to show that there is only a single $\hat{\pi}$ that is supported by $\hat{v} = v$, which then establishes uniqueness v_1 and v_2 (and thus of F as in part 2). By (32) μ_1 strictly increases in $\hat{\pi}$. Equal profits at high and low wage firms implies

$$1-e^{-\mu_2}-\mu_2 e^{-\mu_2} = 1-\mu_1 e^{-\mu_1}/(1-e^{-\mu_1}), \tag{34}$$

which implies that μ_2 is strictly increasing in $\hat{\pi}$, since $\mu_1 e^{-\mu_1}/(1-e^{-\mu_1})$ is strictly decreasing in μ_1. Since $\mu_2 = \gamma_2/\hat{v}_2$, $\hat{v}_2$ is strictly decreasing in $\hat{\pi}$. We have proven the lemma if we can show that also $\hat{v}_1 + \hat{v}_2$ is decreasing in $\hat{\pi}$. Since $\mu_1 = (\gamma_1 + \gamma_2 - \gamma_2 p_2)/\hat{v}_1$ we get $\partial\mu_1/\partial\hat{\pi} = -[\mu_1/\hat{v}_1][\partial\hat{v}_1/\partial\hat{\pi}] - (1/\hat{v}_1)(1-e^{-\mu_2}-\mu_2 e^{-\mu_2})[\partial\hat{v}_2/\partial\hat{\pi}]$. By the prior argument this derivative has to be strictly positive, which together with $\mu_1 > 1-e^{-\mu_2}-\mu_2 e^{-\mu_2}$ implies $\partial\hat{v}_1/\partial\hat{\pi} + \partial\hat{v}_2/\partial\hat{\pi} < 0$. $\mu_1 > 1-e^{-\mu_2}-\mu_2 e^{-\mu_2}$ holds because it is by (34) equivalent to $\mu_1 > 1-\mu_1 e^{-\mu_1}/(1-e^{-\mu_1})$, which is equivalent to $1 > (1-e^{-\mu_1})/\mu_1$. The latter is true for all $\mu_1 > 0$. Since for $\hat{\pi} \to 0$ we have $\hat{v} \to \infty$ and for $\hat{\pi} \to 1$ we have $\hat{v} \to 0$, there is exactly one $\hat{\pi}$ supported by a measure $\hat{v} = v$ of firms. *Q.E.D.*

Proof of Proposition 6:

For $\gamma_0 = 1$ or $v = 0$, the result is trivial as matches are always zero. When workers send one application ($\gamma_2 = 0$), one group of firms with equal hiring probability is optimal because of strict concavity of the matching probability $1-e^{-\lambda}$ (this is a special case of Shimer (2005)).

For $\gamma_2 > 0$ we will use a variational argument to prove that two groups of firms of which one receives all high applications and the other all low applications will be sufficient to achieve the same number of matches as any other optimal wage setting and application behavior $\{F, \boldsymbol{G}\}$. The proof is structured into two parts. First, we show that we can pointwise decompose the application process into two groups without changing the overall matches in the economy. Second, we show that across wages the optimal allocation is random within each group. Take some optimal $\{F, \boldsymbol{G}\}$ as a starting point.

Part 1: We use a pointwise argument for each $w \in \mathcal{V} \cup \mathcal{W}$ – other wages do not contribute to the matching. For technical convenience we have assumed that each wage in $\mathcal{V} \cup \mathcal{W}$ entails a continuum of offers and applications (footnote 9) in order to apply the law of large number convention per wage. All firms at this wage face the same effective queue length $\mu(w)$. We will split the firms at this wage into two *subgroups*, and reshuffle the application behavior of the workers that send applications to this wage, such that their high application is randomly sent to some firm in the first and their low applications to some firm in the second subgroup. Workers that send both applications to w send one to each group, and accept offers from the first over offers from the second. We leave the application behavior towards other wages unchanged.

We will show that for an appropriate choice of the relative size of the subgroups the overall matching is unchanged. Let $\lambda_h(w)$ denote the ratio of workers that only send their high application to wage w to firms offering w under $\{F, \boldsymbol{G}\}$. Let $\lambda_b(w)$ denote the worker/firm ratio for workers that send both applications to w. Let $(1-\bar{\psi})\lambda_l(w)$ be the ratio of workers who send their low or single application to w and do not get a strictly better offer to firms offering that wage. If $\{F, \boldsymbol{G}\}$ is optimal, then neither of these ratios is infinity, and not all of them are zero (except possibly for some wages that attract a zero measure of agents, which we can neglect without loss of optimality). If only $\lambda_h(w)$ (respectively $(1-\bar{\psi})\lambda_l(w)$) is strictly positive, then we can trivially avoid to change the matching by having a zero fraction of the firms in the second (resp. first) subgroup. Therefore consider the case where at least two of the ratios are strictly positive.

We will show that we can equalize the effective queue lengths for both subgroups by adjusting the size of firms in each group appropriately. Let the fraction of firms in the first subgroup be d. Then the effective queue length for these firms is $\mu_h(d) = \frac{\lambda_h(w)+\lambda_m(w)/2}{d}$, because all applications are effective. For those firms in the second group it is $\mu_l(d) = \frac{(1-\bar{\psi})\lambda_l(w)}{1-d} + \frac{1}{2}\frac{1-e^{-\mu_h(d)}}{\mu_h(d)}\frac{\lambda_b(w)}{1-d}$. For d close to zero $\mu_h(d) > \mu_l(d)$, while for d close to 1 $\mu_h(d) < \mu_l(d)$. By the intermediate value theorem it is possible to equalize both at some $d(w)$.

Note that if the two subgroups of firms face some identical effective queue length μ', then $\mu' = \mu(w)$. Assume note, i.e. $\mu' > \mu(w)$. That means that under μ' strictly more firms then under $\mu(w)$ get matched at wage w. On the other hand it becomes strictly harder for workers to get an offer, and since we did not change the application behavior at other wages, strictly less workers get matched at wage w. Since workers and firms are matched in pairs, this yields the desired contradiction. Similarly $\mu' < \mu(w)$ can be ruled out. Since $\mu' = \mu(w)$, we have not changed the overall matching when we assign $d(w)$ firms to the first subgroup.

Part 2: Repeating the step in part 1 pointwise for each w leaves us with a set of low application subgroups and a set of high application subgroups. The next step is to show that random applications within each set does not reduce the number of matches.

Consider first the set of subgroups whose firms receive low (or single) applications. Consider two of these subgroups, one with matching probability $1-e^{-(1-\psi)\lambda}$, and one with $1-e^{-(1-\psi')\lambda'}$. From the larger of the subgroups select a subset of firms with equal size to the smaller subgroup. We will show that in an optimal allocation both have the same effective queue length by shifting firms from one group to the other while leaving the applications that each group receives the same. Let d be the fraction of firms in the first subgroup, and γ and γ' the gross queue length per group. Then the average matchings across both groups is given by

$$d(1-e^{-(1-\psi)\frac{\gamma}{\nu d}}) + (1-d)(1-e^{-(1-\psi')\frac{\gamma'}{\nu(1-d)}}). \tag{35}$$

Since both subgroups have a strictly positive effective queue length, it cannot be optimal to place all firms in only one subgroup (as otherwise few firms placed in the other would be matched nearly for certain). Therefore, to achieve optimal matching d is characterized by the first order condition

$$\nu[(1-e^{-\mu})-(1-e^{-\mu'})-\mu e^{-\mu}+\mu' e^{-\mu'}]=0, \tag{36}$$

where $\mu=(1-\psi)\frac{\gamma}{\nu d}$ and $\mu'=(1-\psi')\frac{\gamma'}{\nu(1-d)}$. Since $1-e^{-\mu}-\mu e^{-\mu}$ is strictly increasing in μ (and similar for μ'), we have $\mu=\mu'$ in the optimal allocation of firms.

Repeating this argument for all sets shows that almost all low or single application firms have to have the same effective queue length in an optimal allocation. This queue length could still result from different ψ and λ combinations. Yet if we let workers randomly send their low or single applications to all firms in low application subgroups, all firms will face an identical ψ and λ. A similar argument as in the last paragraph in Part 1 shows that this does not change the number of matches. Therefore the set of low application firms corresponds to the non-preferred group in the proof for two-group efficiency.

By this construction, for low and single application firms only the *average* matching probability at high application firms matters. If we keep the size of low and single application firms constant and leave the gross queue length for them unchanged, but match more workers already at high wage firms, this clearly improves the matching (despite some negative externality on the low or single application firms).

Therefore, we consider next the set of high application subgroups and maximize the number of matches there. By the strict concavity of $1-e^{-\lambda}$ the average matching probability at high wage firms is maximized if the gross queue length (and thus the effective queue length) is identical for all of them. This result is well known, see e.g. Shimer (2005). The number of matches is again identical to the case in which workers randomly send their high applications to these firms; and this set of firms corresponds to the preferred group in the proof for two-group efficiency.

By the two-group-efficiency of the equilibrium matching, the equilibrium matching is constrained efficient. *Q.E.D.*

Proof of Proposition 7:

Given v and $\boldsymbol{\gamma}$ with $\gamma_2>0$, consider two tuples $\{F',\boldsymbol{G}'\}$ and $\{F'',\boldsymbol{G}''\}$ that lead to equal hiring probabilities η' respectively η'' for all firms. Similar to the argument in the proof of Proposition 6 $\eta'=\eta''=\bar{\eta}$, since otherwise one tuple would match more workers but fewer firms than the other.

We again split the firms into two groups, called first and second, all workers send their high application to the second and their low or single application to the first, and accept offers from the second over those from the first. Let d be the fraction of firms in the second group. Again $\mu_2=\lambda_2=\gamma_2/(vd)$ and $\mu_1=[1-\gamma_2 p_2/(\gamma_1+\gamma_2)]\lambda_1=[\gamma_1+\gamma_2-\gamma_2(1-e^{-\lambda_2})/\lambda_2]/(v(1-d))$. Since for $d\approx 0$ clearly $\mu_2>\mu_1$ and for $d\approx 1$ $\mu_2<\mu_1$, there exists a $\hat{d}$ such that effective queue length and thus the hiring probability of both groups is equalized. It is easy to show that $\mu_1-\mu_2$ is strictly increasing in d around $\mu_2\approx\mu_1$, so that $\hat{d}$ is unique. This two-group process has $\mu_1=\mu_2$, but the optimal two group process fullfils (27), which requires $\mu_1<\mu_2$, i.e. a strictly smaller preferred group. *Q.E.D.*

Explanation to equation (7′):

Consider a particular worker who applies to h firms that offer wage w. For simplicity assume he applies to no other firms. Conditional on the fact that a firm makes an offer to the worker sometime during the recall process, we want to determine the probability σ that this firm hires the worker. The unconditional probability that the worker does not get any offer is $(1-p(w))^h$, and so the unconditional probability that he gets an offer is $\varsigma = 1-(1-p(w))^h$. The unconditional probability of getting an offer from any specific firm is ς/h. This unconditional probability can also be written as the probability of hiring conditional on making an offer multiplied by the unconditional probability of making an offer. So we have $\sigma p(w) = \varsigma/h$, or $\sigma = [1-(1-p(w))^h]/[hp(w)]$, which explains the formula used for equation (7′). The argument is based on insights from Burdett, Shi and Wright (2001).

Proof of Proposition 10:

We start out by fixing $\boldsymbol{\gamma}$ and denote by $\hat{\imath}$ the highest integer for which $\gamma_i > 0$; i.e. $\hat{\imath}$ is the maximum number of applications that workers send. By straightforward extension of the analysis of the workers' best response in Section 3 it can be established that the workers' best response to a wage offer distribution is now given by $\hat{\imath}$ intervals such that every worker sends exactly one application to each interval. More specific, let the utility from sending the first i applications optimally be defined recursively by $u_i \equiv \max_{w\in[0,1]} p(w)w + (1-p(w))u_{i-1}$ for all $i \in \{1,2,...\hat{\imath}\}$, with $u_0 \equiv 0$. Then for any wage offer distribution the effective queue length is characterized by $(u_1,..,u_n,\bar{w}_0,...,\bar{w}_N)$ such that

$$
\begin{aligned}
p(w) &= 1 \qquad \forall w \in [0,\bar{w}_0],\ \textit{and} && (37)\\
p(w)w + (1-p(w))u_{i-1} &= u_i \ \ \forall w \in [\bar{w}_{i-1},\bar{w}_i]\ \ \forall i \in \{1,...,N\}, && (38)
\end{aligned}
$$

where $\bar{w}_0 = u_1$ and $\bar{w}_N = 1$. The indifference implies $\bar{w}_i = u_{i-1} + [u_i - u_{i-1}]^2/(2u_i - u_{i-1} - u_{i+1})$ for intermediate $i \in \{1,...,\hat{\imath}\}$. Clearly $\bar{w}_{\hat{\imath}} \geq \sup \mathcal{V}$, because any wages that are actually offered receive applications from workers that at most send $\hat{\imath}$ applications. At higher (not offered) wages, workers might start sending additional applications. The Market Utility Assumption implies that they cannot receive more than the Market Utility, which implies that $u_i - u_{i-1} = c_i$ for $i > \hat{\imath}$. The indifference then yields $\bar{w}_i = u_{i-1} + [u_i - u_{i-1}]^2/(u_i - u_{i-1} - c_{i+1})$. If this is in $[0,1]$ then this gives the appropriate boundary, otherwise $\bar{w}_i = 1$ and workers would strictly refrain from sending this many applications.

Using (38), we can rewrite the profit function for a firm who offers a wage $w \in [\bar{w}_{i-1},\bar{w}_i]$ with $\bar{w}_{i-1} < 1$ as

$$
\pi(\mu) = (1-e^{-\mu})(1-u_{i-1}) - \mu(u_i - u_{i-1}), \tag{39}
$$

where $\mu = \mu(w)$. The logic is similar to (15). If $\bar{w}_{i-1} = 1$ the profit is trivially zero. Proposition 2, stating that there exists no equilibrium in which only one wage is offered, can now easily be shown with similar techniques whenever $\gamma_i > 0$ for some $i > 1$. By a similar argument it is straightforward that at least i wages have to be offered in equilibrium whenever $\gamma_i > 0$. Given that (39) is strictly concave, it is also immediate that all firms within the same interval will offer the same wage, yielding exactly $\hat{\imath}$ wages when workers send at most $\hat{\imath}$ applications.

We call the group of firms that ends up offering the i'th highest wage as group i and will index all their variables accordingly. It will be convenient to denote by $\Gamma_i = \sum_{k=i}^{\hat{\imath}} \gamma_k$ the fraction of workers who apply to at least i firms. Then at wage i the probability of retaining an applicant is $(1-\psi_i) = \sum_{j=i}^{\hat{\imath}} \frac{\gamma_j}{\Gamma_i} [\prod_{k=i+1}^{j} (1-p_k)]$, since a fraction γ_j/Γ_i of applicants sends j applications and does not get a better job with probability $\prod_{k=i+1}^{j}(1-p_k)$. The effective queue length at wage i is given by $\mu_i = (1-\psi_i)\lambda_i$, where $\lambda_i = \Gamma_i / v_i$ is the gross queue length. For $i < \hat{\imath}$ the unique offered wage in $[\bar{w}_{i-1}, \bar{w}_i]$ is obtained by the first-order-conditions of (39), which are given by

$$u_i - u_{i-1} = e^{-\mu_i}(1 - u_{i-1}). \tag{40}$$

Therefore (39) can be rewritten as

$$\pi_i = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - u_{i-1}). \tag{41}$$

Free entry implies that $\pi_i = K$, which together with (40) implies that $\mu_i = \mu_i^*$ and $u_i = u_i^*$ as defined above. By a similar argument as for (22) and (23) the condition $\pi_i = K$ defines for a given vector $\boldsymbol{\gamma}$ of applications the unique measure v_i of firms in each group, and the wage $w_i = u_{i-1}^* + (u_i^* - u_{i-1}^*)/p_i$ that each group of firms offers (when $i < \hat{\imath}$). Note that the equal profit condition $\pi_i = \pi_{i-1}$ together with (40) implies

$$1 - e^{-\mu_i} - \mu_i e^{-\mu_i} = 1 - \frac{\mu_{i-1} e^{-\mu_{i-1}}}{1 - e^{-\mu_{i-1}}}, \tag{42}$$

which corresponds to (34).

To determine the equilibrium, only $\boldsymbol{\gamma}$ has to still be determined. Recall that i^* denotes the number of applications for which $c_{i^*} < u_{i^*}^* - u_{i^*-1}^*$ and $c_{i^*+1} > u_{i^*+1}^* - u_{i^*}^*$. Consider first the case where $\boldsymbol{\gamma}$ is such that $\hat{\imath} < i^*$. Since profits are determined by first order conditions, there cannot be any wage in $[0, \bar{w}_{\hat{\imath}}]$ that offers higher profits. Therefore a deviating firm has to consider a deviation within $(\bar{w}_{\hat{\imath}}, 1)$. In this region the effective queue length is (at least) given according to $p(w)w + (1 - p(w))u_{\hat{\imath}}^* = u_{\hat{\imath}}^* + c_{\hat{\imath}+1}$ (it may even be larger if workers send two or more additional applications to not-offered high wages). Therefore the profit for a deviating firm is (at least) $\pi(\hat{\mu}) = (1 - e^{-\hat{\mu}} - \hat{\mu}e^{-\hat{\mu}})(1 - u_{\hat{\imath}}^*)$, where $\hat{\mu}$ is given by the first order condition $c_{\hat{\imath}+1} = e^{-\hat{\mu}}(1 - u_{\hat{\imath}}^*)$. Since $c_{\hat{\imath}+1} < e^{-\mu_{\hat{\imath}+1}^*}(1 - u_{\hat{\imath}}^*)$, we have $\hat{\mu} > \mu_{\hat{\imath}}$. This implies $\pi(\hat{\mu}) > (1 - e^{-\mu_{\hat{\imath}+1}^*} - \mu_{\hat{\imath}+1}^* e^{-\mu_{\hat{\imath}+1}^*})(1 - u_{\hat{\imath}}^*) = K$, where the equality follows from the definition of $\mu_{\hat{\imath}+1}^*$. The optimal deviating wage can indeed be shown to lie above $\bar{w}_{\hat{\imath}}$ and therefore the deviation is profitable.

Clearly also $\hat{\imath} > i^*$ cannot be an equilibrium, because for workers who send $\hat{\imath}$ applications the marginal costs of the last application do not cover its marginal benefit.

For the case where $\hat{\imath} = i^*$ all workers want to send exactly i^* applications. We will show that $\bar{w}_{i^*} = 1$, which implies that firms do not have a profitable deviation, which establishes the existence and uniqueness result. If $\bar{w}_{i^*} = 1$, it means that the queue length in $[\bar{w}_{i^*-1}, 1]$ is determined by $p(w)w + (1 - p(w))u^*_{i^*-1} = u^*_{i^*}$. Note that is implies $p(1) = [u^*_{i^*} - u^*_{i^*-1}]/[1 - u^*_{i^*-1}] = e^{-\mu^*_{i^*}}$, where the second equality follows from the definition of $\mu^*_{i^*}$. Determining the effective queue length this way is in accordance with the Market Utility Assumption if and only if it is not profitable to send an additional application. If an additional application is sent, by a logic similar to Lemma 1 it is optimal to send it to the highest wage. The marginal benefit would be $p(1)+(1-p(1))u^*_{i^*} - u^*_{i^*}$, or $e^{-\mu^*_{i^*}}[1-u^*_{i^*}]$. Since $c_{i^*+1} < e^{-\mu^*_{i^*+1}}[1-u^*_{i^*}]$ and $\mu^*_{i^*+1} > \mu^*_{i^*}$ it is not profitable for workers to send another application to a deviant firm. Therefore $\bar{w}_{i^*} = 1$ is indeed the correct specification.

By similar arguments it is easy to see that for $c_{i^*} = u^*_i - u^*_{i-1}$ equilibria exist if and only if $\boldsymbol{\gamma}$ has $\gamma_{i^*} + \gamma_{i^*-1} = 1$, $\gamma_{i^*} \in [0, 1]$; i.e. workers randomize over i^* and $i^* - 1$ applications.

To show constrained efficiency, we will first consider application efficiency for given $\boldsymbol{\gamma}$ and v. Let $\hat{i}$ still denote the maximum number of applications that workers send. Consider $\hat{i}$ groups of firms, with $d_i v$ firms in each groups, which are ordered by their attractiveness for workers. That is, a worker who applies to i firms applies once to each of the lowest i groups and accepts an offer from a higher group over an offer from a lower group. We call an allocation of firms across groups that leads to the maximum number of matches $\hat{i}$-group-efficient. Compare two adjacent groups i and $i-1$ with total measure $\nu = v_i + v_{i-1}$. We show that the only efficient way of dividing this measure up between the two groups is the equilibrium division. The maximal total number of matches within these groups is given by

$$\max_{d\in[0,1]} M(d) = \nu d(1 - e^{-\mu_i}) + \nu(1 - d)(1 - e^{-\mu_{i-1}}). \tag{43}$$

It can be shown that a boundary solution cannot be optimal, as it means that one application is waisted. Noting that $(1-\psi_{i-1}) = \gamma_{i-1}/\Gamma_{i-1} + (1-\psi_i)(1-p_i)\Gamma_i/\Gamma_{i-1}$ we can write $\mu_i = (1-\psi_i)\lambda_i$ and $\mu_{i-1} = [\gamma_{i-1}/\Gamma_{i-1} + (1 - \psi_i)(1 - p_i)\Gamma_i/\Gamma_{i-1}]\lambda_{i-1}$. The first derivative is then

$$\begin{aligned}\frac{\partial M(d)}{\partial d}\frac{1}{\nu} &= 1 - e^{-\mu_i} - (1 - e^{-\mu_{i-1}}) + e^{-\mu_i} d(1 - \psi_i)\frac{\partial \lambda_i}{\partial d} \\ &+ e^{-\mu_{i-1}}(1 - d)\left[(1 - \psi_{i-1})\frac{\partial \lambda_{i-1}}{\partial d} - \frac{\Gamma_i}{\Gamma_{i-1}}(1 - \psi_i)\frac{\partial p_i}{\partial d}\lambda_{i-1}\right]\end{aligned}$$

We can use similar substitutions as for (27), with the adjustment that now $\partial\mu_i/\partial d = -\mu_i/d = -\nu\mu_i^2/[(1 - \psi_i)\Gamma_i]$, to show that the last term in the first line equals $-\mu_i e^{-\mu_i}$, and the second line reduces to $e^{-\mu_{i-1}}[\mu_{i-1} - (1 - e^{-\mu_i} - \mu_i e^{-\mu_i}]$. Therefore we again have

$$\frac{\partial M(d)}{\nu \partial d} = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - e^{-\mu_{i-1}}) - (1 - e^{-\mu_{i-1}} - \mu_{i-1}e^{-\mu_{i-1}}) = 0. \tag{44}$$

The first order condition implies equality with zero. For given ν this uniquely characterizes the optimal interior d, since similar substitutions as above yield $\partial^2 M/\partial d^2 = -\nu[\mu_2^2 e^{-\mu_2}(1 - e^{-\mu_1})/d + e^{-\mu_1}(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2} - \mu_1)^2/(1 - d)] < 0$. It is straightforward to show that for a given measure v of firms there exists an $\hat{i}$-group efficient allocation across all $\hat{i}$ groups. A similar construction as in the proof of Proposition 6 shows that $\hat{i}$ groups are sufficient to achieve the constrained optimal outcome. A similar construction as in Proposition 7 shows that the outcome of a random process (i.e. one wage) could also be achieved with $\hat{i}$ groups, but the division of firms across groups would not be optimal. More generally, such an argument establishes that the optimal allocation cannot be achieved with less than $\hat{i}$ wages given the number of applications summarized in $\boldsymbol{\gamma}$.

It is very tedious to analyze whether (44) - which coincides with profit equality as in (42) - determines the allocation of firms to the $\hat{i}$ groups uniquely for any v. Therefore we will not consider the efficiency of the application process without free entry. We will establish that the overall entry of firms and the measure of firms in each group under equilibrium conditions 1a), 1b), 2a) and 3) yields optimal entry and optimal application decisions simultaneously, taking $\boldsymbol{\gamma}$ as given. The important insight from the previous analysis of constrained optimal applications is that (44) has to hold in the optimal search outcome for all $i \in \{2, ..., \hat{i}\}$, and that we can apply the envelop theorem. Let $\boldsymbol{d}(v) = (d_1(v), d_2(v), .., d_{\hat{i}}(v))$ be the fraction of firms in each of the $\hat{i}$ groups under constrained optimal search given v and $\boldsymbol{\gamma}$. Again let $M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}(v))$ denote the constrained efficient number of matches given v and $\boldsymbol{\gamma}$. Similar to (28) the objective function is given by $\max_{v\geq 0} M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}(v)) - vK$. When $\hat{i} > 0$, then $K < 1$ ensures that the optimal solution is in the interior of $[0, V]$. We will show that the first order condition uniquely determines the solution and corresponds to the free entry condition.

By the envelope theorem the impact of a change of the fraction $d_i(v)$ of firms in each group on the measure of matches can be neglected, i.e. $\frac{\partial M^*}{\partial d_i}\frac{\partial d_i}{\partial v} = 0$ at the $\hat{i}$-efficient d_i. We get as first order condition

$$\partial M^*(\boldsymbol{\gamma}, v, \boldsymbol{d})/\partial v = K, \tag{45}$$

where $\boldsymbol{d} = \boldsymbol{d}(v)$. Writing $M^*(\boldsymbol{\gamma}, v, \boldsymbol{d}) = [1 - \sum_{i=1}^{\hat{i}}[\gamma_i \prod_{j=1}^{i}(1-p_j)]$ we have

$$\begin{aligned}\frac{\partial M^*(\boldsymbol{\gamma}, v)}{\partial v} &= \sum_{i=1}^{\hat{i}}\Big[\gamma_i \sum_{j=1}^{i}[\frac{\partial p_j}{\partial v}\prod_{\substack{k\leq i\\ k\neq j}}(1-p_k)]\Big] \\ &= \sum_{i=1}^{\hat{i}}\Big[\frac{\partial p_i}{\partial v}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)\Big],\end{aligned} \tag{46}$$

where the second line is obtained by rearranging the terms for each $\partial p_i/\partial v$. To simplify notation, define the partial sum

$$\xi_{i'} = \sum_{i\geq i'}^{N}[\frac{\partial p_i}{\partial v}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)]. \tag{47}$$

Since $p_{\hat{i}} = (1-e^{-\mu_{\hat{i}}})/\mu_{\hat{i}}$ we have $\partial p_{\hat{i}}/\partial v = -(1/\mu_{\hat{i}}^2)(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})(\partial\mu_{\hat{i}}/\partial v)$. Since $\mu_{\hat{i}} = \gamma_{\hat{i}}/(d_{\hat{i}}v)$, we have $\partial\mu_{\hat{i}}/\partial v = -\gamma_{\hat{i}}/(d_{\hat{i}}v^2) = -d_{\hat{i}}\mu_{\hat{i}}/\gamma_{\hat{i}}$. So we get $\partial p_{\hat{i}}/\partial v = -d_{\hat{i}}(1-e^{-\mu_{\hat{i}}} -$

$\mu_{\hat{i}}e^{-\mu_{\hat{i}}})/\gamma_{\hat{i}}$. Noting that $\Gamma_{\hat{i}}(1-\psi_{\hat{i}}) = \gamma_{\hat{i}}$, we have established that

$$\xi_{\hat{i}} = d_{\hat{i}}(1 - e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}}) \prod_{k<\hat{i}}(1-p_k). \tag{48}$$

By induction we can establish the following lemma, which we will prove subsequently because it would distract from the argument at this point.

Lemma A1 *For all i it holds that*

$$\xi_i = (\sum_{k=i}^{N} d_k)(1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) \prod_{j<i}(1-p_k). \tag{49}$$

This implies that $\xi_1 = 1 - e^{\mu_1} - \mu_1 e^{-\mu_1}$. The first order condition $\xi_1 = K$ uniquely defines μ_1, and corresponds to the free entry condition of the lowest wage firms. By (44) (or respectively by (42)) it also determines μ_i uniquely for all $i \in 2, ..., \hat{i}$, which in turn determines v_i uniquely for all $i \in 1, ..., \hat{i}$. Thus, the measure of firms in each group under equilibrium conditions $1a$), $1b$), $2a$) and 3) coincides with the measure of firms in each group implied by the first order conditions for optimal entry (incorporating optimal subsequent search). Since there is only one allocation fulfilling the first order conditions, and boundary solutions are not optimal, this again characterizes the global maximum. Thus equilibrium entry and search is constrained optimal given $\boldsymbol{\gamma}$.

Finally, when we endogenize $\boldsymbol{\gamma}$, again note that the number of applications of other workers in equilibrium is not important for the marginal benefits of each individual worker, which are always $u_i^* - u_{i-1}^*$. Therefore again the decision on the number of applications is constrained efficient, establishing constrained efficiency overall. *Q.E.D.*

Proof of Lemma A1:

We are left to show that the following holds for all $i \in \{1, ..., \hat{i}-1\}$:

$$\xi_{i+1} = (\sum_{k=i+1}^{\hat{i}} d_k)(1 - e^{-\mu_{i+1}} - \mu_{i+1}e^{-\mu_{i+1}}) \prod_{j<i+1}(1-p_k). \tag{50}$$

It clearly holds for $i = \hat{i} - 1$ by (48). Now assume it holds for some i. We will consider ξ_i. We know that

$$\xi_i = \xi_{i+1} + \Gamma_i(1-\psi_i)\frac{\partial p_i}{\partial v}\prod_{k<i}(1-p_k). \tag{51}$$

The second summand can be written as

$$\frac{\partial p_i}{\partial v}\prod_{k<i}(1-p_k) \quad = \quad -\frac{1 - e^{-\mu_i} - \mu_i e^{-\mu_i}}{\mu_i^2}[\frac{\partial \mu_i}{\partial v}\prod_{k<i}(1-p_k)] \tag{52}$$

Since $\mu_i = \lambda_i(1-\psi_i) = \lambda_i(\sum_{j=i}^{\hat{i}} \frac{\gamma_j}{\Gamma_i}(\prod_{k=i+1}^{j}(1-p_k)))$ we can write the term in square brackets in (52) as

$$\begin{aligned}\frac{\partial \mu_i}{\partial v}\prod_{k<i}(1-p_k) &= -\frac{\Gamma_i(1-\psi_i)}{d_i v^2}\left[\prod_{k<i}(1-p_k)\right] + \xi_{i+1}\frac{\lambda_i}{\Gamma_i(1-p_i)} \\ &= -\frac{d_i \mu_i^2}{\Gamma_i(1-\psi_i)}\left[\prod_{k<i}(1-p_k)\right] + \xi_{i+1}\frac{\lambda_i}{\Gamma_i(1-p_i)}.\end{aligned}$$

Observing that $\frac{1}{\mu_i}(1-e^{-\mu_i}-\mu_i e^{-\mu_i}) = p_i - e^{-\mu_i}$, we can substitute the prior equation into (52) and multiply by $\Gamma_i(1-\psi_i)$ to get

$$\Gamma_i(1-\psi_i)\frac{\partial p_i}{\partial v}\prod_{k<i}(1-p_k) = \frac{p_i - e^{-\mu_i}}{1-p_i}\xi_{i+1} + d_i(1-e^{-\mu_i}-\mu_i e^{-\mu_i})\prod_{j<i}(1-p_j).$$

We can substitute this into (51), and use (50) and the property of $\hat{i}$-group-efficient search in (42) to obtain

$$\xi_i = (\sum_{k=i}^{N} d_k)(1-e^{-\mu_i}-\mu_i e^{-\mu_i})\prod_{j<i}(1-p_k). \tag{53}$$

Q.E.D.

Proof of Proposition 11:

First we show that for $i^* \to \infty$ the (weakly) shorter side of the market gets matched with probability approaching 1. Since equilibrium search is always more efficient than a process of random applications and acceptances, we will show this for the latter. As $i^* \to \infty$ it cannot happen that workers and firms both are matched with probabilities bounded away from one. If that were the case, than some fraction $\alpha > 0$ of firms would always remain unmatched. But then the chance that a worker applies to such a firm with any given application is α, so that the probability that he applies to such a firm with at least one of his applications converges to 1, yielding a contradiction. With unequal sizes it is obviously the shorter side whose probability of being matched converges to one; with equal sizes the probability of being matched is the same and agents from both sides get matched with probability converging to one.

For the next arguments, recall that the marginal utility gain (excluding the marginal application cost) of the i^*'th application, given by $u^*_{i^*} - u^*_{i^*-1}$, converges to zero as $i^* \to \infty$. We will use this to establish the limit for the average wage if firms are either on the long or on the short side of the market.

Case 1: We will show that $w(i^*) \to 0$ if firms are strictly on the short side of the market. Assume there exists a subsequence of i^*'s such that $v(i^*) < 1-\epsilon$ for all i^* and some $\epsilon > 0$. That implies $\varrho(i^*) < \alpha$ for some $\alpha < 1$. If $w(i^*) \nrightarrow 0$, then there exists a subsequence such that $w(i^*) \to \omega > 0$ and $\pi(i^*) \to 1-\omega$ (since $\eta(i^*) \to 1$). Now consider a deviant firm that always offers wage $w' = \omega/2$. As workers send more applications, the hiring probability for the deviant has to converge to 1. This is due to the fact that for workers the marginal utility of sending the last application converges to zero, which implies that the probability of getting the job at the deviant firm has to become negligible as otherwise each worker would like to send his last application there to insure against the $1-\alpha$ probability of not being hired. With the hiring probability approaching 1 the profit of the deviant converges to $1-\omega/2$, i.e. the deviation is profitable. Thus it has hold that $w(i^*) \to 0$.

Case 2: We will show that $w(i^*) \to 1$ if firms are strictly on the long side of the market. Assume there exists a subsequence of i^*'s such that $v(i^*) > 1+\epsilon$ for all i^* and some $\epsilon > 0$. In this case $\eta(i^*) < \alpha$ for some $\alpha < 1$ and all i^*. If $w(i^*) \nrightarrow 1$, then there exists a subsequence such that $w(i^*) \to \omega < 1$ and $\pi(i^*) \to \pi < \alpha(1-\omega)$. Consider a firm that always offers wage $w' \in (\omega, 1)$ such that $1-w' > \alpha(1-\omega)$. Again the hiring probability of the deviant converges to 1, because if there were a non-negligible chance of getting the job at w' worker's would rather send there last application to this higher than average wage. But then the deviant's profit converges to $1-w'$ and the deviation is profitable. So $w(i^*) \to 1$.

This immediately implies that $v(i^*) \to 1$. Otherwise a subsequence of i^*'s according either to case 1 or to case 2 has to exist, but in case 1 profits are above entry costs and in case 2 they are below entry costs, violating the free entry condition. Finally, since $v(i^*) \to 1$ and firms get matched with probability close to one, $\pi = K$ implies that the average paid wage $w(i^*)$ has to converge to $1-K$. This directly implies that $u^*_{i^*} \to 1-K$.

To show that the individual search effort converges to zero, i.e. that also $U^*(i^*) = u^*_{i^*} - c^{i^*}(i^*) \to 1-K$, rewrite the workers' utility as $U^*(i^*) = \sum_{i=1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i] = \sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] + \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i]$ for some $I \leq i^*$, where $c^{i^*}_i = c^{i^*}(i) - c^{i^*}(i-1)$ again denotes marginal costs. For a given i the difference $u^*_i - u^*_{i-1}$ is simply a number independent of i^* (and the associated cost function). It converges to zero for large i, which entails that $u^*_{i^*} - u^*_{i^*-1} \to_{i^*\to\infty} 0$. Moreover $c^{i^*}_i \leq c^{i^*}_{i^*} \leq u^*_{i^*} - u^*_{i^*-1}$ for all $i \leq i^*$, which only restates that that we consider changing cost functions with $c^{i^*}_i \to_{i^*\to\infty} 0$. Therefore the partial sum $\sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] \to_{i^*\to\infty} \sum_{i=1}^{I}[u^*_i - u^*_{i-1}]$ for any fixed $I \in \mathbb{N}$. On the other hand we have $0 \leq \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i] \leq \sum_{i=I+1}^{\infty}[u^*_i - u^*_{i-1}]$, but $\sum_{i=I+1}^{\infty}[u^*_i - u^*_{i-1}] \to_{I\to\infty} 0$ since $\sum_{i=1}^{\infty}[u^*_i - u^*_{i-1}] \leq 1$. Therefore $\lim_{i^*\to\infty} U(i^*) = \lim_{I\to\infty}\lim_{i^*\to\infty}[\sum_{i=1}^{I}[u^*_i - u^*_{i-1} - c^{i^*}_i] + \sum_{i=I+1}^{i^*}[u^*_i - u^*_{i-1} - c^{i^*}_i]] = \sum_{i=1}^{\infty}[u^*_i - u^*_{i-1}] = \lim_{i^*\to\infty} u^*_{i^*} = 1-K$. *Q.E.D.*

References

[1] Acemoğlu, Daron, and Robert Shimer (2000): "Wage and Technology Disperson," *Review of Economic Studies*, 67, 585-607.

[2] Albrecht James, Pieter A. Gautier, and Susan Vroman (2006): "Equilibrium Directed Search with Multiple Applications," forthcoming in *Review of Economic Studies*, 73, 869-91.

[3] Bulow, Jeremy and Jonathan Levin (2006): "Matching and Price Competition," *American Economic Review*, 63, 652-668.

[4] Burdett, Kenneth and Kenneth L. Judd (1983): "Equilibrium Price Distributions," *Econometrica*, 51, 955-970.

[5] Burdett, Kenneth and Dale T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257-273.

[6] Burdett, Kenneth, Shouyong Shi and Randall Wright (2001): "Pricing and Matching with Frictions," *Journal of Political Economy*, 109, 1060-1085.

[7] Butters, Gerad R. (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44, 465-491.

[8] Chade, Hector, and Lones Smith (2006): "Simultaneous Search," *Econometrica*, 74, 1293-1307.

[9] Cason, Timothy and Charles Noussair (2003): "A Market with Frictions in the Matching Process: An Experimental Study," mimeo.

[10] Delacroix, Alain and Shouyong Shi (2004): "Directed Search On the Job and the Wage Ladder," *International Economic Review*, 45, 579-612.

[11] Diamond, Peter A. (1982): "Wage Determination and Efficiency in Search Equilibrium," *Review of Economic Studies*, 49, 217-227.

[12] Gale, David and Lloyd Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9-15.

[13] Gale, David (1987): "Limit Theorems for Markets with Sequential Bargaining," *Journal of Economic Theory*, 40, 20-54.

[14] Galenianos, Manolis and Philipp Kircher (2006): "Directed Search with Multiple Job Applications," mimeo.

[15] Gautier, Pieter A., and José L. Moraga-González (2005): "Strategic Wage Setting and Coordination Frictions with Multiple Applications," mimeo.

[16] Friedman, Milton (1968): "The Role of Monetary Policy," *American Economic Review*, 58, 1-17.

[17] Hosios, Arthur J. (1990): "Competitive Search Equilibrium," *Review of Economic Studies*, 57, 279-298.

[18] Immorlica, Nicole and Mohammad Mahdian (2005): "Marriage, Honesty, and Stability," *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*.

[19] Kircher, Philipp (2006): *Essays in Equilibrium Search Theory*, Ph.D. dissertation, University of Bonn.

[20] Lauermann, Stephan (2006): "Dynamic Matching and Bargaining Games: A General Approach," mimeo.

[21] Moen, Espen R. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385-411.

[22] Montgomery, James D. (1991): "Equilibrium Wage Dispersion and Interindustry Wage Differentials," *Quaterly Journal of Economics*, 106, 163-179.

[23] Mortensen, Dale T. (1982a): "The Matching Process as a Noncooperative Bargaining Game," in *The Economics of Information and Uncertainty*, ed. John J. McCall, Chicago: University of Chicago Press.

[24] Mortensen, Dale T. (1982b): "Property Rights and Efficiency in Mating, Racing, and Related Games," *Amercian Economic Review*, 72, 968-979.

[25] Mortensen, Dale T. and Randall Wright (2002): "Competitive Pricing and Efficiency in Search Equilibrium," *International Economic Review*, 43, 1-20.

[26] Ochs, Jack (1990): "The Coordination Problem in decentralized Markets: An Experiment," *Quarterly Journal of Economics*, 105, 545-559.

[27] Peters, Michael (1984): " Bertrand Competition with Capacity Constraints and Mobility Constraints," *Econometrica*, 52, 1117-1128.

[28] Peters, Michael (1991): "Ex Ante Price Offers in Matching Games: Non-Steady States," *Econometrica*, 59, 1425-1454.

[29] Peters, Michael (1997): "On the Equivalence of Walrasian and Non-Walrasian Equilibria in Contract Markets: The Case of Complete Contracts," *Review of Economic Studies*, 64, 241-264.

[30] Peters, Michael (2000): "Limits of Exact Equilibria for Capacity Constrained Sellers with Costly Search," *Journal of Economic Theory*, 95, 139-168.

[31] Phelps, Edmund S. (1967): "Phillips Curves, Expectations of Inflation and Optimal Unemployment," *Economica*, 34, 254-281.

[32] Pissarides, Christopher A. (1984): "Search Intensity, Job Advertising, and Efficiency," *Journal of Labor Economics*, 2, 128-143.

[33] Pissarides, Christopher A. (1985): "Short-run Equilibrium Dynamics of Unemployment, Vacancies and Real Wages," *American Economic Review*, 75, 676-690.

[34] Pissarides, Christopher A. (2000): *Equilibrium Unemployment Theory*, 2nd ed., Cambridge, MA: MIT Press.

[35] Roth, Alvin E. (1984): "The evolution of the labor market for medical interns and residents: A case study in game theory," *Journal of Political Economy*, 92, 991-1016.

[36] Roth, Alvin E. and Elliot Peranson (1999): "The redesign of the matching market for american physicians: Some engineering aspects of economic design," *American Economic Review*, 89, 748-780.

[37] Shi, Shouyong (2001): "Fictional Assignment. 1. Efficiency," *Journal of Economic Theory*, 98, 232-260.

[38] Shi, Shouyong (2002): "A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Biased Technology," *Review of Economic Studies*, 69, 467-491.

[39] Shi, Shouyong (2006): "Directed Search for Equilibrium Wage-Tenure Contracts," mimeo.

[40] Shimer, Robert (1996): “Essays in Search Theory,” PhD dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[41] Shimer, Robert (2005): “The Assignment of Workers to jobs in an Economy with Coordination Frictions,” *Journal of Political Economy*, 113/5, 996-1025.

[42] talnaker, John M. (1953): “The Matching Program for Intern Placement: The Second Year of Operation,” *Journal of Medical Education*, 28, 13-19.