Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 08-004

"Efficiency of Simultaneous Search"

by

Philipp Kircher

http://ssrn.com/abstract=1091427

Efficiency of Simultaneous Search*

Philipp Kircher†

First Draft: September 2006
This Version: January 2008

Abstract

We analyze a labor search model in which workers choose their search intensity by deciding how often and where to apply for jobs. They observe firms' wage postings prior to their decision. Due to coordination frictions a firm may not receive any applications; otherwise it is able to hire unless all its applicants have better offers. We show that in equilibrium the entry of firms, the search intensity and the number of filled vacancies are constrained efficient. Wage dispersion creates an (endogenous) safety net against unemployment that is essential for efficiency. As application costs vanish the equilibrium becomes unconstrained efficient.

Keywords (JEL codes): simultaneous search, directed search, efficient wage dispersion, modified Hosios condition, search with stable matchings (J64, C78, D85)

1 Introduction

Search models of the labor market are fraught with externalities. For instance, a new firm that enters the market creates positive externalities for the workers who now have an easier time finding

*This paper circulated under the title "Efficiency 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 a number of conferences.

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

a job, and creates negative externalities for other firms that now have a more difficult time hiring. In random search models in which firms meet workers by chance, this externality is in general not internalized and inefficiencies prevail (see Hosios (1990)). This changes in directed search models in which firms compete for labor through public wage announcements and workers subsequently apply for jobs. Despite the frictions in the decentralized application process, the equilibrium price for labor internalizes the market's externalities and delivers a constrained efficient outcome (see e.g. Hosios (1990), Montgomery (1991), Moen (1997), Shimer (1996, 2005), Acemoğlu and Shimer (1999), Shi (2001, 2002) and Mortensen and Wright (2002)).

The directed search literature has focused mainly on the one-sided externality of entry of firms.[1] There is no "entry" dimension for workers, since by assumption the number of unemployed workers is fixed and each worker is only allowed to send a single application. Yet frictions in the application process leave some of them unemployed. This induces an incentive for the workers to apply for several jobs in order to increase their employment prospects. This naturally introduces search intensity into the model and acts like an entry externality, but on the workers' side of the market: If a worker sends a second application, he creates negative externalities for the other workers who might find it harder to get a job and positive externalities for the firms who have an easier time hiring. This raises the question if a single instrument – the wage – can internalize this additional externality.

Recent contributions provide a negative answer. Work by Albrecht, Gautier and Vroman (2006) and Galenianos and Kircher (2006) shows that the equilibrium interaction is inefficient when workers apply to more than one job. This suggests an efficiency role for labor market interventions, and questions the robustness of the efficiency results obtained in the earlier literature.[2] However, both papers, while relaxing the restriction to a single application, impose another strong assumption. They restrict firms to contact only one of their applicants. If that applicant rejects the offer, the firms remains vacant even if several other workers applied for the job. This restriction exacerbates the negative externality of a second application: the application might "block" other workers from obtaining a job even if the applicant decides not to accept the job offer. It remains an open question whether the inefficiencies are mainly due to this additional restriction, or whether they are a deeper feature induced by the two-sided externality.

This paper investigates the efficiency of the market interaction once firms can contact all their applicants. We propose a theory of market interaction based on the idea that communication leads to a stable matching among the firms and their applicants. Stability means that a firm does not remain

[1]Some analyses such as Shi (2002) and Shimer (2005) also consider the sorting of workers to firms in a model with heterogeneities. A sorting component that we term *search efficiency* is also important in our setting despite homogeneity of agents, because multiple applications lead to endogenous sorting by risk levels.

[2]A minimum wage improves efficiency in Albrecht, Gautier and Vroman (2006). In Galenianos and Kircher (2006) a uniform government-mandated wage or interventions that reduce workers' tendency to apply for high-wage jobs can reduce negative effects associated with wage dispersion.

vacant while one of its applicants starts to work at a lower wage or remains unemployed.[3] If this would happen, both the firm and the worker would be better off by deviating and forming a match amongst themselves. Stability arises naturally if firms offer their job sequentially to applicants, and workers are free to reconsider their options. In such a process a high wage firm will entice a worker away from lower wage competitors, and a firm remains vacant only if all its offers get rejected in favor of better alternatives.[4] In our setup a firm can only communicate with those workers that applied to the firm, and therefore our model naturally differs from the usual stability analysis that presume that all agents in the economy communicate with one-another.

Our main result is that efficiency re-emerges as the equilibrium outcome in our setting. For given market frictions the equilibrium is constrained efficient along the following operative margins. *Entry Efficiency*: the constrained optimal number of firms enter. *Efficiency of Search Intensity*: the number of applications that workers send is constrained efficient. *Search Efficiency*: workers send their applications in a way that induces the optimal number of matches. Thus, even when workers can intensify their search efforts by strategically choosing a *set* of firms to which they apply, market wages are such that workers take the socially optimal application decisions and firms provide the socially optimal number of jobs. Neither interventions that incentivize workers' job search or subsidize firms' entry or wage setting are warranted on efficiency grounds within this model.

The analysis of the efficiency properties of our model reveals two additional insights: First, wage dispersion is crucial to achieve constrained efficiency when workers apply to more than one job – even though workers and firms are homogeneous and risk neutral. Low wage jobs endogenously create a safety net that provide relatively good hiring prospects for workers who happen to be unsuccessful with their applications to high wage firms. We show that at a single market wage the absence of such a safety net would lead to inefficiencies. The presence of multiple wage segments provides the right incentives for workers to apply optimally. Efficient entry of firms arises because each wage segment is governed by a modified version of Hosios' (1990) condition for optimal entry of firms. The idea that matching frictions induce wage dispersion even in settings where workers and firms are homogeneous has been pursued in a large body of work (see e.g. 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); Gaumont, Schindler and Wright (2006); Mortensen (2003)), yet the insight that wage dispersion might arise as the optimal response of the market to deal with the matching frictions in such a setting is as far as we know new to the literature.

Second, the paper also adds to the literature on asymptotic efficiency of search markets. We

[3] We treat wages as fixed in the matching stage. The implications and the applicability of this assumption is discussed in section 4.4. This notion of stability has been successfully applied in many areas; see recently Burlow and Levin (2006) who also deploy matching at fixed wages after non-cooperative wage setting stage but assume that all workers apply to all firms.

[4] Such a sequential process is known as Deferred Acceptance Process pioneered by Gale and Shapley (1962). For finite economies it converges in finite time; in our continuum economy we impose stability by assumption.

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. 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.

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.[5] The paper draws on three strands of literature. We use insights from the directed search literature (e.g. Peters (1984, 1991), 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 derive a simple characterization of the optimal decision rule; Galenianos and Kircher (2006) derive implications in an equilibrium search framework. For the final matching we apply the stability concept pioneered by Gale and Shapley (1962) to the network that formed in the search process.

Since notation and exposition remain much more tractable when we consider at most two applications per worker, Sections 2 to 4 are restricted to this case. Section 2 presents the model. Section 3 characterizes the equilibrium. Section 4 establishes constrained efficiency. In this section we also highlight the role of wage dispersion in achieving efficiency, relate our results to the literature, and discuss the assumptions that lead to efficiency in our model. Section 5 lifts the restriction to two applications per worker and, additionally, discusses convergence for vanishing application costs. Omitted proofs are gathered in the appendix.

2 The Model

2.1 Environment and Strategies

The economy consists of a continuum of unemployed workers and a continuum of vacant firms. The ratio of workers to firms, denoted v, is determined by free entry of firms.[6] Each firm has a single vacancy. If it fills its vacancy, it produces one unit of output. All agents are risk neutral. Firms maximize expected output minus wage and entry costs. Workers maximize expected wage payments minus costs of applying for jobs.

The market interaction proceeds in three stages. First, firms decide whether to enter the market

[5]Gautier and Moraga-González (2005) present a 3-player example with a similar concept for random search.

[6]Due to constant returns to scale in matching it turns out that only this ratio of workers to firms matters in our analysis.

by posting a wage that they commit to pay when they hire a worker. A firm that enters incurs cost $K < 1$ of setting up the job and advertising it to workers. Next, workers observe all wage postings. Each worker decides on the number i of applications he wants to send at cost $c(i)$, which subsumes monetary as well as time and effort costs. A worker who does not send applications does not incur any costs. 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. For now we restrict attention to $i \leq 2$ applications per worker in order to present the basic ideas in the simplest possible context; later we generalize our setup to $i \leq N$ applications. After the applications are sent out they form the links in a network between workers and firms: each worker is linked to the firms to which he applied and each firm is linked to its applicants. In the final stage workers and firms form matches, the announced wage is paid and matched pairs start production.

The matching is assumed to be stable on the network, i.e. matches form such that no firm remains vacant while one of its applicants either remains unmatched or is matched to another firm at a lower wage. This is motivated by the fact that otherwise both could deviate and be jointly better off. Our stability notion presumes that firms cannot react to other firms' offers and are committed to their wage announcement at this final stage - an assumption that we discuss in detail when we point out the forces leading to efficiency in Section 4.4.

The market is large and anonymous. To capture this in the spirit of the directed search literature we consider equilibrium strategies that are symmetric and anonymous. This implies that firms use hiring strategies that do not condition on the identity of the worker. All workers use the same application strategies, and do not condition on the firms' identities. These assumptions create the frictions in the market: the use of identical and anonymous strategies by the workers implies that they sometimes miscoordinate in the sense that several of them apply for the same job that only one of them can get. This arises even when there are fewer workers than firms. Since workers' strategies usually involve some degree of randomization over jobs, they can only anticipate which jobs are more likely to have many applicants but they cannot predict the exact realizations of applications at each vacancy and therefore they cannot target those vacancies that in the end remain vacant.

A pure strategy for an active firm is a wage offer $w \in [0, 1]$. Let F denote the wage distribution and $\mathcal{F}$ its support. That is, $F(w)$ gives the proportion of firms that offer wages below w. A worker observes the distribution of posted wages and then decides on the number of applications that he wants to send. He also decides on the firm to which he wants to apply. Given our focus on anonymous strategies, a worker applies with equal probability to all firms with the same wage.[7] Therefore we can summarize his choice of firm by the wage to which he applies. A mixed strategy for a worker is then given by the tuples $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ and $\boldsymbol{G} = (G^1, G^2)$, where γ_i is the probability of sending i

[7] This is optimal if all other workers also follow anonymous strategies. At the expense of substantial additional notation one can show that symmetry rather than anonymity of the workers' strategy is sufficient to yield identical hiring probabilities to firms with the same wage.

applications. $G^1(w)$ gives the probability that the worker sends his application to a wage below w if he sends one application. $G^2(w_1, w_2)$ gives the probability that the worker applies to a wage below w_1 with the first and below w_2 with the second application when he sends out two applications, where we assume $w_2 \geq w_1$ throughout. In our large economy these probabilities also reflect the fraction of the population that undertake a given action due our assumption that workers use symmetric strategies.

2.2 Expected Payoffs and Equilibrium Definition

To describe the expected payoffs, let $\eta(w)$ denote the probability that a firm that posts wage w hires a worker. Let $p(w)$ be the probability that an applicant at wage w receives an offer given that he does not obtain a job at a higher wage. These probabilities depend on the strategies, as we describe below. The expected profit of a firm posting wage w - omitting entry costs - is

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

The profits comprise the hiring probability multiplied by the profit margin if a worker is hired.

The utility of a worker who sends no application is $U_0 = 0$. A worker who applies with one application to wage w obtains utility

$$U_1(w) = p(w)w - c(1); \tag{2}$$

i.e., the expected wage minus the cost of the application. A worker who applies to wages (w_1, w_2) with $w_2 \geq w_1$ obtains utility

$$U_2(w_1, w_2) = p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1 - c(2). \tag{3}$$

The worker's utility is given by the wage w_2 if he is hired 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.

As in other directed search models, the endogenous probabilities $p(w)$ and $q(w)$ depend on the number of applications relative to number of firms at a given wage. This ratio is called the gross queue length because it describes the average queue of applications per job at wage w. We denote it by $\lambda(w)$. The important novelty in our setup is that a firm might not be able to hire even if it obtains an application, because the applicant might have applied to a higher wage job and is hired there. We call applications "effective" if the applicant does not obtain a better job. Denote by $\psi(w)$ the fraction of applications that are not effective in the sense that the applicant is unavailable for hiring due to better alternatives.

The ratio of effective applications to firms at wage w is given by

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

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

Due to the anonymity of the workers' strategy the effective applications that are sent to firms with wage w are uniformly distributed over the firms that offer this wage. Therefore, the number of effective applications at any individual firm at this wage is random. If the number of effective applications and firms at wage w were finite, the number of effective applications at a given firm would be Binomially distributed. For our large economy where many firms offer each wage in the wage distribution, we consider the limit when the number of both effective applications and firms tends to infinity while retaining a ratio of $\mu(w)$. In the limit the Binomial distribution converges to the much simpler Poisson distribution with parameter $\mu(w)$, and the probability that a firm receives no effective application is $e^{-\mu(w)}$. If the firm receives at least one effective application it is able to fill its vacancy, because once a firm has a worker who is not employed at a better wage it won't remain unemployed by our stability assumption. Therefore the hiring probability for a firm is

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

Next, we 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 those rival applicants that do not have a better offer either, i.e. that are effective. 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[8]

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

with the convention that $p(w) = 1$ if $\mu(w) = 0$.

The hiring probabilities obviously depend on the strategies of the agents in the economy, since the effective queue length depends on the ratio $\lambda(w)$ of gross applications to firms and the probability $\psi(w)$ that a worker gets a better job. The gross queue length is in our large economy characterized by the following condition that relates the measure of applications that workers send to the measure

[8] For a careful and intuitive derivation of the expressions for $\eta(\cdot)$ and $p(\cdot)$ as the limit of the interaction of a finite number of agents at the same wage see Burdett, Shi and Wright (2001).

of applications that firms receive:[9]

$$\gamma_1 G^1(w) + \gamma_2 G_1^2(w) + \gamma_2 G_2^2(w) = v \int_0^w \lambda(\tilde{w})\ dF(\tilde{w}) \quad \forall w \in [0,1], \tag{7}$$

where G_1^2 is the marginal distribution of $G^2(w_1, w_2)$ with respect to w_1 and G_2^2 is the marginal distribution with respect to w_2. The left hand side denotes the 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 to wages up to w, and the probability that workers who send two applications send either their low or their high application up to w. These are dispersed over the firms that offer wages up to wage w. The mass of received applications is specified on the right hand side. It is given by the ratio of applications per firm aggregated over all relevant wages, multiplied by the amount v of active firms in the market.

The probability $\psi(w)$ that an applicant is not available for hiring because he accepts some (weakly) better offer is trivially zero if workers send only one application. Moreover, if no worker applies to wage w then a firm cannot hire and $\psi(w)$ is irrelevant. Otherwise, consider some application sent to wage w in the support of the wage offer distribution, and let $\hat{G}(\tilde{w}|w)$ denote the probability that the sender mailed a second application and sent it to a wage weakly lower then $\tilde{w}$. Let $\hat{g}(w|w)$ be the probability that the sender mailed another application to the same wage.[10] The applicant takes a strictly better job if he applied to a strictly higher wage and receives an offer. This occurs with probability $\int_{\tilde{w}>w}^1 p(\tilde{w}) d\hat{G}(\tilde{w}|w)$. The applicant takes another equally good offer if he applied to it, the other firm also wants to hire him, and the applicant chooses the other firm over this firm. 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. Since workers do not condition their choices on the identities of firms, both firms have equal chances of attracting the worker, and therefore an application is not effective with probability $\hat{g}(w|w)p(w)/2$ due to offers from other equally good firms. Then $\psi(w)$ is given by

$$\psi(w) = \int_{\tilde{w}>w}^1 p(\tilde{w}) d\hat{G}(\tilde{w}|w) + \frac{p(w)}{2} \hat{g}(w|w). \tag{8}$$

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 firms about which worker to hire, a stable matching is clearly not unique. Our specification is based on the standard anonymity assump-

[9] We define $\lambda(\cdot)$ on $\mathbb{R}_+ \cup \{\infty\}$ to account for the case that a zero measure of firms might receive a mass of applications.

[10] The distribution $\hat{G}(\tilde{w}|w)$ depends on the equilibrium strategies as follows: 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 $\hat{G}(\tilde{w}|w) \equiv \sum_{j=1}^2 [\gamma_2 dG_j^2(w)/T(w)] G_{-j}^2(\tilde{w}|w)$, where $-j \in \{1,2\}/\{j\}$. Moreover, $\hat{g}(\tilde{w}|w) \equiv \hat{G}(\tilde{w}|w) - \lim_{\hat{w} \nearrow \tilde{w}} \hat{G}(\hat{w}|w)$ is the size of a possible mass point of $\hat{G}(\cdot|w)$ at $\tilde{w}$.

tion that firms hire one of their effective applicants at random.

Given any specification $\mu(.)$ for the effective queue length, equations (1), (2), (3), (5) and (6) determine the profits and utilities at all wages. We define an equilibrium as follows.

Definition 1 (Equilibrium) *An equilibrium is a tuple* $\{v, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ *such that there exists* $\mu(\cdot)$ *satisfying*

1. *Profit Maximization and Free Entry:*

   (a) $\pi(w) = \pi^* \equiv \max_{w' \in [0,1]} \pi(w')$ *for all* $w \in \mathcal{F}$.

   (b) $\pi^* = K$ *if* $v > 0$*, and* $\pi^* \leq K$ *if* $v = 0$.

2. *Utility Maximization for a Given Number of Applications and Optimal Choice of the Number of Applications:*

   (a) For any $i \in \{1, 2\}$*:* $U_i(\boldsymbol{w}) = U_i^* \equiv \max_{\boldsymbol{w'} \in [0,1]^i} U_i(\boldsymbol{w'})$ *for all* $\boldsymbol{w} \in suppG^i$.

   (b) $U_i^* = \max_{j \in \{0,1,2\}} U_j^*$ *if* $\gamma_i > 0$.

3. *Consistency:* $\mu(\cdot)$ *is consistent with (4) - (8).*

Condition 1a) specifies that firms set only those wages that maximize their profits. Condition 1 b) specifies a zero profit condition for free entry and ensures that firms only abstain from the market if they cannot earn positive profits. Condition 2a) implies that workers who send i applications send them optimally to the options available in the market. Condition 2b) ensures that workers send out the optimal number of applications. Condition 3 reiterates the conditions on 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.

We conclude this section with a discussion of the off-equilibrium-path restrictions that are imbedded in our equilibrium definition. The equilibrium definition allows firms to post any wage in $[0, 1]$. To know the profitability of a wage offer, the firm needs to know the effective queue length. Yet conditions (7) for the gross queue length and (8) for the probability that applicants take better jobs are informative only at wages that are in the support of the firms' wage offers. Off the equilibrium path, i.e. at wages $w \notin \mathcal{F}$ that no firm offers in equilibrium, firms have to form a belief about the reaction of workers. Without any restrictions on these beliefs a multitude of equilibria can be sustained by postulating that firms expect a zero queue length at off-equilibrium wages, i.e. expect not to be able to trade. Yet in the subgame after the wages are announced, workers have an incentive to apply to a deviant if it offers a high enough wage. The equilibrium definition indirectly embeds a notion of subgame perfection by ensuring that the effective queue length at out-of-equilibrium wages cannot be too low. In particular, condition $2a$) specifies that at no wage (including off-equilibrium wages) the workers' utility can be higher than at those wages to which the workers apply on the equilibrium path. In particular this rules out the belief that the queue length is zero at high wages, since workers would like to apply if the queue length were indeed zero which would violate condition $2a$).

A slightly more stringent concept is the Market Utility Assumption that is regularly applied in directed search models to capture the idea of subgame perfection more precisely. It is based on the idea that workers who can obtain the Market Utility U^* in equilibrium would not settle for a lower level of utility at any firm: *Given some Market Utility level U^* that workers obtain in equilibrium, the effective queue length at any wage w is such that a worker who applies to w (and possibly to the most suitable second wage $w' \in \mathcal{F}$) also achieves utility U^*. If there is no queue length that achieves such indifferent, the queue length at w is zero.* For wages that are offered in equilibrium this is implied by worker optimality. The Market Utility Assumption extends this to wages offered by deviating firms. We discuss the connection to subgame perfection as well as the exact mathematical formulation in the appendix. In some derivations it will be convenient to adopt this assumption, and therefore we will adopt it throughout. We show in the appendix that the set of equilibria is not affected by imposing this additional assumption.

3 Equilibrium Characterization

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

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

To establish this formally, we proceed in in three subsections. First, we analyze the workers' search behavior for a given distribution of wages and for a given number of applications that the workers send. For this we follow the approach in Chade and Smith (2006) and Galenianos and Kircher (2006). Then we analyze the distribution of wages that firms post. Finally, we determine the equilibrium number of applications.

3.1 Workers' Search Decision

Consider a single worker who observes all wages and – in equilibrium – knows the probability of success at each wage. Equilibrium condition $2a$) implies that workers apply optimally. For a worker who sends one application, applying to wage w' is optimal if and only if

$$p(w')w' = u_1 \equiv \max_w p(w)w. \tag{9}$$

That is, u_1 is the highest expected return that a worker can generate with one application. If the worker sends two applications, the analysis is a bit more involved. Let $\bar{w}$ be the highest wage that delivers u_1, i.e. $\bar{w} = \sup\{w \in [0,1] | p(w)w = u_1\}$.

Lemma 1 *The optimal choice for a worker who sends two applications 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 two applications maximizes

$$\max_{(w_1,w_2)} p(w_2)w_2 + (1 - p(w_2))p(w_1)w_1, \tag{10}$$

where we can omit the costs 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 (10) without the restriction that $w_1 \leq w_2$. Nevertheless, it is immediate that a worker who has the choice between two wages always accepts the higher wage over the lower wage. Therefore any solution to (10) has $w_1 \leq w_2$.

Since w_1 is only exercised if w_2 failed, (10) immediately implies that the choice of w_1 only depends on the expected return $p(w_1)w_1$. Therefore, the worker's optimal decision resembles that of workers with a single application and he chooses w_1 such that

$$p(w_1)w_1 = u_1. \tag{11}$$

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

$$p(w_2)w_2 + (1 - p(w_2))u_1 = u_2 \equiv \max_w \{p(w)w + (1 - p(w))u_1\}, \tag{12}$$

i.e. if it yields the highest payoff given that the low application was sent optimally. Clearly any w_1 solving (11) combined with a w_2 solving (12) yields the highest possible payoff and provides a solution to maximization problem (10). Since the optimal solution can never include $w_1 > w_2$, this implies that any solution to (11) has to be weakly lower than any solution to (12). The proposition follows since $\bar{w}$ was defined as the supremum of the wages that solve (11). *Q.E.D.*

Since all workers face the same maximization problem, the effective queue length $\mu(w)$ and the corresponding probability of being hired $p(w) = (1 - e^{-\mu(w)})/\mu(w)$ at any wage w are governed by either one of three possibilities: Very low wages are unattractive and do not receive applications, and therefore the effective queue length is zero. Wages in the intermediate range receive the low applications that workers are only willing to send if optimality condition (11) for low wage applications holds, and high wages receive high applications only if the optimality condition (12) for high wage applications is met. We summarize this in the next proposition, which is formally proven in the appendix.

Proposition 1 (Workers' Application Behavior) *In any equilibrium in which some workers send applications, i.e.* $\gamma_1 + \gamma_2 > 0$*, the following conditions for the job finding probability* $p(w)$ *hold, which by* (6) *also determine* $\mu(w)$*:*

$$\begin{aligned} p(w) &= 1 \;\; \forall\, w \in [0, u_1] && (13)\\ p(w)w &= u_1 \; \forall\, w \in [u_1, \min\{\bar{w}, 1\}] && (14)\\ p(w)w + (1 - p(w))u_1 &= u_2 \; \forall\, w \in [\bar{w}, 1], && (15) \end{aligned}$$

for some tuple (u_1, u_2) *and* $\bar{w} = u_1^2/(2u_1 - u_2)$*. If no workers apply twice, i.e.* $\gamma_2 = 0$*, then* $u_2 = u_1 + c_2$*.*

The proposition includes specifications for at all wages, including those off the equilibrium path. By the Market Utility Assumption workers are willing to apply to wages off the equilibrium path as long as the wages are not too low, and therefore exactly the same logic applies as for wages on the equilibrium path. It is worth pointing out that even if workers only send one application in equilibrium, if a firm would offer a very high wage workers might be willing to send a second application there. Workers are just willing to do this if the marginal utility of the second application is exactly equal to the additional cost, i.e. $u_2 - u_1 = c_2$. This case will be relevant whenever $\bar{w} \leq 1$, otherwise the interval $[\bar{w}, 1]$ is empty.

Proposition 1 is important for the subsequent derivations because it summarizes the schedule of job finding probabilities $p(\cdot)$ uniquely by the two numbers (u_1, u_2) that represent the workers utility from sending one and two applications (omitting application costs). These numbers will in equilibrium be dependent on the wage distribution that firms offer. Since the effective queue length is related to the probability of getting a job via $p(w) = (1 - e^{-\mu(w)})/\mu(w)$, also $\mu(\cdot)$ is uniquely determined by (u_1, u_2).

For given utility numbers we can illustrate the relationship between the effective queue length and the wage in the two-dimensional plane of Figure 1. The curve $IC_1 - IC_1$ describes the indifference curve of workers for their low wage application, i.e. it describes all combinations of queue length and wage that fulfill equation (14). The curve $IC_2 - IC_2$ describes the workers indifference curve over the high applications as given in equation (15). IC_2 is steeper than IC_1 because of the fallback option due to the low application, which induces workers to give up more of their employment chances for an increase in the wage when the high wage is concerned than when the low wage is concerned. At any wage the realized queue length has to lie on one of the two curves, because some worker has to be willing to send either his low or his high application to that wage. Moreover, the realized queue length can never be below either of the two curves as otherwise workers would have a profitable deviation. Therefore, the queue length has to lie on the upper contour of the IC_1 and the IC_2 curve, which is indicated by the dashed line.

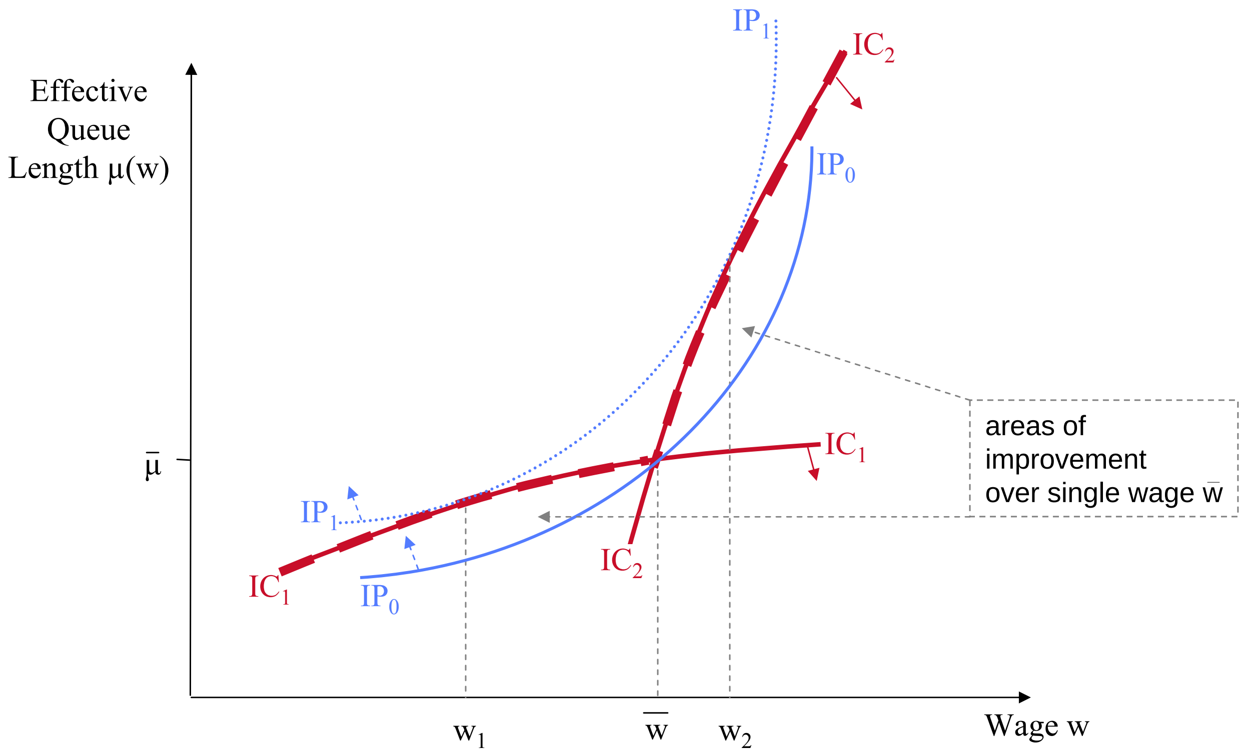

Figure 1: Illustration of market interaction for given (u_1, u_2). IC_1 and IC_2: Worker's indifference curve for the low and high wage applications, respectively. IP_0: Isoprofit curve when all firms post the same wage. IP_1: Isoprofit curve in an equilibrium with two wages. Note that the tuple (u_1, u_2) and, thus, IC_1 and IC_2 will in general change when a positive measure of firms change their wage.

3.2 Firms' Wage Setting

We now turn to the firms' problem. Their iso-profit curves $(1 - e^{-\mu})(1 - w) = \pi^*$ are smooth convex functions. The iso-profit curves are illustrated by curves $IP_1 - IP_1$ and $IP_2 - IP_2$ in Figure 1, where IP_2 is the curve with the higher profit level. Since an individual firm cannot influence the workers utility levels (u_1, u_2), an individual firm knows that the effective queue length that it will obtain is given by the workers response, i.e. by the dashed line in Figure 1. The individual firm tries to find the point on the dashed line that maximizes its profit. From the figure, we can obtain three insights that we will subsequently verify analytically.

First, if some workers send two applications, then there has to be wage dispersion in equilibrium. If firms would offer the same wage, the dashed line will adjust exactly such that the kink is at the offered wage because only at the kink are workers willing to send both their applications to this wage. The firms obtain the expected profit level associated with $\bar{w}$ and $\bar{\mu}$ and their iso-profit curve is IP_0. Since the iso-profit curve is smooth but the dashed curve is kinked because workers treat high and low wages differently, a deviant firm can move to higher profits along the dashed line. In the figure both higher and lower wages are profitable deviations. This is not generally the case, yet one direction always remains profitable.

Second, the figure illustrates that a single wage leads to an inefficient outcome *when some workers send two applications*. The area between the workers indifference curves and the firms' iso-profit curve is an area of mutual benefit for workers and firms. We explore the reasons for the inefficiency at a single wage in Section 4, yet note here that the firms' incentives to exploit the unrealized gains is exactly what rules out the single-wage equilibrium. A way to achieve efficiency is to have two market wages such that the iso-profit curve is tangent to the dashed (combined) indifference curve. This is illustrated by iso-profit curve IP_1.[11]

Third, we see that two wages can indeed constitute an equilibrium: since the isoprofit curve is tangent to both of the workers' indifference curves, a deviation that places a firm at any other point of the worker's indifference curves moves the deviant to a strictly lower level of profits.

In the following we will establish these three insights analytically. We start out by considering some tuple (u_1, u_2) and associated cutoff wage $\bar{w}$ which characterize the workers' application behavior by Proposition 1. We call firms that offer wages below $\bar{w}$ low wage firms, and those offering wage above $\bar{w}$ high wage firms. The problem for each individual firm is to maximize profits $\pi(w) = (1-e^{-\mu(w)})(1-w)$, given that $\mu(w)$ is governed by the application behavior of the workers summarized in equations (13), (14) and (15). Using these equations to substitute out the wage and recalling that $p(w) = (1 - e^{-\mu(w)})/\mu(w)$, we can write the firm's maximization problem in terms of the effective queue length

$$
\begin{aligned}
\pi(\mu) &= 1 - e^{-\mu} - \mu u_1 && \forall\ \mu \in [0, \bar{\mu}], && (16)\\
\pi(\mu) &= (1 - e^{-\mu})(1 - u_1) - \mu(u_2 - u_1) && \forall\ \mu \in [\bar{\mu}, \mu(1)], && (17)
\end{aligned}
$$

where $\bar{\mu} = \mu(\bar{w})$. In this formulation, the firm decides on the effective queue length it wants to obtain. Each queue lengths is associated with a unique wage that the firm has to pay in order to obtain it. Therefore each queue length is associated with well defined expected profits. In terms of Figure 1, the dashed curve tells the firm which wage it has to pay for given queue length. Using this way of describing the individual firm's profits we can show that any equilibrium features wage dispersion when at least some workers apply twice:

Proposition 2 (Wage Dispersion) *In any equilibrium with $\gamma_2 > 0$ the set of offered wages $\mathcal{F}$ cannot be a singleton.*

[11] In general the workers' indifference curves IC_1 and IC_2 change when the wage distribution changes because utilities (u_1, u_2) will change. Under special conditions (i.e. particular choices of wages and the number of entering firms) the utility levels remain constant, and we use this special case for easier graphical representation.

The proof in the appendix follows the logic presented above: Assume there is a pooling wage that all firms offer, and assume that it is not profitable to locally deviate to a lower wage. Due to the kink in the workers' response we show that a deviation to a higher wage has to be profitable for an individual firm, yielding the desired contradiction. We note that this reasoning differs from most other work on wage dispersion among homogeneous agents, which usually relies on a profit discontinuity of the following kind: If all firms offer the same wage, sometimes a firm cannot hire because its applicant accepts some other equally good alternative, and so a deviation to a slightly higher wage yields a jump in profits because now every applicant prefers this firm.[12] In our setting the workers understand that better hiring chances for the deviant mean lower chances for themselves, and they adjust their application behavior accordingly. Therefore profits are continuous. Dispersion arises nevertheless since workers trade off wage and acceptance probability differently for upward deviations than for downward deviations due to the availability of the fallback option at high wages, which leads to discontinuity in the derivative of the profit function. This suffices to induces wage dispersion.

Next we show that in an equilibrium in which some workers send two applications exactly two wages are 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 are offered, i.e. $\mathcal{F} = \{w_1^*, w_2^*\}$, and $w_1^* < \bar{w} < w_2^*$.*

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 profits as a function of the queue length as in (16). The function is strictly concave on $[0, \bar{\mu}]$. Therefore all low wage firms choose the same queue length and, thus, the same wage. For high wage firms profits can be written as in (17), which is strictly concave on $[\bar{\mu}, \mu(1)]$. Therefore all high wage firms choose the same queue length and, thus, the same wage. Finally, assume one group, say low wage firms, choose queue length $\bar{\mu}$ with associated wage $\bar{w}$. Since there is wage dispersion, high wage firms offer $w_2^* > \bar{w}$ at queue length $\mu_2^* > \bar{\mu}$. But since their problem is strictly concave on $[\bar{\mu}, 1]$, they make strictly higher profits than firms at $\bar{\mu}$, which yields the desired contradiction. A similar argument rules out that $\bar{\mu}$ is chosen by high wage firms. *Q.E.D.*

Given that only two wages are offered in equilibrium, it will be convenient for the subsequent derivations to simplify notation by indexing variables that refer to low wage firms by *1* and those that

[12]This logic applies 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).

refer to high wage firms by *2*.[13] Let ρ be the equilibrium fraction of firms offering the high wage. Then $\rho_2 = \rho$ and $\rho_1 = 1 - \rho$ are the fractions offering each of the two wages, and $v_i = v\rho_i$ is the measure of firms at the respective wage. Equation (7) implies gross queue lengths $\lambda_2 = \gamma_2/v_2$ and $\lambda_1 = (\gamma_1 + \gamma_2)/v_1$, where each expression reflects the measure of applications at the respective wage divided by the number of firms that offer that wage. Workers that apply to the high wage w_2 accept an offer for sure and therefore all gross applications are also effective applications, i.e. $\mu_2 = \lambda_2$. Therefore, the probability of obtaining a job at the high wage $p_2 = (1 - e^{-\mu_2})/\mu_2$ is fully specified by γ_2 and v_2. At the low wage, a fraction $\gamma_2/(\gamma_1 + \gamma_2)$ of the applicants also applies to w_2 and are unavailable for hiring if they get the better job. So by (8) a fraction $\psi_2 = p_2\gamma_2/(\gamma_1 + \gamma_2)$ of applicants cannot be hired due to better offers. Therefore the effective number of applications per firm at the low wage is $\mu_1 = (1 - p_2\gamma_2/(\gamma_1 + \gamma_2))\lambda_1$, which then defines the probability $p_1 = (1 - e^{-\mu_1})/\mu_1$ of getting a job at the low wage as a function of γ_1, γ_2, v_1 and v_2. With these definitions we can express equilibrium profits and wages:

Corollary 1 *In an equilibrium with $\gamma_2 > 0$ profits and wages for high and low wage firms, respectively, are uniquely determined by γ_1, γ_2, v_1 and v_2 as*

$$\begin{aligned} \pi_1 &= 1 - e^{-\mu_1} - \mu_1 e^{-\mu_1}, && (18)\\ w_1^* &= \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1}), && (19)\\ \pi_2 &= (1 - e^{-\mu_2} - \mu_1 e^{-\mu_1})(1 - e^{-\mu_1}), \ \textit{and} && (20)\\ w_2^* &= \mu_2 e^{-\mu_2}(1 - e^{-\mu_1})/(1 - e^{-\mu_2}) + e^{-\mu_1}. && (21) \end{aligned}$$

Proof: We know that neither low wage nor high wage firms are constrained by the boundaries of their maximization problem, 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, for low wage firms the first order condition of their profits (16) with respect to μ is necessary for optimality. It yields immediately

$$u_1 = e^{-\mu} = e^{-\mu_1}. \qquad (22)$$

The second equality follows since in equilibrium all low wage firms choose the same queue length, or rather the wage associated with it. When substituted into (16) this leads to the expression for the profits. By (14) we know that $w_1^* = u_1/p_1$ and we immediately get the corresponding wage. For high

[13] 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.

wage firms, the first order condition of profits (17) with respect to μ implies

$$u_2 - u_1 = e^{-\mu}(1 - u_1) = e^{-\mu_2}(1 - e^{-\mu_1}). \tag{23}$$

Substitution back into (17) yields the expression for the profits. By (15) 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.*

If workers no workers send two applications, then the arguments above easily establish that only one wage is offered ($v_2 = 0$). The determination of the wage is identical to the low wage in the previous corollary, i.e. it is determined by (19) and induces profits given by (18). Given that workers only apply once any offer leads to a hire, and therefore $\mu_1 = \lambda_1$ and the distinction between effective and gross queue length disappears. Therefore, for $\gamma_2 = 0$ we get the standard result from one-application models such as Burdett, Shi and Wright (2001). The introduction of a second application changes the equilibrium and essentially generates two markets. The profits in each are given by

$$\pi_i = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - u_{i-1}). \tag{24}$$

In the low wage 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 wage 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.

3.3 Equilibrium Outcome

In this section we show existence and uniqueness of the equilibrium. Before we turn to the full equilibrium, it is instructive to consider the case with exogenous search intensity γ. We show existence and uniqueness of an equilibrium with and without free entry.

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

1. *for given entry $v > 0$ there exists unique $(F, \boldsymbol{G})$ that fulfills the appropriate equilibrium conditions* 1*a*)*,* 2*a*) *and* 3*;*
2. *with free entry there exists unique $(v, F, \boldsymbol{G})$ that fulfills the appropriate equilibrium conditions* 1*a*)*,* 1*b*)*,* 2*a*) *and* 3*.*

Proof: We show part 2 here; part 1 is relegated to the appendix. The effective queue length μ_1 at the low market wage is given by the zero profit condition, which by (18) yields

$$1 - e^{-\mu_1} - \mu_1 e^{-\mu_1} = K. \tag{25}$$

The left hand side of (25) is strictly increasing in μ_1, is zero for $\mu_1 = 0$ and one for $\mu_1 \to \infty$. Therefore, μ_1 is uniquely determined. This yields w_1 by (19).

If $\gamma_2 > 0$, the profits for high wage firms is given by (20) and free entry implies

$$(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})(1 - e^{-\mu_1}) = K. \tag{26}$$

Since μ_1 is unique, μ_2 is unique. The high wage is given by (21). 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. Workers send send their low application to the low wage, and if they have a second application they send it to the high wage. By construction there are not profitable deviations from this profile: firms are willing to offer these wages since the wages were determined by the appropriate first order conditions and in each region the maximization problem is concave. Workers are willing to apply in this fashion because we used their indifference conditions (14) and (15) to construct the profits. *Q.E.D.*

Now we turn to the analysis of the equilibrium when the fraction of agents that send zero, one or two applications is endogenous. The free entry conditions (25) and (22) determine the effective queue lengths at the high and low wage solely as a function of the exogenous entry cost K, which in turn determine the utility gain that workers can get by sending an additional application by (22) and (23). In analogy to these conditions, we can define the following four numbers μ_1^*, μ_2^*, u_1^* and u_2^* recursively as follows: Let μ_1^* be the queue length that solves equation (25) for free entry at low wages. Given this number, the utility that a worker can obtain with one application is (by the first order condition (22)) represented by number $u_1^* = e^{-\mu_1^*}$. Let μ_2^* be the queue length that solves equation (26) for entry at the high wage when μ_1 is replaced by number μ_1^*. This represents the queue length at the high wage if such wage is offered, and the worker obtains (by first order condition (23)) an additional utility gain represented by number $u_2^* - u_1^* = e^{-\mu_2^*}(1 - u_1^*)$ when he sends a second application. The numbers u_1^* and $u_2^* - u_1^*$ represent the only candidates for the marginal utility gain for workers for the first and second application. They are independent of the exact structure of search intensity $\boldsymbol{\gamma}$. It turns out that the equilibrium is determined by comparing these marginal gains with the marginal costs of sending the application. The following proposition is a stronger version of Summary 1 with which we started the Section 3.

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

1. *For $c_1 > u_1^*$ in the unique equilibrium no firm enters and no application is sent.*

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.*

The key reason for uniqueness is that firms anticipate that workers will send additional applications when they offer high wages (this is captured by the Market Utility Assumption). Even if in (a candidate) equilibrium only one wage is offered and workers only send one application, a firm that deviates and offers a sufficiently high wage expects that workers will send a second application to this very high 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 whenever the marginal cost c_2 is below the marginal benefit $u_2^* - u_1^*$.

For completeness we note that in the case where $c_1 = u_1^*$ we have a continuum 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 at all. If $c_2 = u_2^* - u_1^*$ an equilibrium exists in which workers randomize between one and two applications, i.e. for any $\gamma_2 \in [0, 1]$ and $\gamma_1 = 1 - \gamma_2$ an equilibrium exists.

4 Efficiency

To discuss the efficiency properties of the equilibria just characterized, we follow Pissarides (2000) and others. We restrict the planner to the same frictions as the market; i.e., we assume that the planner can only assign symmetric strategies to workers, who are then matched according to the same stable assignment process that we applied to the decentralized economy. We call an outcome efficient if it yields the highest output in the economy, taking into account the costs of entry and the costs of sending applications.

While in equilibrium at most two wages are offered, a planner could potentially specify any number of wages in the pursuit of an optimal way by which workers apply to firms. To avoid some technical complications, we restrict attention to finite sets of wages over which the planner maximizes. Let $\mathcal{W}$ contain all finite sets of wages $W = \{w_1, w_2, ..., w_M\}$ with $0 \le w_i < w_{i+1} \le 1$. Wages themselves do not enter the efficiency notion directly since they only affect the division of the surplus. They only enter the problem indirectly as they define the rank-order of the firms for the stable matching. Denote by Φ_i^W the set of cumulative distribution functions with support on the set W^i. The planner can use any finite set of wages $W \in \mathcal{W}$, and assigns some level of entry $v \in \mathbb{R}_+$ and some wage distribution $F \in \Phi_1^W$. The planner also assigns the probability of sending zero, one or two applications and determines where these applications are sent. That is, he chooses $\boldsymbol{\gamma}$ in the three dimensional probability simplex $\triangle_3$ and $\boldsymbol{G} \in \Phi_1^W \times \Phi_2^W$. We denote the planners choice set by $\mathcal{P}^W$.

In the following formal definition of efficiency let $M(v, F, \boldsymbol{\gamma}, \boldsymbol{G}) = v \int_0^1 \eta(w) dF$ denote the number of matches when $\eta(w)$ is determined by (4) - (8) under tuple $\{v, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$. The constrained efficient outcome maximizes the number of matches minus entry and application costs:

Definition 2 (Constrained Efficiency) *Tuple $\{v, F, \boldsymbol{\gamma}, \boldsymbol{G}\} \in \mathcal{P}^W$, $W \in \mathcal{W}$, is constrained efficient* *if*

$$
\begin{aligned}
& M(v, F, \boldsymbol{\gamma}, \boldsymbol{G}) - vK - \gamma_1 c(1) - \gamma_2 c(2) \\
\geq\ & M(v', F', \boldsymbol{\gamma}', \boldsymbol{G}') - v'K - \gamma_1' c(1) - \gamma_2' c(2)
\end{aligned} \tag{27}
$$

for any $\{v', F', \boldsymbol{\gamma}', \boldsymbol{G}'\} \in \mathcal{P}^{W'}$ *with* $W' \in \mathcal{W}$.

The model has three margins of efficiency. 1. The number of matches that are achieved. 2. The entry by firms. 3. The search intensity. We show that the economy achieves efficiency along all three dimensions, which implies that interventions in the labor market are not warranted on efficiency grounds unless they tackle the underlying frictions.

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

We prove the proposition in the next three subsections. We first show that the number of matches is optimal under the appropriate subset of equilibrium conditions even if the number of firms and the search intensity are fixed. Then we additionally analyze entry by including the appropriate free entry condition and show that also entry is optimal. Finally, we also consider the endogenous search intensity and show that the entire equilibrium is constrained efficient. We discuss the assumptions that lead to efficiency and the difference to models without stable assignment in the forth subsection.

4.1 Search Efficiency

For a given vector of search intensity $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2)$ and given entry of firms v, we show that the way workers apply for jobs as characterized in the first part of Lemma 2 is constrained efficient.

We start by analyzing a narrow concept that we call *two-group-efficiency.* In analogy to the decentralized outcome first assume that the planner uses only two wages $W = \{w_1, w_2\}$ and assigns application strategies such that workers with one application apply to the low wage and those with two applications apply to both wages.[14] This restricts the planner to use only two groups of firms, one that receives high and one that receives low or single applications. Throughout the efficiency analysis we will adopt the same notation that we used for the analysis of equilibrium play in connection with Corollary 1. The planners' only choice variable under these restrictions is the fraction of firms that offers each of the two wages. We say that the search process is two-group-efficient if the fraction ρ of the high wage firms is chosen to maximize the number of matches given the assumptions just made about the workers' application strategy.

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* 1*a*), 2*a*) *and* 3) *yields two-group-efficiency.*

[14] That is, $suppG^1 = \{w_1\}$, $suppG_1^2 = \{w_1\}$ and $suppG_2^2 = \{w_2\}$

Proof: Two-group-efficiency is obtained as a solution to

$$\max_{\rho\in[0,1]} M(\rho) = \rho v(1-e^{-\mu_2}) + (1-\rho)v(1-e^{-\mu_1}), \tag{28}$$

where the first term reflects the size of high wage firms times their matching probability. Similarly, the second term accounts for the low wage firms. Recalling the notation, we have $p_i = (1-e^{-\mu_i})/\mu_i$ with $\mu_2 = \lambda_2 = \gamma_2/(v\rho)$, $\mu_1 = (1-\gamma_2 p_2/(\gamma_1+\gamma_2))\lambda_1$ and $\lambda_1 = (\gamma_1+\gamma_2)/(v(1-\rho))$. The first derivative of (28) is

$$\frac{dM}{d\rho} = \left[(1-e^{-\lambda_2}) - (1-e^{-\mu_1}) + \rho e^{-\lambda_2}\frac{d\lambda_2}{d\rho} + \frac{e^{-\mu_1}}{v\lambda_1}\left[-\gamma_2\lambda_1\frac{dp_2}{d\rho} + (\gamma_1+\gamma_2-\gamma_2 p_2)\frac{d\lambda_1}{d\rho}\right]\right]v.$$

The third term equals $-\lambda_2 e^{-\lambda_2}$ and the forth term equals $e^{-\mu_1}[-(1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2})+\mu_1]$. This yields

$$\frac{dM}{d\rho}\frac{1}{v} = (1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2})(1-e^{-\mu_1}) - (1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) = 0, \tag{29}$$

where the last equality gives the first order condition. (29) coincides with the equal profit condition between high and low wage firms. By the proof of Lemma 2 part 1 we know that this uniquely determines the measure of firms in each of the groups. Therefore, the number of high wage firms in equilibrium coincides with the number of high wage firms that a planner would choose. Optimality is ensured by global concavity since $d^2M/d\rho^2 = -v[\lambda_2^2 e^{-\lambda_2}(1-e^{-\mu_1})/\rho + e^{-\mu_1}(1-e^{-\lambda_2}-\lambda_2 e^{-\lambda_2}-\mu_1)^2/(1-\rho)] < 0$. *Q.E.D.*

Next we show that the search outcome is constrained efficient without a priori restricting the planner to use two wages. The proof in the appendix relies on a variational argument that establishes that any search strategy can be weakly improved upon by the two-group-efficient one.

Proposition 6 (Search Efficiency) *For given entry v and given search intensity $\boldsymbol{\gamma}$, the search process is constrained efficient, i.e. it holds that*

$$M(v,F,\boldsymbol{\gamma},\boldsymbol{G}) \geq \max_{F'\in\Phi_1^{W'},\boldsymbol{G}'\in\Phi_1^{W'}\times\Phi_2^{W'}} M(v,F',\boldsymbol{\gamma},\boldsymbol{G}'),$$

for any $W'\in\mathcal{W}$ when $\{F,\boldsymbol{G}\}$ conform to equilibrium conditions 1*a*), 2*a*) *and* 3.

Finally, we show that differences in hiring probabilities between different firms are indeed necessary to obtain the efficient solution when some workers send two applications. Differences in wages induce workers to search in a way that implements these differences in hiring probabilities. If only a single wage were offered, i.e. if wage offer distribution has support $\mathcal{F}=\{w\}$ for some w, then all firms would have the same hiring probability, which we show to be inefficient.

Proposition 7 (Inefficiencies without Wage Dispersion) *If some workers send two applications, identical hiring probabilities for all firms cannot be constrained efficient. In specific, a single wage at which all firms face the same effective queue length is not constrained efficient. Formally, given* v *and* $\boldsymbol{\gamma}$ *with* $\gamma_2 > 0$*, if* F *and* $\boldsymbol{G}$ *are such that* $\eta(w) = \bar{\eta}$ *for all* $w \in \mathcal{F}$ *then* $M(v, F, \boldsymbol{\gamma}, \boldsymbol{G}) < \max_{F' \in \Phi_1^{W'}, \boldsymbol{G}' \in \Phi_1^{W'} \times \Phi_2^{W'}} M(v, F', \boldsymbol{\gamma}, \boldsymbol{G}')$ *when* W' *has at least two elements.*

In the proof we show that even when we use two wages (instead of e.g. one wage) and have all workers send their high application to the high wage and their low or single application to the low wage, we can replicate a situation in which all firms have identical hiring probabilities. This can be achieved by placing the appropriate fraction of firms at each wage. Then we show that such arrangement is not optimal because too few firms offer the low wage. The reasoning is the following: Workers that end up taking jobs with low wage firms were unsuccessful with their high wage applications. The low wage firms are their last chance to avoid unemployment. On the other hand those workers that take a job at high wage firms might have gotten a job at a firm in the low wage group. Increasing workers' matching probability at the low wage at the cost of decreasing their matching probability at high wage firms improves matching for those workers for whom it is the last option to avoid unemployment. This benefit outweighs the expense of a lower matching probability for those who still might have another option.

Wage dispersion decentralizes the efficient outcome by attracting fewer applications per firm at the low wage. This provides an endogenous safety net at low wage jobs because they are relatively easy to obtain. Without wage dispersion workers equalize the queue length at all firms and a safety net would not be possible, rendering efficiency unattainable.[15] This result is surprising because with a single application different hiring probabilities are only warranted when there are productivity differences.[16] Here the source for different hiring probabilities is a sorting externality: Different hiring probabilities help to distinguish between workers who might be able to obtain some other job and workers who have no other options. Figure 1 illustrates that two wages enable a situation without efficiency gains since the iso-profit and the indifference curves are tangent.[17] With only one wage, areas for improvement always arise because workers want to obtain "insurance" with their low applications, while they do not mind to take more "risk" with the other application as they internalize that the low application acts as a fall-back option. The workers' interest is consistent with the social benefit of matching agents

[15]Our assumption of anonymity trivially induces equal hiring probabilities at firms with the same wage. Even if we only impose symmetric application strategies it is clear that the queue length at all firms with the same wage has to be equal, otherwise workers would prefer to change their behavior and apply to the firms with the lower effective queues.

[16]Shimer (2005) analyzes productivity differences when only one application is possible. If workers and firms are homogenous an identical hiring probability among all firms and all workers, respectively, is efficient (similar to our case for $\gamma_1 > 0 = \gamma 2$), and only with heterogeneities differences in hiring probabilities are efficient.

[17] In models where firms only make one offer as in Albrecht, Gautier and Vroman (2006) and Galenianos and Kircher (2006) such a figure would look very differently: next to the wage both the gross queue length $\lambda(\cdot)$ and the retention probability $\psi(\cdot)$ remain separately important in such models, and three dimensions are needed.

with higher probability at their low application, yet a single wage does not allow such a distinction in matching probabilities.

4.2 Entry efficiency

We now add the entry decision to our efficiency analysis. The intuition for the following efficiency result can be gained by considering equation (24) that describes the equilibrium profits of firms in each interval of wages. The social benefit that is created by a match is $1 - u_i$, i.e. the excess productivity over the expected output at the next lower wage.[18] The share of this benefit that accrues to the firm is $1 - e^{-\mu_i} - \mu_i e^{-\mu_i}$, which coincides with the elasticity of the matching function. This share takes into account that workers might obtain jobs at higher wages. Hosios' (1990) well-known efficiency condition for models with one application requires such a split of the surplus for efficient entry. Since this modified version of the Hosios' condition is satisfied for each wage segment, the efficient measure of firms enters each segment.

Let $M^*(v, \boldsymbol{\gamma})$ be the number of matches given search intensity $\boldsymbol{\gamma}$ and entry v when search behavior is constrained efficient as analyzed in the preceding subsection. For given $\boldsymbol{\gamma}$ we have determined the unique level of entry in the second part of Lemma 2. Here we show:

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

$$M^*(v, \boldsymbol{\gamma}) - vK = \max_{v' \geq 0} M^*(v', \boldsymbol{\gamma}) - v'K, \tag{30}$$

where v arises when equilibrium conditions $1a)$, $1b)$, $2a)$ 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 low wage and high wage firms, respectively, under 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

$$\begin{aligned} K &= \gamma_2[dp_1/dv](1 - p_2) + \gamma_2[dp_2/dv](1 - p_1) + \gamma_1[dp_1/dv] \\ &= [dp_1/dv](\gamma_1 + \gamma_2)(1 - \psi_1) + [dp_2/dv](1 - p_1). \end{aligned} \tag{31}$$

p_i depends on v directly since the measure ρv of high wage firms depends on v directly. It also depends on v indirectly since the two-group-efficient fraction ρ of high wage firms is a function of v. Since this fraction maximizes the number of matches given v, by the envelop theorem the indirect effect is zero. Consider the first term on the right hand side first. We can write $dp_1/dv = [dp_1/d\mu_1][d\mu_1/dv]$.

[18] Under free entry firms obtain no surplus, and the workers utility u_i indeed reflects productivity.

One can show that $[d\mu_1/dv](\gamma_1+\gamma_2)(1-\psi_2) = -(1-\rho)\mu_1^2 - \gamma_2\mu_1[dp_2/dv]$. Noting that $dp_1/dv = -[(1-e^{-\mu_1}-\mu_1 e^{-\mu_1})/\mu_1^2][d\mu_1/dv] = -[(p_1-e^{-\mu_1})/\mu_1][d\mu_1/dv]$ we obtain

$$[dp_1/dv](1-p_2) = (1-\rho)(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(p_1-e^{-\mu_1})[dp_2/dv].$$

Then (31) reduces to

$$\begin{aligned} K &= (1-\rho)(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \gamma_2(1-e^{-\mu_1})[dp_2/dv] \\ &= (1-\rho)(1-e^{-\mu_1}-\mu_1 e^{-\mu_1}) + \rho(1-e^{-\mu_1})(1-e^{-\mu_2}-\mu_2 e^{-\mu_2}) \\ &= 1-e^{-\mu_1}-\mu_1 e^{-\mu_1}, \end{aligned}$$

where the second line follows by taking the appropriate derivative and the last line follows as a consequence of efficient search (see (29)). The last line also denotes the profits of low wage firms in equilibrium. Applying (29) 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 there is a unique entry v that implies zero profits (Lemma 2, part 2). Therefore, 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 account for the associated entry of firms and the search outcome, and therefore also immediately establish the overall constrained efficiency of the equilibrium. This proves Proposition 5.

To gain intuition for the efficiency of the number of applications, consider the case where two wages are offered and a worker contemplates whether he should send two applications rather than one. The second application has by (23) a private marginal benefit of $e^{-\mu_2^*}(1-e^{-\mu_1^*})$. Private marginal costs are $-c_2$. Additional production arises only if the high firm does not have another effective applicant but the low firm does, which has probability $e^{-\mu_2^*}(1-e^{-\mu_1^*})$. Social and private benefits coincide. Note that the marginal benefit is independent of the economy wide vector of search intensity $\boldsymbol{\gamma}$, and therefore the decisions of other workers summarized in $\boldsymbol{\gamma}$ provide no externality on the individual. This is due to the optimal entry of firms, which absorbs any such externalities.

Let $v(\boldsymbol{\gamma})$ be the entry for a given vector of search intensity $\boldsymbol{\gamma}$ as implied by equilibrium conditions 1*a*), 1*b*), 2*a*) 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 search decisions as analyzed in the preceding subsection.

Proposition 9 (Efficiency of Search Intensity and Overall Efficiency)

The equilibrium vector of search intensity $\boldsymbol{\gamma}$ *is constrained efficient, i.e.* $\boldsymbol{\gamma}$ *solves the program* $\max_{\boldsymbol{\gamma}' \in \triangle_3} M^{**}(\boldsymbol{\gamma}') - Kv(\boldsymbol{\gamma}') - \gamma_1' c(1) - \gamma_2' c(2)$.

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. The efficient search intensity is, therefore, given when workers send applications whenever their marginal benefit is larger than their marginal cost. Proposition 4 specify such socially optimal behavior. 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.*

4.4 Discussion of the Efficiency Properties

Efficiency in our model arises because firms can "price" the good they are interested in: the queue of *effective* applications. The pricing works through the reaction by workers. Workers care only about rival applicants that are effective, i.e. they care about rival applicants that do not get jobs at higher wages. Only such applicants make it difficult to get a job at any wage. Applicants that are not effective because they accept better jobs do not prevent other workers from getting a job. If a firm changes its wage, workers react by changing the queue of *effective* applications until they are indifferent between this wage and the other wages offered in the market. This allows firms to price the effective applications at marginal cost and achieve efficiency.

In models without a stable assignment as in Albrecht, Gautier and Vroman (2006) and Galenianos and Kircher (2006) firms cannot price the good they are interested in. If each firm can only make a single offer and remains vacant if that offer gets rejected even when the firm has other applicants, then workers care about *all* rival applicants. Any applicant that applies for a job might receive an offer and blocks that job even if he decides to take a different position. Therefore, when a firm raises its wage workers respond by raising the queue of total (gross) applications to a new level at which they are indifferent this job and the other options in the market. This may not, however, result in more effective applications. If workers change their application behavior to *other* firms such that they get higher wage jobs more easily, the effective number of applications may actually go down. Firms cannot price the queue of effective applications that they are interested in, and efficiency does not obtain.

This distinction between gross and effective applications has not been considered in the literature because in models in which workers only have one application the two notions coincide: With one application no worker rejects an offer, and every application is effective.[19] Our model encompasses the

[19]This also holds for models that incorporate search intensity in a more reduced-form way through a scalar

one application case when the cost of the first application is small while the cost for two applications is prohibitively high ($c(1) \to 0$ and $c(2) \to \infty$). When we allow for general cost structures, the model extends to more applications and more wages, but within each wage segment the interaction remains similar to the one application model. In particular, the Hosios (1990) condition that ties the division of the match surplus to the elasticity of the matching function and ensures that private and social surplus in the market coincide holds per wage segment once the appropriate notion of effective applications is applied. Therefore, earlier results on efficient entry in one-application models naturally extend to our setup, and efficiency even carries over to the additional margins of number of applications and where to send the various applications.

The efficiency property of this model relies on the combination of wage commitments and stability. Wage commitments are important because they determine the terms of trade in a competitive manner before the market splits into small groups in which market forces do not lead to efficient outcomes. Weakening commitment by introducing methods for adjusting the terms of trade after the applications are sent might therefore change the results regarding efficiency.[20] In the theoretical analysis wage commitments have the additional advantage of allowing for a particularly tractable notion of stability that allows us to study the earlier stages of firm entry, wage setting and application behavior analytically. The assumption of wage commitments might be particularly applicable in markets in which large firms apply hiring policies uniformly to the entire organization. It seems also reasonable when the "wage" offers are interpreted more broadly as investments into prestige or amenities which are valued by the workers but cannot easily be adapted to the specific bargaining situation. It also applies when working conditions including wages are specified in advance of the application and matching process, as e.g. in the US market for medical residents.[21] Since other environments might not feature full commitment, it would be interesting to investigate forms of adjustments of the wage announcement in the final stage according to the application conditions that each firm and worker faces. This introduces strategic elements into the last step whose feedbacks to earlier stages render such analysis beyond the scope of this paper.

---

that increases the matching probability but yields at most one job at a time (see e.g. Moen (1995)).

[20] In Albrecht, Gautier and Vroman (2006) firms' wage announcements are not complete commitments, but firms bid the wage up to the worker's marginal product if two firms make an offer to the same worker. This has the effect that workers are mainly interested in two offers even if the announced wages are very low, which contributes to depressed wages and excessive entry.

[21] Ross (1984, p. 995-996) points out that in this market hospitals specify the job requirements and wages, *then* residents have to apply and interview with hospitals, and finally both sides of the market submit rankings of those partners *with whom they have interviewed* for a stable matching. While this paper is targeted to general job markets and might not reflect the details of this specific market – in particular we neglect heterogeneities here – it highlights the efficiency properties even of the decentralized application procedure.

5 Generalization to $N > 2$ applications

In this section we dispense with the restriction that workers cannot send more than two applications. Understanding the general case is important for several reasons. First, it allows us to assess the scope of the results of the earlier sections. The analysis reveals that the number of applications and the corresponding number of wages increase as the costs of applying decrease, and thus is not endogenously limited to the search for at most two alternatives as in other simultaneous search models (see e.g. Burdett and Judd (1983), Acemoglu and Shimer (2000)). Second, it has been argued that explicit forms of search intensity based on simultaneous search can be useful to understand the response of workers to varying labor market conditions (see Shimer (2004)), and we will see that the formulation in a directed search framework like ours remains particularly tractable and is, thus, potentially useful for the study of such wider questions. Finally, the general case allows us to study the market interaction as applications costs vanish, and we show that this induces convergence to the unconstrained efficient Walrasian outcome.

We consider the case where workers can send any number $i \in \mathbb{N}_0$ 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 $i \in \mathbb{N}$. Together with increasing marginal costs this implies a largest integer N such that $c(N) \leq 1$. Clearly, it is neither individually nor socially optimal to send more than N applications because application costs would exceed the value created in a match. Thus, the number of applications per individual remains bounded. Apart from the larger number of possible applications that each worker is allowed to send the model remains unchanged. The formal adjustments that are necessary to extend the setup to more than two applications are presented in the appendix.

The characterization of the equilibrium extends by analogy to the previous section. Again, the free entry condition defines the effective queue length in the various wage segments that arises in equilibrium. In particular

$$(1 - e^{-\mu_i^*} - \mu_i^* e^{-\mu_i^*})(1 - u_{i-1}^*) = K$$

and

$$u_i^* = e^{-\mu_i^*}(1 - u_{i-1}^*) + u_{i-1}^*$$

recursively define the effective queue length μ_i^* at the $i^t h$ highest wage as a function of entry cost K, given initial condition $u_0^* = 0$.[22] Equation (32) equates the competitive profits to the entry cost in analogy to equations (25) and (26), while equation (32) captures the outside option that lower wage segments induce for higher wage segments in analogy to equations (22) and (23). Note that $u_i^* - u_{i-1}^*$

[22] Associated wages are $w_i = (\mu_i^* e^{-\mu_i^*}/(1 - e^{-\mu_i^*}))(1 - u_{i-1}^*) + u_{i-1}^*$.

is strictly decreasing in i, while c_i is weakly increasing. In the appendix we 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 are 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, and is relegated to the appendix. The workers again partition the wages into intervals related 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.

5.1 Convergence to the Competitive Outcome

We now show that the equilibrium allocation converges to the unconstrained efficient allocation of a competitive economy when application costs become small. A competitive economy achieves the following allocation: Since entry costs are below the productivity of a match firms enter until the measure of firms equals the measure of workers. All workers and all firms get matched. Since free entry places firms on the long side of the market, firms are just compensated for their entry cost K. The market wage is then $1 - K$ and coincides with the utility of each worker.

We 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 of functions directly, it is convenient to simply consider the associated equilibrium number i^* of applications that each worker sends.[23] Vanishing costs amounts to $i^* \to \infty$. Let $v(i^*)$ denote the equilibrium measure of active firms, while $\eta(i^*) = \int \eta(w)dF$ and $\sigma(i^*) = v(i^*)\eta(i^*)$ denote the average probability of being 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 show that

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

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

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 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.

5.2 Conclusion

This paper incorporates a micro-foundation for search intensity into a directed search framework. In this setting directed search can be interpreted as strategic but frictional link formation between workers and firms.[24] Search intensity can be viewed as a choice on the number of links by the worker. We consider a stable allocation on the network. 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 by providing an endogenous safety net at low wages for applicants that were unsuccessful at obtaining high wage jobs. The stable resolution of the final matching problem allows the wage to endogenize all other externalities that are present in earlier stages.

An interesting avenue for future work is the integration of a time dimension into the hiring process that allows workers to apply again if they fail to secure a job. Simultaneous applications to multiple jobs are likely to be important when the time delay between the application process and the final hiring decision is non-trivial. Delays might be pronounced in many occupations, especially in the high skill sector. Delays are obviously severe when institutional restrictions confine hiring to designated dates, as in occupations with annual job markets. Another area for future research is the incorporation of heterogeneity into the model, which would yield additional insights into the interplay between the final matching stage and the earlier application stage.

6 Appendix

Definition and Discussion of the Market Utility Assumption:

Subgame perfection cannot be straightforwardly applied to large directed search economies for a technical reason: The restriction to symmetric strategies by workers implies that all workers have to

[24]The large literature in network formation usually deploys solution concepts that eliminate any randomness on which a frictional nature of unemployment could be based because they require that no (pair of) individuals would choose links differently after the network has been realized (see e.g. Dutta and Jackson (2000), Jackson (2005)). Our equilibrium notion requires ex-ante optimality before applications are sent out and uses miscoordination to retain frictions in the spirit of the search literature on unemployment.

apply with equal probability to a deviant. With a continuum of workers each one of them assigns negligible probability to the deviant that cannot be well specified.

The standard approach in the literature to capture the idea of subgame perfection is the Market Utility Assumption. In our multiple application case, workers can obtain utility $U_i^* = \max_{\boldsymbol{w} \in \mathcal{F}^i} U_i(\boldsymbol{w})$ from sending i applications to non-deviant firms. By choosing the number $i \in I$ of applications, they can obtain a *Market Utility* of $U_I^* = \max_{i \in I} U_i^*$. Now consider some deviant firm with wage $w \notin \mathcal{F}$ and assume his queue length were $\mu(w) \in [0, \infty]$, which defines the workers' probability $p(w)$ of getting an offer as in (6). At this queue length a worker who applies there with one application obtains utility $\hat{U}_1 = U_1(w)$ given by (2). A workers who applies to w and to some wage offered by another firm obtains at best $\hat{U}_2 = \max\{\sup_{\tilde{w} \in \mathcal{F}} U_2(\tilde{w}, w), \sup_{\tilde{w} \in \mathcal{F}} U_2(\tilde{w}, w)\}$, where the two expressions in the outer maximization operator distinguish the case where the other firm offers a lower wage from the case where the other firm offers a higher wage. The maximum utility from applying to the deviant is then $\hat{U}_I = \max_{i \in I} \hat{U}_i$, where $\hat{U}_0 = 0$.

Definition 3 (Market Utility Assumption)
For any $w \in [0,1]$, the expected queue $\mu(w) > 0$ of effective applicants is such that applicants obtain the Market Utility, i.e. $U_I^* = \hat{U}_I(w)$. If $U_I^* < \hat{U}_I(w)$ for all $\mu(w) > 0$, then $\mu(w) = 0$.

The assumption specifies that workers obtain the Market Utility not only at offered wages, but also at wages offered by deviating firms. This is based on the idea that if the queue length were higher workers would want to withdraw applications and thus decrease the queue length, while they would like to send more if the queue were lower and thus increase the queue length.[25] Throughout our analysis we impose the Market Utility Assumption. The set of application choices is $I = \{1, 2\}$ for Sections 3 and 4 and $I = \{1, ..., N\}$ for section 5. In the special case of lemma 2 where we analyze the interaction for an exogenously fixed vector of applications $\{\gamma_0, \gamma_1, \gamma_2\}$ we restrict I for logical consistency to include only those indexes i with $\gamma_i > 0$, since no workers are allowed to pursue other options. To impose the Market Utility Assumption beyond the equilibrium requirements does not change the set of equilibria:

Lemma 4 *Any equilibrium $\{v, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ can be sustained with a belief according to the Market Utility Assumption.*

Proof: Consider any equilibrium $\{v, F, \boldsymbol{\gamma}, \boldsymbol{G}\}$ which is supported by belief $\mu(\cdot)$. Consider first positive entry $v > 0$. Observe that $\mu(w)$ conforms by equilibrium condition $2a$) to the Market Utility Assumption at all wages $w \in \mathcal{G} \equiv suppG^1 \cup suppG_1^2 \cup suppG_2^2$ to which workers apply. If $\mathcal{F} \nsubseteq \mathcal{G}$ then by (7) some firms have zero queue length and make zero profits, and then by equilibrium condition 1 all firms do, which means $v = 0$ and thus a contradiction. Therefore, $\mathcal{F} \subseteq \mathcal{G}$, and $\mu(w)$ conforms to the market utility assumption at all $w \in \mathcal{F}$.

Now define a different belief $\hat{\mu}(\cdot)$ that conforms to the Market Utility Assumption at all $w \in [0,1]$. On $\mathcal{F}$ the trade-offs for all agents remain unchanged. Under $\hat{\mu}(\cdot)$ workers are indifferent to applying to wages in $[0,1]$ by construction, and so condition $2a$) is not violated. Now observe that $\hat{\mu}(w) \leq \mu(w)$ for all $w \in [0,1] \backslash \mathcal{F}$. To see this, assume $\hat{\mu}(w) > \mu(w)$ at some $w \in [0,1] \backslash \mathcal{F}$. Since under $\hat{\mu}(w)$ workers

[25] While the arguments presented here only intuitively appeal to "reasonable" responses of workers to deviating wage offers, papers by Peters (1991, 1997, 2000) and Burdett, Shi and Wright (2001) 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. Multiple applications induce additional complications in finite economies that only disappear in the limit, since success at one application is revealing about the success of the other in finite environments (see Albrecht et al. (2005)).

are indifferent to applying to the off-equilibrium wage w, they strictly prefer the off-equilibrium w under $\mu(w)$ over the wages in the support of their randomization. This violates $2a$) and $\mu(\cdot)$ cannot have sustained the equilibrium. Since $\hat{\mu}(w) \leq \mu(w)$ for all $w \in [0,1]/\mathcal{F}$, deviations by firms are weakly less attractive under $\hat{\mu}(w)$ than under $\mu(w)$, and equilibrium condition 1 is also fulfilled.

Now consider $v = 0$. Then $2b$) and 3 imply $\gamma_0 = 1$. By $2b$) $\mu(w)$ has to be such that $p(w)w \leq c_1$. The Market Utility Assumption yields effective queue length $\hat{\mu}(w)$ such that $\hat{p}(w)w = c_1$ if $\hat{\mu}(w) > 0$. By construction no worker wants to change his behavior. Moreover, $\hat{p}(w) \geq p(w)$ implies $\hat{\mu}(w) \leq \mu(w)$ and firms are less willing to enter under $\hat{\mu}$. *Q.E.D.*

Proof of Proposition 1:

At any wage $w < u_1$ the Market Utility cannot be obtained, yielding $\mu(w) = 0$ by the Market Utility Assumption. Also 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 (11) has to hold for all wages in $(u_1, \bar{w}]$ in order to provide the Market Utility.

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 (12). 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 neighborhood of $\bar{w}$ would offer a utility different from the Market Utility. Therefore $\bar{w}$ is determined as the wage where both (11) and (12) hold.

Even when $\gamma_2 = 0$ the workers might prefer to send a second application if a high wage were offered. Assume the queue length would be governed by (11) for all wages in $(u_1, 1]$. If $p(w)w + (1 - p(w))u_1 - c_2 \geq u_1$, workers would strictly like to send a second application, which contradicts with equilibrium condition 2a). In this case at higher wages the queue length is again governed by (12), only that $u_2 = u_1 + c_2$ ensures that workers are indifferent between sending two or one application in accordance with the Market Utility Assumption. *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 (7) 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, which leads by (7) to a zero probability of getting a job and violates workers optimality $2b$). Then profits for wages above w^* are given by (17), for wages in $[p(w^*)w^*, w^*]$ by (16), 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 (16) to get $\pi'_-(\bar{\mu}) = e^{-\bar{\mu}} - u_1$, and the right derivative by differentiating (17) which yields $\pi'_+(\bar{\mu}) = e^{-\bar{\mu}}(1 - u_1) - (u_2 - u_1)$. In equilibrium it needs to hold that firms 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

$$-e^{-\bar{\mu}}u_1 - (u_2 - u_1) \leq -u_1. \tag{32}$$

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 (32) reduces to

$$(1 - e^{-\bar{\mu}} - \bar{\mu}e^{-\bar{\mu}})u_1 \leq 0. \tag{33}$$

We know that $u_1 > 0$ since $w^* > 0$. Since workers send a strictly positive measure of applications we have $\bar{\mu} > 0$, and it is easily shown that the term in brackets is strictly positive in this case, 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 (14), wages below one provide profitable deviations. Now consider $w^* = 0$. Equilibrium profits are strictly smaller than one because not all firms get matched. (15) 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 (25) and (26), we now have

$$\begin{aligned} 1 - e^{-\mu_1} - \mu_1 e^{-\mu_1} &= \hat{\pi}, && (34)\\ (1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})(1 - e^{-\mu_1}) &= \hat{\pi}, && (35)\end{aligned}$$

for some endogenous profit $\hat{\pi}$. Consider $\hat{\pi}$ as a free parameter. For a given $\hat{\pi}$ (34) and (35) 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 (34) μ_1 strictly increases in $\hat{\pi}$. Equal profits at high and low wage firms implies

$$1 - e^{-\mu_2} - \mu_2 e^{-\mu_2} = 1 - \mu_1 e^{-\mu_1}/(1 - e^{-\mu_1}), \tag{36}$$

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 $d\mu_1/d\hat{\pi} = -[\mu_1/\hat{v}_1][d\hat{v}_1/d\hat{\pi}] - (1/\hat{v}_1)(1 - e^{-\mu_2} - \mu_2 e^{-\mu_2})[d\hat{v}_2/d\hat{\pi}]$. 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 $d\hat{v}_1/d\hat{\pi} + d\hat{v}_2/d\hat{\pi} < 0$. $\mu_1 > 1 - e^{-\mu_2} - \mu_2 e^{-\mu_2}$ holds because it is by (36) 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 4:

We first show that an equilibrium without entry exists if $c_1 > u_1^*$, while it does not exist if $c_1 < u_1^*$. Consider some (candidate) equilibrium without entry, i.e. $v = 0$, sustained by some function $\mu(.)$ for the effective queue length. Then by equilibrium conditions $2a$) and $2b$) the conjecture about the queue that would prevail under entry cannot be so low that workers would like to send applications. In particular, it has to hold that

$$\frac{1 - e^{1-\mu(w)}}{\mu(w)} w \leq c_1 \tag{37}$$

for all $w \in [0, 1]$, and (37) has to hold with equality for all $\mu(w) > 0$ by the Market Utility Assumption.[26] Substitution into the firm's profit function and taking first order conditions implies that the highest profit for an entering firm is at wage w' such that $e^{-\mu(w')} = c_1$, yielding profits $\pi(w') = 1 - e^{-\mu(w')} - \mu(w')e^{-\mu(w')}$. Note that $c_1 < u_1^*$ implies $e^{-\mu(w')} < e^{-\mu_1^*}$ or $\mu(w') > \mu_1^*$, while $c_1 > u_1^*$ conversely implies $\mu(w') < \mu_1^*$.

In case 1 we have $c_1 > u_1^*$ and therefore $\pi(w') = 1-e^{-\mu(w')}-\mu(w')e^{-\mu(w')} < 1-e^{-\mu_1^*}-\mu_1^* e^{-\mu_1^*} = K$, where the inequality follows since $1 - e^{-\mu} - \mu e^{-\mu}$ is increasing in μ on $\mathbb{R}_+$. Therefore, $v = 0$ does not violate equilibrium condition 1*b*). Since positive entry would result in marginal benefits $u_1^* < c_1$ that would not sustain any application behavior, there cannot be any equilibria with entry.

In cases 2 and 3 we have $c_1 < u_1^*$ and thus $\pi(w') = 1 - e^{-\mu'} - \mu' e^{-\mu'} > 1 - e^{-\mu_1^*} - \mu_1^* e^{-\mu_1^*} = K$. Therefore $v = 0$ cannot be an equilibrium as a firm that would enter would make strictly positive profits, violating equilibrium condition 1*b*). Therefore, assume an equilibrium with with $v > 0$ exists (as we will show in the following). Since u_1^* is the marginal utility at the low wage under firms free entry condition, all workers will apply at least once since $u_1^* > c_1$. The question is whether also a high wage will be offered. Let's assume only one wage is offered and workers only send one application. Workers do not want to deviate if it is not profitable to send another application to the offered wage (because at other wages the queue length is determined by their indifference by the Market Utility Assumption). It can be shown that a worker would only deviate if $c_2 < e^{-\mu_1^*}(\mu_1^* - 1 + e^{-\mu_1^*})/\mu_1^*$, which is stronger than $c_2 < u_2^* - u_1^*$. So in case 1 equilibrium conditions 2*a*) and 2*b*) will be fulfilled, and even for some parameters in case 2 workers would not start sending additional applications even if all firms offered only one wage.

Firms would deviate from the only candidate for a single wage $w_1^* = [\mu_1^* e^{-\mu_1^*}]/[1 - e^{-\mu_1^*}]$ (see (19)) despite the fact that this candidate 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 happens in region $[\bar{w}, 1]$ according Proposition 1. Since by construction the wage w_1^* is 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 (17) 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

$$\pi(\hat{\mu}) = (1 - e^{-\hat{\mu}} - \hat{\mu} e^{-\hat{\mu}})(1 - e^{-\mu_1^*}). \tag{38}$$

Comparing (38) with (26) 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 exists in case 2 [and entry is such that $v = 1/\mu_1^*$], 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 exists in case 3 since $u_2^* - u_1^* > c_2$ [with entry $v = v_1 + v_2$, $v_2 = 1/\mu_2^*$ and $v_1 = (1 - (1 - e^{-\mu_2^*})/\mu_2^*)/\mu_1^*$]. 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.*

[26]This captures the idea of subgame perfection similar to the Market Utility Assumption: workers are indifferent between applying to the deviant that entered and not applying at all, which is the prerequisite for having some workers apply while other do not.

Proof of Proposition 6:

For $\gamma_0 = 1$ or $v = 0$, the result is trivial since matches are always zero. When workers send one application ($\gamma_2 = 0$), one group of firms with equal hiring probability as in the decentralized equilibrium is indeed optimal because of strict concavity of the firms' matching probability $1 - e^{-\lambda}$. This is proven, e.g., in Shimer (2005).

For $\gamma_2 > 0$, we first note that for any $\{F, \boldsymbol{G}\} \in \Phi_1^W \times \Phi_1^W \times \Phi_2^W$ on $W = \{w_1, w_2, ..., w_M\}$ the effective queue length $\mu(w_i)$ and its components $\lambda(w_i)$ and $1 - \psi(w_i)$ are uniquely defined by equations (4) to (8) at all wages that are offered.[27] We prove the result using a variational argument. Let $\{F, \boldsymbol{G}\} \in \Phi_1^W \times \Phi_1^W \times \Phi_2^W$ be some possibly optimal wage setting and application behavior on $W = \{w_1, w_2, ..., w_M\}$ with $0 < w_i < w_{i+1} < 1$, and define $w_0 = 0$ for notational convenience.[28] We will show that we can find a two-group efficient matching (with two wages) that weakly improves on $\{F, \boldsymbol{G}\}$. The proof is structured into two steps. First, we show that for each wage in W we can decompose the application process such that some firms only receive high and others only receive low applications without changing the overall matches in the economy. In the second part of the argument we apply the decomposition from step 1 successively to all wages and rearrange wages. This leaves us with some firms at high wages that only receive the high application of any worker who sends two applications, and some firms at low wages that only receive the low (or single) application of every worker. We then show that it weakly improves the matching if workers send the applications that go to low wage firms randomly, and send the applications that go to high wage firms randomly. Therefore, we only need one high and one low wage to generate the optimal matching, and two-group-efficiency is indeed the best possible outcome.

Step 1: Restrict W to consist only of those wages w that are offered by firms, i.e. $w \in \mathcal{F}$, and to which some workers apply, i.e. $w \in suppG^1 \cup suppG_1^2 \cup suppG_2^2$. All other wages do not contribute to the matching in the market and can be removed from W. This immediately implies that at each remaining wage $w \in W$ we have $\lambda(w) \in (0, \infty)$. $\{F, \boldsymbol{G}\}$ can be completely described by the measure $f(w_i)$ of firms offering wage w_i, the measure $g^1(w_i)$ of workers that send a single application and send it to w_i and the measure $g^2(w_i, w_j)$ of workers that apply to the combination of wages (w_i, w_j), $i \leq j$.[29] By construction $g^2(w_i, w_j) = 0$ if $i > j$. The measure of workers that send two applications and send their low application to w_i is then given by the marginal $g_1^2(w_i) = \sum_j g(w_i, w_j)$, while the measure that send their high application to w_i is given by $g_2^2(w_j) = \sum_i g(w_i, w_j)$.

Consider some wage $w_i \in W$ that receives both high as well as low (or single) applications by agents, i.e. $g_2^2(w_i) > 0$ and either $g_1^2(w_i) > 0$ or $g^1(w_i) > 0$. Let $\lambda_h(w_i) = \frac{g_2^2(w_i) - g(w_i, w_i)}{f(w_i)}$ be the gross queue of high applications to firms at w_i, excluding the applications from agents that sent both applications to the same wage. Let $\lambda_l(w_i) = \frac{g^1(w_i) + g_1^2(w_i) - g(w_i, w_i)}{f(w_i)}$ be the gross queue of low or single

[27] $\lambda(w_i)$ is simply the measure of applications sent to w_i divided by the measure of firms offering this wage. $\psi(w_i)$ is determined as a fixed point to (8) since $p(w_i) = (1 - e^{-\mu(w_i)})/\mu(w_i)$ and $\mu(w_i) = (1 - \psi(w_i))\lambda(w_i)$. Since $\int_{\tilde{w} > w_i}^1 p(\tilde{w}) d\hat{G}(\tilde{w}|w) =: A$ is exclusively determined at higher wages and $\hat{q}(w_i|w_i) =: B$ is determined by $\boldsymbol{G}$, both independent of $\psi(w_i)$, we have a fixed point $\psi(w_i) = A + (B/2)(1 - e^{-(1-\psi(w_i))\lambda(w_i)})/[(1 - \psi(w_i))\lambda(w_i)]$. The left hand side is linear in $\psi(w_i)$, the right hand side strictly convex in $\psi(w_i)$, which can be used to show existence of a unique fixed point. The system can be solved recursively starting at the highest wage.

[28] Strict inequalities at the boundaries does not restrict optimality, as the only purpose of the wages is to determine the rank in the matching, and it facilitates the specification of a larger grid in the later exposition.

[29] That is, $f(w_i) = v(F(w_i) - F(w_{i-1}))$, $g^1(w_i) = \gamma_1(G^1(w_i) - G^1(w_{i-1}))$ and $g^2(w_i, w_j) = \gamma_2[G^2(w_i, w_j) - G^2(w_{i-1}, w_j) - G^2(w_i, w_{j-1}) + G^2(w_{i-1}, w_{j-1})]$.

applications to firms at w_i, again excluding agents that sent both applications to the same wage. Let $\lambda_m(w_i) = g(w_i, w_i)/f(w_i)$ be the ratio of workers that send both applications to w_i to firms offering that wage. Each of these latter workers sends two applications to this wage, and therefore $\lambda(w_i) = \lambda_h(w_i) + \lambda_l(w_i) + 2\lambda_m(w_i)$. Let $\bar{\psi}_i$ be the average probability that someone who applied to w_i *and* to a strictly higher wage obtains an offer at the high wage.

We want to show that we can achieve the same number of matches by introducing an additional wage $w_i^L \in (w_i, w_{i-1})$ and have all high wage applications still be sent to w_i while all low (or single) applications are now sent to w_i^L, while neither changing the application behavior nor the measure of firms at higher or lower wages. That is, we assign new strategies F' and G' on $W \cup \{w_i^L\}$ with associated f' and g'.[30] Only the measure of firms at w_i and w_i^L is not determined by this specification, i.e. we are free to choose $f'(w_i)$ and $f'(w_i^L)$ as long as they add to the original measure $f(w_i)$.

We first notice that the effective queue length $\mu'(w_j)$ under the new strategies coincides with $\mu(w_j)$ for all $w_j > w_i$, as we did not change the application behavior at higher wages and the matching rates at higher wages are not influenced by changes at lower wages. Therefore, the average probability that somebody who sends his application to w_i^L and to a wage strictly above w_i gets the high wage offer remains $\bar{\psi}_i$. Next, we observe that we can find assign measure $f'(w_i)$ and $f'(w_i^L)$ adding to $f(w_i)$ such that $\mu'(w_i) = \mu'(w_i^L)$, i.e. the effective queue length at both wages is equalized. Let $\mu_h(\rho)$ and $\mu_h(\rho)$ denote the effective queue length at w_i and w_i^L, respectively, given that a fraction ρ of the $f(w_i)$ firms offers the higher of the two wages. $\mu_h(\rho) = \frac{\lambda_h(w_i)+\lambda_m(w_i)}{\rho}$, because all applications are effective. For those firms in the second group it is $\mu_l(\rho) = \frac{(1-\bar{\psi})\lambda_l(w_i)}{1-\rho} + \frac{1-e^{-\mu_h(\rho)}}{\mu_h(\rho)}\frac{\lambda_m(w_i)}{1-\rho}$. For ρ close to zero $\mu_h(\rho) > \mu_l(\rho)$, while for ρ close to 1 $\mu_h(\rho) < \mu_l(\rho)$. By the intermediate value theorem it is possible to equalize both at some $\rho' \in (0, 1)$. Let $f'(w_i) = \rho' f(w_i)$ and $f'(w_i^L) = (1-\rho)f(w_i)$.

Note that the effective queue length $\mu'(w_i)$ under the new strategies is identical to $\mu(w_i)$ under the original strategies. Assume not, e.g. $\mu'(w_i) = \mu'(w_i^L) > \mu(w_i)$. That means that under μ' strictly more of the measure $f(w_i)$ of firms get matched then under μ at wages w_i and w_i^L. 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 these wages. Since workers and firms are matched in pairs, we cannot obtain more matches for firms than for workers. Similarly $\mu'(w_i) < \mu(w_i)$ can be ruled out. Therefore, $\mu'(w_i) = \mu(w_i^L) = \mu(w_i)$ and we have not changed the overall matching probabilities at any of the wages, because firms at wages below w_i^L face exactly the same probabilities that their workers take other jobs than under the original strategy.

Step 2: Repeating step 1 successively for each $w_i \in W$ leaves us with a set of low application wages $\{w_1^L, ..., w_M^L\}$ and a set of high application wages $\{w_1, ..., w_M\}$ with effective queue lengths $\mu'(w_i) = \mu'(w_i^L) = \mu(w_i)$ and measures of firms $f'(w_i) + f'(w_i^L) = f(w_i)$, and since the firms still face the same effective queue length we have an unchanged number of matches in the economy. Now we can rearrange the wages such that any w_i^L wage is below any w_j wage without changing the number of matches, since any pair of wages (w_i^L, w_j) with $g'(w_i^L, w_j) > 0$ already has by construction $w_j > w_i^L$. That is, we can use a new set of wages $\check{W} = \{\check{w}_1, ..., \check{w}_{2M}\}$ such that $0 < \check{w}_1 < ... < \check{w}_{2M} < 1$, a mapping κ that assigns $\kappa(\check{w}_i) = w_i^L$ if $i \le M$ and $\kappa(\check{w}_i) = w_{i-M}$ if $i \ge M$, and strategies $\check{f}(\check{w}_i) = f'(\kappa(\check{w}_i))$, $\check{g}^1(\check{w}_i) = g^{1\prime}(\kappa(\check{w}_i))$ and $\check{g}^1(\check{w}_i, \check{w}_j) = g^{1\prime}(\kappa(\check{w}_i), \kappa(\check{w}_j))$ that leave the overall matching unchanged.

The next step is to show that it is sufficient to have only one low wage and one high wage. We say that we "collapse" two wages $\check{w}$ and $\check{w}'$ into one if we assign new strategies $\tilde{f}$ and $\tilde{g}$ that are identical

[30] $f'(w) = f(w)$ and $g^{1\prime}(w) = g^1(w)$ for all $w \in W/\{w_i\}$, $g^{2\prime}(w_a, w_b) = g^2(w_a, w_b)$ for all $(w_a, w_b) \in (W/\{w_i\})^2$, $g^{1\prime}(w_i^L) = g^1(w_i)$ and $g^{1\prime}(w_i) = 0$, $g^{2\prime}(w_i^L, w_b) = g^2(w_i, w_b)$ for $w_b \in W/\{w_i\}$, $g^{2\prime}(w_a, w_i) = g^2(w_a, w_i)$ for $w_a \in W/\{w_i\}$, $g^{2\prime}(w_i^L, w_i) = g^2(w_i, w_i)$, and $g^{2\prime}(w_a, w_b) = 0$ for all other $(w_a, w_b) \in W'^2$.

to the strategies $\check{f}$ and $\check{g}$ except that all firms that offered $\check{w}'$ now also offer $\check{w}$ and all applications that were sent to $\check{w}'$ are now sent to $\check{w}$.[31] We will show that by collapsing all low wages and collapsing all high wages we weakly improve the number of matches.

Consider first two low wages $\check{w}$ and $\check{w}'$ in $\{\check{w}_1, ..., \check{w}_M\}$, i.e. wages at which firms only receive low (or single) applications. At the former the matching probability for a firm is given by $1 - e^{-(1-\psi)\lambda}$ for the appropriate average ψ and average gross queue length λ under $\tilde{f}$ and $\tilde{g}$, while the matching probability is $1 - e^{-(1-\psi')\lambda'}$ at the latter. We show that in an optimal allocation firms at either wage have the same queue length by shifting firms from one group to the other while leaving the applications that each group receives the same. That is, we retain $\tilde{f}$ and $\tilde{g}$, except for $\tilde{f}(\check{w})$ and $\tilde{f}(\check{w}')$ for which we retain the overall measure $\nu = \tilde{f}(\check{w}) + \tilde{f}(\check{w}')$ but assign a fraction ρ to $\check{w}$ and the remainder to $\check{w}'$ so as to get to optimal number of matches. Let γ and γ' denote the gross measure of applications at $\check{w}$ and $\check{w}'$. We consider ν, γ, γ' strictly positive, as otherwise at least one wage does not contribute to the number of matches and can be dropped form the analysis. Then the number of matches across both groups is given by

$$\nu[\rho(1 - e^{-(1-\psi)\frac{\gamma}{\nu\rho}}) + (1-\rho)(1 - e^{-(1-\psi')\frac{\gamma'}{\nu(1-\rho)}})]. \tag{39}$$

Since both subgroups have a strictly positive measures of applications, 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 ρ is characterized by the first order condition

$$\nu[(1 - e^{-\mu}) - (1 - e^{-\mu'}) - \mu e^{-\mu} + \mu' e^{-\mu'}] = 0, \tag{40}$$

where $\mu = (1-\psi)\frac{\gamma}{\nu\rho}$ and $\mu' = (1-\psi')\frac{\gamma'}{\nu(1-\rho)}$. 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. That means that in the optimal allocation both groups have the same hiring probability, and by collapsing the two wages they remain to have the same hiring probabilities (which is identical to the case prior to the collapsing as otherwise one would match a different measure of firms than workers – as outlined in the last paragraph of step 1). Therefore the overall matching is unchanged. Repeating this argument, we can collapse all wages in $\{\check{w}_1, ..., \check{w_M}\}$ to a single wage.

By this construction, for low and single application firms only the *average* matching probability at high application firms matters. If we keep the measure 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 wages $\{\check{w}_{M+1}, ..., \check{w}_{2M}\}$ and maximize the number of matches there. At high wages the gross and the effective queue length coincide. 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). Since the queue length at all high length firms is identical in the optimal matching, we can collapse them successively to a single wage without changing the number of matches in the economy.

[31] That is, $\tilde{f}$ and $\tilde{g}$ coincide at all wage (combinations) with $\check{f}$ and $\check{g}$ that do not involve $\check{w}$ or $\check{w}'$ and have $\tilde{f}(\check{w}) = \check{f}(\check{w}) + \check{f}(\check{w}')$, $\tilde{f}(\check{w}') = 0$, $\tilde{g}^1(\check{w}) = \check{g}^1(\check{w}) + \check{g}^1(\check{w}')$, $\tilde{g}^1(\check{w}') = 0$, $\tilde{g}^2(\check{w}, \check{w}) = \check{g}^2(\check{w}, \check{w}) + \check{g}^2(\check{w}, \check{w}') + \check{g}^2(\check{w}', \check{w})$, $\tilde{g}^2(\check{w}, \check{w}_b) = \check{g}^2(\check{w}, \check{w}_b) + \check{g}^2(\check{w}', \check{w}_b)$ when $\check{w}_b \notin \{\check{w}, \check{w}'\}$, and $\tilde{g}^2(\check{w}_a, \check{w}) = \check{g}^2(\check{w}_a, \check{w}) + \check{g}^2(\check{w}_a, \check{w}')$ when $\check{w}_a \notin \{\check{w}, \check{w}'\}$.

This leaves us with two wages, one that attracts only low (and single) applications and one that attracts only high applications, and we have either not changed the number of matches compared to those arising from $\{F, \mathbf{G}\}$ or we have improved upon them. Our analysis of two-group efficiency revealed that the equilibrium achieves the highest number of matches in this two-wage environment, which means that the equilibrium weakly improves upon $\{F, \mathbf{G}\}$. *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 $\eta' = \eta''$, since otherwise one tuple would match more workers but fewer firms than the other. Therefore, all strategies that have identical matching probabilities for all firms achieve the same measure of matches $v\eta'$.

Assume we have a grid W at our disposal that has at least two elements w_1 and $w_2 > w_1$. We use a strategy that assigns some firms to w_1 and all others to w_2, and all workers send their high application to the second and their low or single application to the first wage. Let ρ be the fraction of firms at the high wage. We have $\mu_2 = \lambda_2 = \gamma_2/(v\rho)$ and $\mu_1 = [1 - \gamma_2 p_2/(\gamma_1 + \gamma_2)]\lambda_1 = [\gamma_1 + \gamma_2 - \gamma_2(1 - e^{-\lambda_2})/\lambda_2]/(v(1 - \rho))$. Since for $\rho \approx 0$ clearly $\mu_2 > \mu_1$ and for $\rho \approx 1$ $\mu_2 < \mu_1$, there exists a $\hat{\rho}$ 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 ρ around $\mu_2 \approx \mu_1$, so that $\hat{\rho}$ is unique. This two-group process has $\mu_1 = \mu_2$, but the optimal two group process fulfills (29), which requires $\mu_1 < \mu_2$, i.e. a strictly smaller fraction of firms at the high wage compared to the assignment that equalizes matching probabilities at both wages. *Q.E.D.*

Extended Setup for Section 5:

We generalize the setting to the case where workers can send any number $i \in \mathbb{N}_0$ of applications, at a cost $c(i)$. We have $c(0) = 0$, increasing marginal costs $c_i = c(i) - c(i-1)$, and largest integer N such that $c(N) \leq 1$. Since many arguments are straightforward generalizations of that special case, we focus mainly on the changes that are necessary to adapt the prior setup.

The extension requires mainly adaptations of the workers' setup, while it remains essentially unchanged for firms. The workers' strategy is now a tuple $(\boldsymbol{\gamma}, \boldsymbol{G})$, where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, ..., \gamma_N) \in \triangle_N$ and $\boldsymbol{G} = (G^1, G^2, ..., G^N) \in \times_{i=1}^N \Phi^i$, where $\triangle_N$ is the N-dimensional unit simplex and Φ^i the set of cumulative distribution functions of $[0,1]^i$. γ_i denotes the probability of sending i applications, and G^i denotes the cumulative distribution function over $[0,1]^i$ that describes the application behavior. Let $(w_1, ..., w_i)$ 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. A worker who applies to $(w_1, ..., w_i)$ attains in analogy to (3) the utility

$$U_i(w_1, ..., w_i) = \sum_{j=1}^{i} \left[\prod_{k=j+1}^{i} (1 - p(w_k)) \right] p(w_j) w_j - c(i). \tag{3'}$$

A worker who applies nowhere attains $U_0 = 0$. Instead of (7) the relevant condition is now

$$\sum_{i=1}^{N} \left[\gamma_i \sum_{j=1}^{i} G^i_j(w) \right] = v \int_0^w \lambda(\tilde{w}) \, dF(\tilde{w}). \tag{7'}$$

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 (8) we now specify

$$\psi(w) = \int \left[1 - \frac{1-(1-p(w))^{h(\tilde{\mathbf{w}}|w)}}{p(w)h(\tilde{\mathbf{w}}|w)} \prod_{\tilde{w}_j > w} [1 - p(\tilde{w}_j)]\right] d\hat{G}(\tilde{\mathbf{w}}|w). \tag{8$'$}$$

The product $\prod_{\tilde{w}_j > w}[1 - p(\tilde{w}_j)]$ describes the probability that the applicant does not take a job at a strictly better wage. Its multiplier gives the probability that a worker does not turn down a job offer because of a job at another firm with the same wage, conditional on failing at higher wages (see e.g. Burdett, Shi and Wright (2001), equation (6)). Then the integrand gives the probability that the worker takes the job at a different firm.

The definitions for all other variables, i.e. μ, p and η and π remain unchanged. With these adjustments the equilibrium definition extends to this section.

Proof of Proposition 10:

Consider some (candidate) equilibrium with some $\boldsymbol{\gamma}$ that we fix for the moment. Denote by $\hat{\imath}$ the highest integer for which $\gamma_i > 0$. By straightforward generalization of Proposition 1 we have

$$p(w) = 1 \qquad \forall w \in [0, \bar{w}_0], \text{ and} \tag{41}$$
$$p(w)w + (1-p(w))u_{i-1} = u_i \quad \forall w \in [\bar{w}_{i-1}, \bar{w}_i] \quad \forall i \in \{1, ..., N\}, \tag{42}$$

where $u_i \equiv \max_{w \in [0,1]} p(w)w + (1-p(w))u_{i-1}$ for all $i \in \{1, 2, ...\hat{\imath}\}$, $u_0 \equiv 0$, $\bar{w}_0 = u_1$. The Market Utility Assumption implies that workers cannot receive more than the Market Utility, which implies that $u_i - u_{i-1} = c_i$ for $i > \hat{\imath}$. 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 $[\bar{w}_i, \bar{w}_{i+1}]$ is empty.

Using (42), 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

$$\pi(\mu) = (1 - e^{-\mu})(1 - u_{i-1}) - \mu(u_i - u_{i-1}), \tag{43}$$

where $\mu = \mu(w)$. The logic is similar to (17). 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 (43) is strictly concave, it is also immediate that all firms within the same interval offer the same wage, yielding exactly $\hat{\imath}$ wages when some workers send $\hat{\imath}$ applications.

We call the group of firms that ends up offering the i'th highest wage as group i and index all their variables accordingly. It is 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 (43), which are given by

$$u_i - u_{i-1} = e^{-\mu_i}(1 - u_{i-1}). \tag{44}$$

Therefore (43) can be rewritten as

$$\pi_i = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - u_{i-1}). \tag{45}$$

Free entry implies that $\pi_i = K$, which together with (44) implies that $\mu_i = \mu_i^*$ and $u_i = u_i^*$ as defined in the main body. By a similar argument as for (25) and (26) the condition $\pi_i = K$ defines for a given vector $\boldsymbol{\gamma}$ of applications the unique measure v_i of firms in each group. Existence and – except for the case where $c_{i^*} = u_i^* - u_{i-1}^*$ – uniqueness follow by similar arguments as in the case of at most two applications.

To show constrained efficiency, we first consider search efficiency for given $\boldsymbol{\gamma}$ and v. Let $\hat{\imath}$ still denote the maximum number of applications that workers send. We first consider a planner that only uses $\hat{\imath}$ wages and, thus, $\hat{\imath}$ groups of firms, with $\rho_i v$ firms in each groups. That is, a worker who applies to i firms applies once to each of the lowest i wages and accepts an offer from a higher wage firm over an offer from a lower wage firm. We call an allocation of firms across groups that leads to the maximum number of matches $\hat{\imath}$-group-efficient. Compare two adjacent groups of firms 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

$$\max_{\rho \in [0,1]} M(\rho) = \nu\rho(1 - e^{-\mu_i}) + \nu(1-\rho)(1 - e^{-\mu_{i-1}}). \tag{46}$$

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

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

We can use similar substitutions as for (29), with the adjustment that now $d\mu_i/d\rho = -\mu_i/\rho = -\nu\mu_i^2/[(1-\psi_i)\Gamma_i]$, 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 obtain the following first order condition

$$\frac{dM(\rho)}{\nu d\rho} = (1 - e^{-\mu_i} - \mu_i e^{-\mu_i})(1 - e^{-\mu_{i-1}}) - (1 - e^{-\mu_{i-1}} - \mu_{i-1}e^{-\mu_{i-1}}) = 0. \tag{47}$$

For given ν this uniquely characterizes the optimal interior ρ, since similar substitutions as above yield $d^2M/d\rho^2 = -\nu[\mu_2^2 e^{-\mu_2}(1-e^{-\mu_1})/\rho + e^{-\mu_1}(1-e^{-\mu_2}-\mu_2 e^{-\mu_2}-\mu_1)^2/(1-\rho)] < 0$. A similar construction as in the proof of Proposition 6 shows that $\hat{\imath}$ groups are sufficient to achieve the constrained optimal outcome.

Next, we 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. Let $\boldsymbol{\rho}(v) = (\rho_1(v), \rho_2(v), .., \rho_{\hat{i}}(v))$ be the fraction of firms in each of the $\hat{i}$ groups under constrained optimal search given v and $\boldsymbol{\gamma}$. Again let $M^*(\boldsymbol{\gamma}, v, \boldsymbol{\rho}(v))$ denote the constrained efficient number of matches given v and $\boldsymbol{\gamma}$. Similar to (30) the objective function is given by $\max_{v\geq 0} M^*(\boldsymbol{\gamma}, v, \boldsymbol{\rho}(v)) - vK$. When $\hat{i} > 0$, then $K < 1$ ensures that the optimal solution is in the interior of $[0, V]$. We 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 $\rho_i(v)$ of firms in each group on the measure of matches can be neglected, i.e. $\frac{dM^*}{d\rho_i}\frac{d\rho_i}{dv} = 0$ at the $\hat{i}$-group-efficient ρ_i. We get as first order condition

$$dM^*(\boldsymbol{\gamma}, v, \boldsymbol{\rho})/dv = K, \tag{48}$$

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

$$\frac{dM^*(\boldsymbol{\gamma}, v)}{dv} = \sum_{i=1}^{\hat{i}}\left[\gamma_i \sum_{j=1}^{i}\left[\frac{dp_j}{dv}\prod_{\substack{k\leq i\\ k\neq j}}(1-p_k)\right]\right] = \sum_{i=1}^{\hat{i}}\left[\frac{dp_i}{dv}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)\right], \tag{49}$$

where the equality line is obtained by rearranging the terms for each dp_i/dv. To simplify notation, define the partial sum

$$\xi_{i'} = \sum_{i\geq i'}^{N}\left[\frac{dp_i}{dv}\Gamma_i(1-\psi_i)\prod_{k<i}(1-p_k)\right]. \tag{50}$$

Since $p_{\hat{i}} = (1-e^{-\mu_{\hat{i}}})/\mu_{\hat{i}}$ we have $dp_{\hat{i}}/dv = -(1/\mu_{\hat{i}}^2)(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})(d\mu_{\hat{i}}/dv)$. Since $\mu_{\hat{i}} = \gamma_{\hat{i}}/(\rho_{\hat{i}}v)$, we have $d\mu_{\hat{i}}/dv = -\gamma_{\hat{i}}/(\rho_{\hat{i}}v^2) = -\rho_{\hat{i}}\mu_{\hat{i}}/\gamma_{\hat{i}}$. So we get $dp_{\hat{i}}/dv = -\rho_{\hat{i}}(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})/\gamma_{\hat{i}}$. Noting that $\Gamma_{\hat{i}}(1-\psi_{\hat{i}}) = \gamma_{\hat{i}}$, we have established that

$$\xi_{\hat{i}} = \rho_{\hat{i}}(1-e^{-\mu_{\hat{i}}} - \mu_{\hat{i}}e^{-\mu_{\hat{i}}})\prod_{k<\hat{i}}(1-p_k). \tag{51}$$

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

Lemma A1 *For all i it holds that*

$$\xi_i = \left(\sum_{k=i}^{N}\rho_k\right)(1-e^{-\mu_i} - \mu_i e^{-\mu_i})\prod_{j<i}(1-p_k). \tag{52}$$

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 (47) 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 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 $i \in \{1, ..., \hat{i} - 1\}$:

$$\xi_{i+1} = (\sum_{k=i+1}^{\hat{i}} \rho_k)(1 - e^{-\mu_{i+1}} - \mu_{i+1} e^{-\mu_{i+1}}) \prod_{k<i+1} (1 - p_k). \tag{53}$$

It clearly holds for $i = \hat{i} - 1$ by (51). Now assume it holds for some i. We consider ξ_i. We know that

$$\xi_i = \xi_{i+1} + \Gamma_i (1 - \psi_i) \frac{dp_i}{dv} \prod_{k<i} (1 - p_k). \tag{54}$$

The second summand can be written as

$$\frac{dp_i}{dv} \prod_{k<i} (1 - p_k) = -\frac{1 - e^{-\mu_i} - \mu_i e^{-\mu_i}}{\mu_i^2} [\frac{d\mu_i}{dv} \prod_{k<i} (1 - p_k)] \tag{55}$$

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 (55) as

$$\frac{d\mu_i}{dv} \prod_{k<i} (1 - p_k) = \frac{\lambda_i \ \xi_{i+1}}{\Gamma_i (1 - p_i)} - \frac{\Gamma_i (1 - \psi_i)}{\rho_i v^2} \prod_{k<i} (1 - p_k) = \frac{\lambda_i \ \xi_{i+1}}{\Gamma_i (1 - p_i)} - \frac{\rho_i \mu_i^2}{\Gamma_i (1 - \psi_i)} \prod_{k<i} (1 - p_k).$$

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 (55) and multiply by $\Gamma_i (1 - \psi_i)$ to get

$$\Gamma_i (1 - \psi_i) \frac{dp_i}{dv} \prod_{k<i} (1 - p_k) = \frac{p_i - e^{-\mu_i}}{1 - p_i} \xi_{i+1} + \rho_i (1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) \prod_{k<i} (1 - p_k).$$

We can substitute this into (54), and use (53) and the property of $\hat{i}$-group-efficient search in (47) to obtain

$$\xi_i = (\sum_{k=i}^{N} \rho_k)(1 - e^{-\mu_i} - \mu_i e^{-\mu_i}) \prod_{k<i} (1 - p_k). \tag{56}$$

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 (as would happen if all firms offered the same wage), we 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 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 $\sigma(i^*) < \alpha$ 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 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 costs 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 cost. 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 $u^*_{i^*} - u^*_{i^*-1} \to_{i^* \to \infty} 0$. Due to increasing marginal costs and optimality of sending $i*$ applications we have $c^{i^*}_i \leq c^{i^*}_{i^*} \leq u^*_{i^*} - u^*_{i^*-1}$ for all $i \leq i^*$, which which together with $u^*_{i^*} - u^*_{i^*-1} \to_{i^* \to \infty} 0$ 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 (1999): "Holdups and Efficiency with Search Frictions." *International Economic Review*, 40, 827-849.

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

[3] Albrecht, James, Pieter A. Gautier, Serene Tan, and Susan Vroman (2005): "Matching with Multiple Applications Revisited," *Economic Letters*, 84, 311-314.

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

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

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

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

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

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

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

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

[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] Damien Gaumont, Martin Schindler and Randall Wright (2006): "Alternative Theories of Wage Dispersion," *European Economic Review*, 50, 831848.

[16] Dutta, Bhaskar and Matthew Jackson (2000): "The Stability and Efficiency of Directed Communication Networks," *Review of Economic Design*, 5, 251272.

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

[18] Hosios, Arthur J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57, 279-298.

[19] Jackson, Matthew O. (2005): "A Survey of Models of Network Formation: Stability and Efficiency," in Gabrielle Demange and Myrna Wooders (eds.): *Group Formation in Economics: Networks, Clubs, and Coalitions*, 2005, Cambridge University Press, Cambridge.

[20] Kircher, Philipp (2007): "Technical Appendix to 'Efficiency of Simultaneous Search'," mimeo.

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

[22] Moen, Espen R. (1995): "Essays on Matching Models of the Labour Market," Ph.D. dissertation, London School of Economics.

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

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

[25] Mortensen, Dale T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, Zeuthen Lecture Book Series.

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

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

[32] 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.

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

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

[35] Shi, Shouyong (2006): “Directed Search for Equilibrium Wage-Tenure Contracts,” mimeo.

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

[37] Shimer, Robert (2004): “Search Intensity,” mimeo.

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