Ex Ante and Ex Post Inefficiency in Search and Matching Models*

Ronald P. Wolthoff†
University of Chicago

Preliminary

February 14, 2009

Abstract

I present a directed search model of the labor market in which workers can send multiple applications and firms can make multiple job offers. The model incorporates many existing models as special cases, including the standard directed search model with one application and one job offer and the stable matching outcome. I characterize the equilibrium and argue that the number of applications sent and the number of job offers made generate enough variation in the equilibrium outcomes to structurally estimate both parameters. In that way, the model can answer the question whether ex ante frictions (i.e. on the worker side) or ex post frictions (i.e. on the firm side) contribute more to the output loss compared to the frictionless Walrasian world.

Keywords: search and matching, stable matching, simultaneous search, directed search, inefficiency, frictions
JEL codes: J64, J31, D83

1 Introduction

In this paper I present a directed search model of the labor market in which workers can send multiple applications and firms can make job offers to multiple applicants. The main contribution of the paper is to show that both parameters have distinct effects on the equilibrium outcomes. Estimation of the model on wage and duration data can therefore shed new light on the matching process of workers and firms.

*I thank Pieter Gautier, Ali Hortacsu, Philipp Kircher and Robert Shimer for valuable comments.
†University of Chicago, Department of Economics, 1126 E. 59th Street, Chicago, IL 60637, USA, wolthoff@uchicago.edu

The question how the demand and supply side meet and match is a fundamental one in the analysis of economic markets. Traditional Walrasian models rely on a hypothetical auctioneer who registers demand and supply, and adjusts the price until the market-clearing price is reached. In the early sixties, models were developed with more explicit descriptions of the meeting process. Examples include Stigler (1961, 1962), Karlin (1962), and Gale and Shapley (1962). Since then, a large search and matching literature has developed with a considerable diversity in model specifications. One important source of variation is the number of potential contacts between buyers and sellers. In the seminal (male-optimal) Gale and Shapley (1962) algorithm, men propose to one woman at the time and women are engaged to at most one men at the time. However, men can make as many proposals as they need to find a partner, which effectively means that there are no frictions and that all men can contact all women and vice versa. As a result, a stable matching obtains, i.e. a pairing in which no man prefers a different wife that also prefers him over her current husband.

Directed search models in the spirit of Moen (1997), Acemoglu and Shimer (1999a,b), and Burdett et al. (2001) are very different. Buyers still observe all sellers and the prices that they charge, but they can contact only one of them.[1] Likewise, each seller can only contact one potential buyer. As a result of theses frictions the equilibrium outcome is fundamentally different. For example, on both sides of the market some agents remain unmatched involuntarily. The assumption of one contact per seller is not restrictive in these models. Sellers never need to approach a second potential buyer, since they have unit capacity only and the first buyer will always be willing to trade. However, the assumption is typically maintained in richer models. Especially in the job search literature some progress has been made on analyzing models in which workers (buyers) can apply to multiple but a finite number of firms (sellers). Examples of such contributions are Albrecht et al. (2006) and Galenianos and Kircher (2008). In such a setting, the assumption that a firm can contact one applicant only is much more restrictive. The firm might lose the applicant to a competing firm that offered a higher wage. In that case the firm would certainly like to make an offer to a potential second applicant.[2] Nevertheless, literature on such recall of applicants is limited. The only exception is Kircher (2007) who studies a model in which firms can contact all applicants if necessary. In other words, firms offer the job to one applicant at the time, but can go to applicant $n+1$ if applicant n rejects. This process continues until an applicant accepts, or the firms runs out of applicants.

Which model is the right one may vary across markets. In some markets there are few frictions, since the cost of both an application and an offer are small.[3] In that case the Gale and Shapley (1962) outcome might be an adequate description. Potential examples of such markets include financial markets (see e.g. Sorensen, 2007) and dating websites (see Hitsch et al., 2008). In other markets, frictions are typically larger. For example, in the academic job market the number of applications

[1] This is only explicit in last paper. The other papers do not give a microfoundation for the matching function, but the conditions imposed on the matching function only hold if the setting is similar to Burdett et al. (2001).

[2] In an extension of their main model, Albrecht et al. (2006) discuss how their results change if firms can make a second offer. It turns out that the tractability of this extended model is limited.

[3] The right terminology varies by market. For reasons of consistency, I continue to use labor market terminology.

is typically very high, but the number of offers relatively small, since there is a high cost of recall. The opposite pattern is observed in the housing market. Buyers typically make a bid at one house at a time, but sellers can relatively easily contact other bidders if the negotiations with the first bidder break down.

In this paper, I focus on the labor market. However, even within such a well-defined market, the level of frictions may vary across different submarkets. The job market for assistant professors has very different characteristics than the recruiting process of e.g. steel workers. Hence, the exact extend to which contacts from both sides of the market are possible and the exact level of frictions in a specific (sub)market are empirical questions. The model that I present here makes it feasible to answer those questions. I derive the equilibrium outcomes for a model in which workers send an arbitrary number of applications and firms make an arbitrary number of offers. The model therefore incorporates many of the above models as special cases. Estimation of the model would then show which of those models is empirically relevant in the specific market that I study. Moreover, having estimates of the number of applications and the number of offers makes it possible to calculate the relative impact of both on search frictions and the output loss compared to a Walrasian world. So far, it is an open question whether ex ante frictions (i.e. on the worker side) or ex post frictions (i.e. on the firm side) are a larger source of inefficiency.

A few other papers are related to what I do here. In the competing auction model of Julien et al. (2000, 2006) workers do not search themselves but are contacted randomly by firms. This is basically equivalent to a setting in which all workers apply to all firms, but in which firms can again make one offer only. Gautier and Moraga-González (2005) and Kaas (2007) present random search models of the labor market with multiple applications per worker and one offer per firm. Gautier et al. (2007) present a model in the same spirit and structurally estimate it in order to assess the level of inefficiency in the labor market. Gautier and Teulings (2006) also try to quantify search frictions, but in an assignment model. Parallel to the search literature, a large literature on matching has been developed with extensions, refinements and applications of the original model by Gale and Shapley (1962). See Roth (2008) for a recent overview.

The structure of the paper is as follows. Section 2 presents the model. The equilibrium is characterized in section 3. Section 4 discusses the empirical content of the model and section 5 concludes. Proofs are relegated to the appendix.

2 Model

2.1 Setting

Consider a labor market with a measure 1 of workers and a positive measure θ of firms. Both types of agents are risk-neutral and maximize their expected payoff. To keep the exposition as simple as possible, I initially focus on a one-period model. Extension to the dynamic case is relatively straightforward.[4] All firms and workers are identical. All workers supply one indivisible unit of labor, whereas each firm has a position that can be filled by exactly one worker. A worker that matches with a firm receives a wage w. The firm receives the difference between the output that is produced, which is normalized to 1, and the wage w that it pays.

I assume that unemployed workers and firms with vacancies meet each other via a process of directed search. Firms with a vacancy post a wage w to which they commit. After that an application phase starts. Workers observe all vacancies and their associated wages and decide to which ones they want to apply. All workers send A applications. It is straightforward to extend the model to allow for heterogeneity in the number of applications, as shown in Kircher (2007).

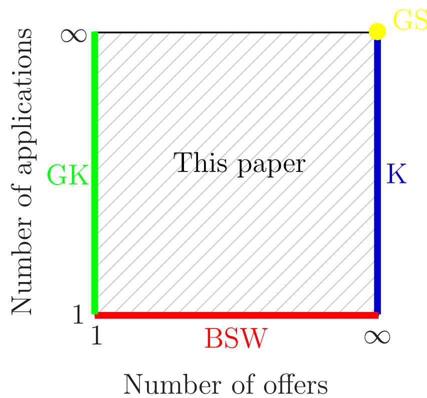
After the applications are sent, the recruiting phase starts. The recruiting process consists of several rounds as well. In the first round, each firm randomly selects one of it applicants, if any, and offers her the job. The applicant in question might get multiple offers in this first round. She compares them, tentatively accepts the best one and rejects all others. Then the second round starts for the firms that are not tentatively matched with a worker. A fraction τ of them experiences a termination shock, which implies that they cannot continue recruiting and reject all remaining applicants. The firms not hit by the shock, i.e. a fraction $1-\tau$, again randomly selects one of its (remaining) applicants and offers her the job. Workers again compare all offers, potentially including the one tentatively accepted in the previous round, accept the best one for the moment and reject all other. The process continues until no firm can make an offer anymore. So, the number of offers that a firm can potentially make follows a geometric distribution with termination probability τ. All offers that are tentatively accepted at the end of the recruiting phase become matches between the worker and firm in question.

Abstract for the moment from the fact that search is directed. Then the matching technology that I present here includes several models from literature as special cases. First of all, the model corresponds with the seminal stable matching outcome of Gale and Shapley (1962) for $A \to \infty$ and $\tau = 0$. This case, which has been applied extensively in the analysis of many different markets, allows all workers to apply to all firms and all firms to contact all workers. If one thinks of the labor market as a network, this would be equivalent to a complete bipartite graph.

The graph is always bipartite in our model, but not necessarily always complete. The number of links can be significantly smaller if $A < \infty$ and $\tau > 0$. During the application phase a certain number of links are generated between the workers and the firms. During the recruiting phase a random number of those links is destroyed while the other ones are kept. The resulting matching is a stable matching on this sparser graph. If $A = 1$ and $\tau = 1$, all workers send one application and all firms can make one offer. This is the standard urn-ball matching technology, which was introduced by Butters (1977) and has been used extensively in the labor literature, for example in Burdett et al. (2001). Julien et al. (2000) study a model in which workers apply to all firms, and firms can bid for the services of one worker, i.e. $A \to \infty$ and $\tau = 1$. Albrecht et al. (2006) also impose $\tau = 1$, but study the case $A = 2$.[5]

[4] See Galenianos and Kircher (2005, 2008).

[5] In an extension of their model, they briefly discuss what happens if firms can contact a second applicant.



The vertical axis shows the number of applications A that workers send out. The horizontal axis displays the expected number of offers $\frac{1}{\tau}$ that a firm can make. GS = Gale and Shapley (1962), BSW = Burdett et al. (2001), GK = Galenianos and Kircher (2008), K = Kircher (2007). In Burdett et al. (2001), firms never make more than one offer, since the first offer always gets accepted. The equilibrium outcome does therefore not depend on the number of offers that firms can make.

Figure 1: Related literature

Galenianos and Kircher (2008) and Gautier and Moraga-González (2005) both describe models in which workers send A applications and no recall is possible (i.e. $\tau = 1$). Gautier et al. (2007) and Kaas (2007) also exclude the possibility of recall, but allow for heterogeneity in A among the workers. Finally, Kircher (2007) studies a model with A applications and perfect recall ($\tau = 0$). Figure 1 shows how this paper relates to some of these models.

In section 3 I derive the workers' and firms' optimal strategies. I focus on symmetric equilibria, i.e. identical agents have identical strategies. Further, I require the strategies to be anonymous, in the sense that they cannot be conditioned on the identity of a specific worker or firm.

2.2 Expected Payoffs

Firms aim to maximize expected profit, which we define as the difference between output produced and the wage that it pays to its worker. I normalize output to 1. Let $\eta(w)$ denote the matching probability of a firm posting a wage w. This probability is endogenously determined in equilibrium by the firms' and workers' strategies. Given this, expected profit for the firm equals

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

Hence, the strategy for a firm is the wage w that it posts. Denote the distribution of wage offers by F and its support by $\mathscr{F}$. Since wages below zero or above one can never be optimal, $\mathscr{F} \subset [0,1]$.

Next, consider the workers. They observe all posted wages and decide where to apply. As a result of the anonymity assumption, workers apply to all firms posting the same wage with equal probability. Hence, the strategy for a worker can be summarized by vector of wages $\mathbf{w} = (w_1, w_2, \ldots, w_A)$, indicating to what wage levels she wants to send her applications. Without loss of generality I order the elements of the vector $\mathbf{w}$ in such a way that $w_1 \leq \ldots \leq w_A$. I denote the distribution of workers' strategies by G. In other words $G(\widetilde{\mathbf{w}})$ denotes the probability that the worker sends her first application to a wage below $\widetilde{w}_1$, her second application to a wage below $\widetilde{w}_2$, et cetera. Since workers can only apply to wage that are actually offered, G indirectly depends on F. The support of G is denoted by $\mathscr{G}$.

Let $\psi(w)$ denote the (endogenous) probability that an application to a firm offering w results in a job offer during any of the firm's recruiting rounds. A worker accepts the best job offer that she gets. Hence, a worker applying to $\mathbf{w}$, accepts w_i if she gets a job offer from that firm, which happens with probability $\psi(w_i)$, and if none of the applications to higher wages resulted in an offer. The probability that this happens equals $\prod_{j=i+1}^{A}(1-\psi(w_j))\,\psi(w_i)$. A worker who fails to get a job offer, gets a payoff equal to her outside option, which I denote by u_0. This occurs with probability $\prod_{j=1}^{A}(1-\psi(w_j))$. Hence, the worker's expected payoff can be written as follows

$$u(\mathbf{w}) = \sum_{i=1}^{A}\prod_{j=i+1}^{A}(1-\psi(w_j))\,\psi(w_i)\,w_i + \prod_{j=1}^{A}(1-\psi(w_j))\,u_0. \tag{2}$$

3 Equilibrium

3.1 Queue Lengths and Equilibrium Definition

It is well known that in the analysis of search models with a continuum of agents the so-called queue length is a key variable. In general there is a one-to-one relationship between the wage that a firm posts and this variable that measures the expected number of applications that this firm receives.[6] In this paper not one but three different queue lengths play a role. The first queue length, i.e. $\lambda(w)$, is the standard one, defined as the total number of applications sent to firms posting a wage w divided by the number of firms doing so. I refer to $\lambda(w)$ as the *total queue length*. It is this queue length that determines the job offer probability in a world without recall. The second queue length is denoted by $\mu(w)$ and represents the expected number of applicants per vacancy that would actually accept a job offer from this firm since they do not get a better offer. As shown by Kircher (2007), this is the relevant notion in a world with perfect recall, in which case a (constrained) efficient outcome is obtained. Therefore, I call $\mu(w)$ the *efficient queue length*.[7] Let $\Psi(w)$ denote the fraction of applicants that rejects a wage w, because they getter a better job offer. Then, $1-\Psi(w)$ is called the retention probability, which links the total queue length to the efficient queue length in the following way:

$$\mu(w) = \lambda(w)(1-\Psi(w)). \tag{3}$$

[6] See for example Acemoglu and Shimer (1999b), Burdett et al. (2001), Galenianos and Kircher (2008), and Kircher (2007). Alternatively, some authors write their models in terms of the labor market tightness, which in case of sequential search or $A = 1$ is simply the inverse of the queue length. See Moen (1997) for a seminal example.

[7] Kircher (2007) calls $\mu(w)$ the *effective queue length*. I reserve this name for the third queue length.

In this paper, I consider partial recall, which means that $\lambda(w)$ and $\mu(w)$ are limiting cases of a more general concept. Therefore, I introduce a third queue length which I denote by $\kappa(w)$ and which is a weighted average of the total and the efficient queue length with τ and $1-\tau$ being the respective weights

$$\kappa(w) = \tau\lambda(w) + (1-\tau)\mu(w). \tag{4}$$

As I will show below, the job offer probability in my model solely depends on $\kappa(w)$. For that reason, I call $\kappa(w)$ the *effective queue length*. In order to derive expressions for the workers' job offer probability $\psi(w)$ and the firms' matching probability $\eta(w)$, I consider a firm posting a wage w and I assume for the moment that the total queue length $\lambda(w)$ and the efficient queue length $\mu(w)$ are exogenously given. This also pins down the queue length $\nu(w) = \lambda(w) - \mu(w)$ of applicants not willing to accept a job offer. Given these queue lengths, the actual number of interested applicants m and the actual number of uninterested applicants n follow a Poisson distribution with means equal to $\mu(w)$ and $\nu(w)$ respectively. This observation allows me to derive an expression for both the probability η that the firm will match and the probability ψ that a worker who applies to the firm gets a job offer.

Lemma 1 *Consider a firm posting a wage w with total queue length $\lambda(w)$ and efficient queue length $\mu(w)$. The firm hires with probability*

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

A worker that applies to the firm gets a job offer with probability

$$\psi(w) = \frac{1}{\kappa(w)}\left(1 - e^{-\kappa(w)}\right). \tag{6}$$

The lemma shows that the job offer probability $\psi(w)$ and the firm's matching probability $\eta(w)$ indeed have the same structure as in standard direct search models. The only difference concerns the queue length that is used. The job offer probability solely depends on the effective queue length and the matching probability depends on the effective and the efficient queue length. Note that the definition of $\kappa(w)$ indeed incorporates the total and the efficient queue length as limiting cases. If $\tau = 1$, the recruiting process terminates for sure after the first round and $\kappa(w) = \lambda(w)$. This leads to the standard case in which the job offer probability solely depends on the total queue length $\lambda(w)$ and the firm's matching probability depends on $\lambda(w)$ and $\mu(w)$. On the other hand, if $\tau = 0$, firms can continue contacting applicants until they have exhausted their entire list, i.e. $\kappa(w) = \mu(w)$. This is the perfect recall case analyzed in Kircher (2007) in which both probabilities only depend on the efficient queue length.

It is important to stress that the scope of lemma 1 is not limited to directed search models, since the lemma describes the recruiting phase. The applications that workers have sent during the application phase have created a network. The structure that is imposed by the symmetry and anonymity assumption imply that this network can be fully characterized by two variables only: the total queue length $\lambda(w)$ and the efficient queue length $\mu(w)$. Lemma 1 takes these variables as given and therefore holds for other application processes, e.g. random search, as well.

The application process enters in deriving expressions for the total and the efficient queue length. For example, under random search the total queue length would be independent of the wage offered by a firm, i.e. $\lambda(w) = \lambda = \frac{A}{\theta}$.[8] This obviously changes if workers can direct their search, as is the case here. Then, the total queue length is formally defined by the integral equation

$$\theta \int_0^w \lambda(\omega)\, dF(\omega) = \sum_{i=1}^{A} G_i(w),$$

where $G_i(w)$ denotes the marginal distribution of $G(\mathbf{w})$ with respect to w_i. The right hand side denotes the total mass of applications that are send by the workers to wages no higher than w. The left hand side represents the mass of applications received by firms posting a wage no higher than w. Obviously, both masses need to be the same for each possible value of w.

Equation 3 shows that knowing the total queue length at a certain wage w, only the retention probability is required to calculate the efficient queue length. Consider a worker who gets a job offer w with his i-th application. Let $\hat{G}_i(w_{-i}|w)$ denote the conditional distribution of the remaining applications, given that the i-th application was sent to a w. The worker will accept the wage offer w if and only if all applications sent to higher wages do not result in a job offer. This is the case with probability $Q_i(w_i, w_{-i}) = \prod_{j>i}(1-\psi(w_j))$. Finally, let $P(i|w)$ denote the conditional probability that the worker who got the wage offer w applied there with his i-th application. Then the retention probability equals

$$1-\Psi(w) = \sum_{i=1}^{A} P(i|w) \int Q_i(w, w_{-i})\, d\hat{G}_i(w_{-i}|w).$$

Given the retention probability, equations 3, (4), (5), and (6) determine the workers' job offer probabilities and the firms' matching probabilities for any wage w. Subsequently, equations (1) and (2) determine the firms' profit and the workers' utility for any strategy. I now define an equilibrium as follows.

Definition 1 *An equilibrium is a set of strategies* $\{F, G\}$ *such that*

1. *Profit maximization:* $\pi(w) = \pi^* \equiv \max_{w' \in [0,1]} \pi(w')$ *for all* $w \in \mathscr{F}$.
2. *Utility maximization:* $u(\mathbf{w}) = u^* \equiv \max_{\mathbf{w}' \in [0,1]^A} u(\mathbf{w}')$ *for all* $\mathbf{w} \in \mathscr{G}$.

Hence, in equilibrium no firm must be able to make larger profits by deviating from F. Likewise, no worker must be able to obtain a higher utility by applying in a way that differs from G.

[8] It is well known that the equilibrium wage offer density $f(w)$ has continuous support in that case (see Burdett and Judd, 1983; Gautier and Moraga-González, 2005). For a worker to be willing to accept a job offer w, none of the $A-1$ other applications may result in a higher wage offer. Hence, the efficient queue length equals $\mu(w) = \lambda(w)\left(1-\int_w^{\bar{w}} \psi(\omega) f(\omega)\, d\omega\right)^{A-1}$, where $\bar{w}$ denotes the upper bound of the support of $f(w)$.

3.2 Workers' Application Behavior

I start the derivation of the equilibrium by considering the application decision of a worker. Since the worker side of the model is similar to Galenianos and Kircher (2008) and Kircher (2007), this section resembles their analysis.

A worker observes all wages and in equilibrium she knows what job offer probability ψ is associated with each wage. Given this information, she selects A wage levels to apply to. Formally, she solves the following maximization problem

$$\max_{(w_1,\ldots,w_A)} \sum_{i=1}^{A} \prod_{j=i+1}^{A} (1-\psi(w_j))\,\psi(w_i)\,w_i + \prod_{j=1}^{A} (1-\psi(w_j))\,u_0.$$

Note that this optimization problem has a recursive structure. The application to a firm offering w_i is only relevant if all applications to higher wage firms, i.e. the applications $w_{i+1},\ldots,w_A$, fail to result in a job offer. This simplifies the derivation of the workers' optimal strategy. Define u_n in the following way

$$u_n \equiv \max_{(w_1,\ldots,w_n)} \sum_{i=1}^{n} \prod_{j=i+1}^{n} (1-\psi(w_j))\,\psi(w_i)\,w_i + \prod_{j=1}^{n} (1-\psi(w_j))\,u_0.$$

Then the workers optimal choice for the i-th application ($i \in \{1,\ldots,A\}\}$) solves

$$\max_{w} \psi(w)\,w + (1-\psi(w))\,u_{i-1}.$$

Let $\bar{w}_i$ denote the highest wage that yields utility u_i if the outside option equals $\bar{u}_{i-1}$, i.e.

$$\bar{w}_i = \sup\{w \in [0,1]\,|\,\psi(w)\,w + (1-\psi(w))\,u_{i-1} = u_i\} \;\forall i \in \{1,\ldots,A\}$$

Additionally, define $\bar{w}_0$ as the lowest wage that would receive applications. The following lemma states that the values $\bar{w}_i$ bound the interval to which application i can be sent in equilibrium.

Lemma 2 *For any distribution of posted wages, the optimal application strategy for any worker is to send application i to a wage in the interval $[\bar{w}_{i-1},\bar{w}_i]$ for $i \in \{1,\ldots,A\}$.*

Since this lemma restricts the workers' application behavior, it has implications for the job offer probability $\psi(w)$ in equilibrium. For example, no worker will apply to wages below $\bar{w}_0$. Therefore the queue length at firms offering such wages equals zero and the corresponding job offer probability equals 1. Firms posting wages in $[\bar{w}_0,\bar{w}_1]$ receive the first application of each worker. This application must yield a utility u_1 to the worker, hence the job offer probability is such that the condition $\psi(w)\,w + (1-\psi(w))\,u_0$ holds. In a similar manner conditions for the remaining intervals can be derived. This is summarized in the following lemma.

Lemma 3 *In any equilibrium, the job offer probabilities $\psi(w)$ satisfy the following conditions*

$$\begin{aligned}\psi(w) &= 1 \ \forall w \in [0, \bar{w}_0] \\ \psi(w)\, w + (1-\psi(w))\, u_{i-1} &= u_i \ \forall w \in [\bar{w}_{i-1}, \bar{w}_i] \ \forall i \in \{1, \ldots, A\}\end{aligned}$$

3.3 Firms' Wage Setting

I now consider the firm's strategy. The firm chooses what wage to post, taking into account the behavior of the other firms as well as the application behavior of the workers. The firm's goal is to maximize its expected profit, which is defined to be equal to the firm's matching probability times the difference between output and the wage it posts. Hence, the firm solves

$$\max_{w} \eta(w)(1-w).$$

Equation (5) shows that $\eta(w) = \frac{\mu(w)}{\kappa(w)}\left(1-e^{-\kappa(w)}\right)$. Firms that post the highest wage in the market know for sure that their wage offer will be accepted. As a result, $\lambda(w) = \mu(w) = \kappa(w)$ must hold for these firms. In that case, $\eta(w)$ reduces to $\left(1-e^{-\kappa(w)}\right)$ and the firm solves

$$\begin{aligned}&\max_{w \in [\bar{w}_{A-1}, \bar{w}_A]} \left(1-e^{-\kappa(w)}\right)(1-w) \\ \text{s.t.} \quad &\psi(w)\, w + (1-\psi(w))\, u_{A-1} = u_A\end{aligned} \tag{7}$$

Rewriting the constraint yields $w = u_{A-1} + \frac{u_A - u_{A-1}}{\psi(w)}$ and substituting this into the objective function gives an expression in $\kappa(w)$ only. Hence, we can maximize over the queue length κ rather than the wage w

$$\max_{\kappa > \bar{\kappa}_{A-1}} \left(1-e^{-\kappa}\right)(1-u_{A-1}) - \kappa(u_A - u_{A-1}),$$

where $\bar{\kappa}_{A-1} = \kappa(\bar{w}_{A-1})$. This function is strictly concave in κ, which means that the optimal value κ_A^* is either determined by the solution $\hat{\kappa}_A$ to the first order condition or equals the boundary value $\bar{\kappa}_{A-1}$.

Consider a type $A-1$ firm next. This firm has to take into account that it might lose its applicant to a type A firm. By lemma 3 we know that the probability that the firm loses its candidate is constant within the interval $[\bar{w}_{A-2}, \bar{w}_{A-1}]$ and equal to

$$\Psi(w) = \Psi_{A-1} \equiv \psi(w_A^*) \ \forall w \in [\bar{w}_{A-2}, \bar{w}_{A-1}].$$

This implies that for all wages in the interval $[\bar{w}_{A-2}, \bar{w}_{A-1}]$, the ratio $\frac{\mu(w)}{\kappa(w)}$ can be written as

$$\frac{\mu(w)}{\kappa(w)} = \frac{\lambda(w)(1-\Psi_{A-1})}{\tau\lambda(w) + (1-\tau)\lambda(w)(1-\Psi_{A-1})} = \frac{1-\Psi_{A-1}}{1-(1-\tau)\Psi_{A-1}},$$

which is independent of the exact value of w. Hence, the first factor in the profit maximization problem serves as a constant and the firm effectively solves

$$\max_{w \in [\bar{w}_{A-2}, \bar{w}_{A-1}]} \left(1 - e^{-\kappa(w)}\right)(1-w)$$
$$\text{s.t.} \quad \psi(w)\, w + (1-\psi(w))\, u_{A-2} = u_{A-1},$$

which closely resembles (7). The same argument applies for firms of type $i \in \{1, \ldots, A-2\}$, with the retention probability for these firms being equal to

$$1 - \Psi_i = \prod_{j=i+1}^{A} \left(1 - \psi\left(w_j^*\right)\right).$$

Hence, in each interval the optimal effective queue length either equals the interior value $\hat{\kappa}_i$ or the boundary value $\bar{\kappa}_{i-1}$.

Now consider the firms to which workers send their first application. For these firms, the interior solution obtains. Hence, $w_1^* = \hat{w}_1$, $\kappa_1^* = \hat{\kappa}_1 = \kappa(\hat{w}_1)$, and $\mu_1^* = \hat{\mu}_1 = \mu(\hat{w}_1)$. The FOC of the maximization problem implies

$$e^{-\hat{\kappa}_1}(1-u_0) = u_1 - u_0.$$

Substituting this into the profit function yields

$$\pi_1 \equiv \pi(w_1^*) = \frac{1-\Psi_1}{1-(1-\tau)\Psi_1}\left(1 - e^{-\hat{\kappa}_1} - \hat{\kappa}_1 e^{-\hat{\kappa}_1}\right)(1-u_0).$$

Next consider a type 2 firm. If $\kappa_2^* = \hat{\kappa}_2$, then the firm's profit equals

$$\pi(\hat{w}_2) = \frac{1-\Psi_2}{1-(1-\tau)\Psi_2}\left(1 - e^{-\hat{\kappa}_2} - \hat{\kappa}_2 e^{-\hat{\kappa}_2}\right)(1-u_1).$$

Galenianos and Kircher (2008) show that for $\tau = 1$, $\pi(\hat{w}_2)$ is always strictly smaller than π_1. Hence, $\kappa_2^* = \hat{\kappa}_2$ violates the equal profit condition and cannot be part of an equilibrium. Therefore, κ_2^* must equal $\bar{\kappa}_1$. In the same way $\kappa_i^* = \bar{\kappa}_{i-1}$ for all $i \in \{3, \ldots, A\}$. The corresponding wage is $w_i^* = \bar{w}_{i-1}$, which by definition must satisfy

$$\psi(\bar{w}_{i-1})\, \bar{w}_{i-1} + (1-\psi(\bar{w}_{i-1}))\, u_{i-2} = u_{i-1}.$$

Since all firms follow the same strategy $\psi(w_i^*)$ equals $\frac{1}{\bar{\kappa}_i}\left(1 - e^{-\bar{\kappa}_i}\right)$ and profit is given by

$$\begin{aligned}
\pi_i \equiv \pi(w_i^*) &= \frac{1-\Psi_i}{1-(1-\tau)\Psi_i}\left(1 - e^{-\kappa_i^*}\right)(1-w_i^*) \\
&= \frac{1-\Psi_i}{1-(1-\tau)\Psi_i}\left(\left(1 - e^{-\kappa_i^*}\right)(1-u_{i-2}) - \kappa_i^*(u_{i-1} - u_{i-2})\right)
\end{aligned}$$

On the other hand, if $\tau = 0$, then the optimal queue length for a type i firm is given by $\kappa_i^* = \hat{\kappa}_i$, as shown by Kircher (2007). Profits are equal to

$$\pi(\hat{w}_i) = \frac{1-\Psi_i}{1-(1-\tau)\Psi_i}\left(1-e^{-\hat{\kappa}_i}-\hat{\kappa}_i e^{-\hat{\kappa}_i}\right)(1-u_{i-1})$$

and the optimal wage equals

$$\hat{w}_i = \frac{\hat{\kappa}_i e^{-\hat{\kappa}_i}}{1-e^{-\hat{\kappa}_i}}(1-u_{i-1})+u_{i-1}.$$

For intermediate values of τ, as is the case in this paper, such a general result does not exist. For values close enough to 0, the interior solution is the relevant one, whereas the boundary solution is the equilibrium value for τ sufficiently close to 1.

Proposition 1 *An equilibrium exists and it is unique.*

[to be extended]

4 Empirical Content

The equilibrium described in the previous section depends on two fundamental parameters that determine the extent to which contact between workers and firms is possible: the number of applications A and the termination probability τ. In this section, I discuss how change in these two parameters affect the equilibrium outcome, as a first step towards a structural estimation of both parameters.

Consider first some special cases of the model that correspond to models that have described in literature before. If all workers send one application, i.e. $A = 1$, then the value of τ is irrelevant. Each firms knows that any applicant will accept the job for sure, since the applicant cannot receive competing offers. Ex ante competition still takes place since workers observe the wage offers before they apply. This is the standard directed search model, described by Burdett et al. (2001). The equilibrium wage offer distribution consists of a single mass point.

If we allow A to be larger than 1 and fix $\tau = 1$, then the model of Galenianos and Kircher (2008) obtains. As shown in their paper, the equilibrium wage offer distribution has A points of support in this case. Both the density of wage offers and the density of accepted wages are strictly decreasing. The job offer probability $\psi(w)$ is also strictly decreasing in wages, while the firms' matching probability $\eta(w)$ is larger for higher values of w. The latter is necessary to make sure that all firms make the same profit in equilibrium. Finally, high wage firms receive more applications, i.e. the total queue length $\lambda(w)$ strictly increases in w.

These results are shown in figure 2, together with the cases $\tau = 0$ and $\tau = 0.25$. Imposing $\tau = 0$ leads to the model by Kircher (2007). It turns out that in this case the wage offer density and the density of accepted wages have quite a different shape: they are strictly increasing. The degree of wage dispersion has diminished. The job offer probability continues to be smaller for larger wages and the matching probability continues to increase in wages. However, the decline in the job offer probability goes much faster now than for $\tau = 1$: $\psi(w)$ is a convex rather than a concave function of w. The shape of the total queue length changes as well. With recall, it is decreasing in wages.

I present one intermediate value for the termination probability: $\tau = 0.25$. This value of τ implies that firms with many applicants can contact on average 4 job candidates before their recruitment phase ends. The figure shows that in this case the equilibrium outcomes indeed lie between the two extreme cases. The wage density and the density of accepted wages are now (slightly) non-monotonic: decreasing in w first, but increasing again after reaching a minimum. Also the total queue length is non-monotonic, but this variable increases first and decreases only after that. The job offer probability and the firms' matching probability continue to be strictly increasing, but at a level that lies between the curves for $\tau = 0$ and $\tau = 1$.

In general, the equilibrium outcomes for $\tau = 0.25$ resemble the case without recall more than the full recall case. The reason for this is that even for $\tau = 0.25$, many firms are still constrained in the number of job offers that they can make. This is a direct result of the following lemma.

Lemma 4 *Consider a firm with total queue length λ that gets rejected by all applicants. The expected number of job offers that the firm will make as a fraction of the total queue length equals $\frac{1-e^{-\lambda\tau}}{\lambda\tau}$.*[9]

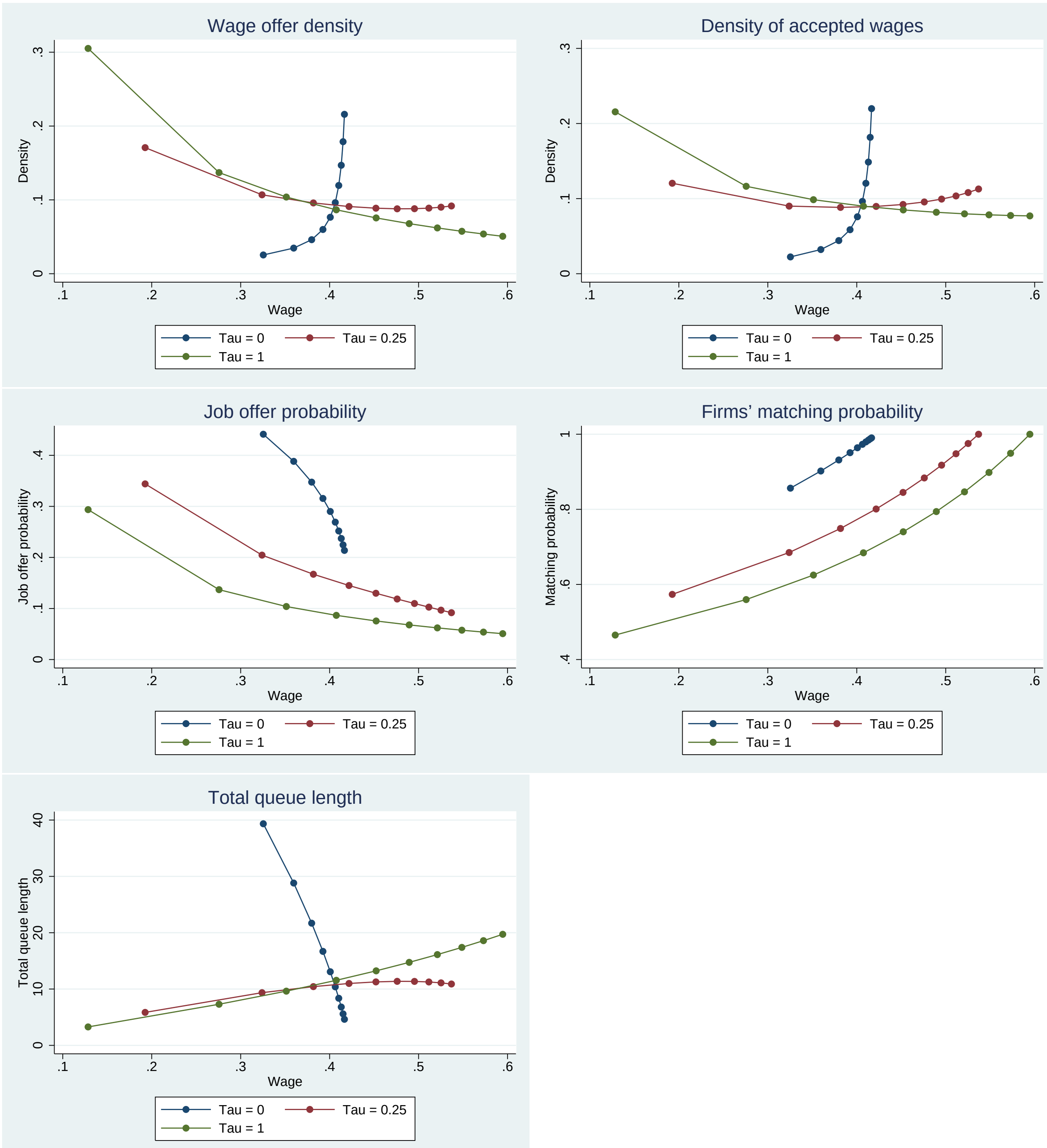
The lemma states that the expected number of applicants that firms can contact is a function of the product $\lambda\tau$ only. However, the total queue length λ at a specific firm is an endogenous variable and depends in equilibrium on τ. Figure 2 shows that for $\tau = 0.25$ many firms have a total queue length equal of approximately 11. Evaluating the expression in lemma 4 in $\tau = 0.25$ and $\lambda = 11$ shows that firms can contact on average only 34% of their applicants.

[to be continued]

5 Conclusion

This papers extends the existing literature on directed search with a model that allows firms to contact a new applicant if the first job offer gets rejected. I allow for an arbitrary number of job offers, which implies that the model contains several existing models as special cases. I show that the degree of recall has distinct effects on the equilibrium outcomes. The next step will be an structural estimation of the model in order to obtain estimates for the number of applications that workers send and the number of job offers that firms make. These estimates can then be used to quantify the relative contribution of both sides of the market on the welfare loss compared to a Walrasian world.

[9]This lemma speaks about the potential number of job offers. In general the firm will stop contacting applicants once one accepts.



Equilibrium outcomes for $A = 10$, $\theta = 1$, and various values of τ. The case $\tau = 0$ corresponds to a situation with perfect recall, while $\tau = 1$ implies no recall. In a world with $\tau = 0.25$, firms can contact four candidates on average. Only a discrete number of wages is offered in equilibrium (characterized by dots). The lines connecting the dots serve illustrative purposes only.

Figure 2: Effect of recall on equilibrium outcomes

A Proofs

A.1 Proof of Lemma 1

Proof. Consider the job offer probability ψ first.[10] When the worker applies to the firm, she will face in general a number of competitors, i.e. other applicants at the same firm. Let m denote the number of competitors who are interested in the job in the sense that they will accept the job if offered to them. On the other hand, let n denote the number of uninterested competitors who will reject the job in favor of a better offer. As described above, the recruiting process can exist of multiple rounds if applicants reject the job offer. The process ends when the firm selects an interested applicant, or when the exogenous termination shock occurs.

For the worker it is not relevant in which round she gets the job offer. Getting a job offer in round i and getting one in round j are mutually exclusive events, since the firm will offer the job to each worker at most once. Hence, the probability that the worker gets a job offer equals the probability that she gets an offer in round i, summed over all possible i. In order for the worker to get a job offer in round i, several things must happen: (i) only uninterested workers must have been selected in round $1,\ldots,i-1$, (ii) the recruiting process must not have terminated up to round i and (iii) the worker gets selected in round i itself. The probability that this occurs therefore equals

$$(1-\tau)^{i-1}\frac{n}{m+n+1}\cdot\frac{n-1}{m+n}\cdot\ldots\cdot\frac{n+2-i}{m+n+3-i}\cdot\frac{1}{m+n+2-i}=(1-\tau)^{i-1}\frac{n!}{(n+1-i)!}\frac{(m+n+1-i)!}{(m+n+1)!},$$

as long as $i\leq n+1$ and zero otherwise. Hence, the total probability that the worker gets a job offer equals

$$\hat{\psi}(m,n|\tau)=\sum_{i=1}^{n+1}(1-\tau)^{i-1}\frac{n!}{(n+1-i)!}\frac{(m+n+1-i)!}{(m+n+1)!},$$

which goes to $\frac{1}{m+1}$ for $\tau\to 0$.

In reality, the worker faces a random number of competing applicants. Both m and n follow Poisson distributions with respective parameters μ and $\nu=\lambda-\mu$. Hence, the probability to compete with exactly m interested and n uninterested applicants equals

$$z(m,n|\mu,\nu)=e^{-\mu}\frac{\mu^m}{m!}e^{-\nu}\frac{\nu^n}{n!}.$$

The ex ante probability to get a job offer therefore equals

$$\psi=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\hat{\psi}(n,m|\tau)\,z(m,n|\mu,\lambda-\mu).$$

[10] In this proof, I omit the dependence of the job offer probability, the matching probability and the queue lengths on the wage in order to keep notation simple.

Substituting the expression for $\hat{\psi}(m,n|\tau)$ and changing the order of summation yields

$$\psi = \sum_{i=1}^{\infty} (1-\tau)^{i-1} \sum_{m=0}^{\infty} \sum_{n=i-1}^{\infty} \frac{n!}{(n+1-i)!} \frac{(m+n+1-i)!}{(m+n+1)!} z(m,n|\mu,\lambda-\mu).$$

The last part of the right hand side can be rewritten as follows

$$\sum_{m=0}^{\infty} \sum_{n=i-1}^{\infty} \frac{n!}{(n+1-i)!} \frac{(m+n+1-i)!}{(m+n+1)!} z(m,n|\mu,\lambda-\mu) = \frac{(\lambda-\mu)^{i-1}}{\lambda^i} \left(1 - e^{-\lambda} \sum_{j=0}^{i-1} \frac{\lambda^j}{j!}\right).$$

Substituting this into the expression for ψ and applying a change in the order of summation gives

$$\psi = \sum_{i=1}^{\infty} \frac{(1-\tau)^{i-1} (\lambda-\mu)^{i-1}}{\lambda^i} - e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \sum_{i=j+1}^{\infty} \frac{(1-\tau)^{i-1} (\lambda-\mu)^{i-1}}{\lambda^i}.$$

Since $\sum_{i=j+1}^{\infty} \frac{(1-\tau)^{i-1}(\lambda-\mu)^{i-1}}{\lambda^i} = \frac{1}{\kappa} \left(\frac{(\lambda-\mu)(1-\tau)}{\lambda}\right)^j$ with $\kappa = \tau\lambda + (1-\tau)\mu$, this implies

$$\begin{aligned} \psi &= \frac{1}{\kappa} \left(1 - e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda-\mu)^j (1-\tau)^j}{j!}\right) \\ &= \frac{1}{\kappa} \left(1 - e^{-\kappa}\right). \end{aligned}$$

An expression for the firm's matching probability η can be derived in a similar way. A firm with m interested and n uninterested applicants, hires in round i with probability

$$(1-\tau)^{i-1} \frac{n}{m+n} \cdot \frac{n-1}{m+n-1} \cdot \ldots \cdot \frac{n+2-i}{m+n+2-i} \cdot \frac{m}{m+n+1-i} = (1-\tau)^{i-1} \frac{m \cdot n!}{(n+1-i)!} \frac{(m+n-i)!}{(m+n)!},$$

as long as $i \leq n+1$ and zero otherwise. Hence, given m and n the total hiring probability equals

$$\hat{\eta}(m,n|\tau) = \sum_{i=1}^{n+1} (1-\tau)^{i-1} \frac{m \cdot n!}{(n+1-i)!} \frac{(m+n-i)!}{(m+n)!},$$

which goes to 1 for $\tau \to 0$.

Both m and n follow Poisson distributions with respective parameters μ and $\nu = \lambda - \mu$. Hence, the ex ante matching probability is equal to

$$\begin{aligned} \eta &= \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} \hat{\eta}(n,m|\tau) z(m,n|\mu,\lambda-\mu) \\ &= \sum_{i=1}^{\infty} (1-\tau)^{i-1} \sum_{m=1}^{\infty} \sum_{n=i-1}^{\infty} \frac{m \cdot n!}{(n+1-i)!} \frac{(m+n-i)!}{(m+n)!} z(m,n|\mu,\lambda-\mu). \end{aligned}$$

The last part of the right hand side can be rewritten as follows

$$\sum_{m=1}^{\infty}\sum_{n=i-1}^{\infty}\frac{m\cdot n!}{(n+1-i)!}\frac{(m+n-i)!}{(m+n)!}z(m,n|\mu,\lambda-\mu)=\frac{\mu(\lambda-\mu)^{i-1}}{\lambda^{i}}\left(1-e^{-\lambda}\sum_{j=0}^{i-1}\frac{\lambda^{j}}{j!}\right),$$

which implies

$$\begin{aligned}\eta &= \mu\psi \\ &= \frac{\mu}{\kappa}\left(1-e^{-\kappa}\right).\end{aligned}$$

■

A.2 Proof of Lemma 2

Proof. See Galenianos and Kircher (2008) and Kircher (2007) [to be extended]. ■

A.3 Proof of Lemma 3

Proof. See Galenianos and Kircher (2008) and Kircher (2007) [to be extended]. ■

A.4 Proof of Proposition 1

Proof. The proof is a straightforward extension of Galenianos and Kircher (2008) and Kircher (2007) [to be extended]. ■

A.5 Proof of Lemma 4

Proof. If a firm gets rejected by all applicants, the number of offers that it can make solely depends on the number of applications and the moment at which the termination shock occurs. Suppose a firm has i applicants and can recruit for j periods. Then, the firm can make at most $\min\{i,j\}$ job offers. The probability that the firm has i applicants is given by $e^{-\lambda}\frac{\lambda^{i}}{i!}$, where as the probability that the termination shock occurs after period j is equals $(1-\tau)^{j-1}\tau$. Hence, the expected potential number of offers equals

$$\begin{aligned}
\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\min\{i,j\}\,e^{-\lambda}\frac{\lambda^i}{j!}(1-\tau)^{j-1}\tau &= \sum_{i=0}^{\infty}e^{-\lambda}\frac{\lambda^i}{i!}\sum_{j=0}^{i-1}j(1-\tau)^{j-1}\tau+\sum_{i=0}^{\infty}ie^{-\lambda}\frac{\lambda^i}{i!}\sum_{j=i}^{\infty}(1-\tau)^{j-1}\tau \\
&= \sum_{i=0}^{\infty}e^{-\lambda}\frac{\lambda^i}{i!}\frac{1-(1-\tau)^{i-1}(1-\tau+i\tau)}{\tau}+\sum_{i=0}^{\infty}ie^{-\lambda}\frac{\lambda^i}{i!}(1-\tau)^{i-1} \\
&= \frac{1-e^{-\lambda\tau}-\lambda\tau e^{-\lambda\tau}}{\tau}+e^{-\lambda\tau}\lambda \\
&= \frac{1-e^{-\lambda\tau}}{\tau}.
\end{aligned}$$

Dividing by λ gives the desired result. ■

References

Acemoglu, D. and Shimer, R. (1999a). Efficient unemployment insurance. *Journal of Political Economy*, 107:893–928.

Acemoglu, D. and Shimer, R. (1999b). Holdups and efficiency with search frictions. *International Economic Review*, 40:827–851.

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

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51:955–969.

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

Butters, G. (1977). Equilibrium distributions of sales and advertising prices. *Review of Economic Studies*, 44(3):465–491.

Gale, D. and Shapley, L. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15.

Galenianos, M. and Kircher, P. (2005). Directed search with multiple job applications. Technical report, PIER working paper.

Galenianos, M. and Kircher, P. (2008). Directed search with multiple job applications. *Journal of Economic Theory*, forthcoming.

Gautier, P. A. and Moraga-González, J. L. (2005). Strategic wage setting and coordination frictions with multiple job applications. mimeo.

Gautier, P. A., Moraga-González, J. L., and Wolthoff, R. P. (2007). Structural estimation of search intensity: Do non-employed workers search enough? mimeo.

Gautier, P. A. and Teulings, C. N. (2006). How large are search frictions? *Journal of the European Economic Association*, 4(6):1193–1225.

Hitsch, G. J., Hortaçsu, A., and Ariely, D. (2008). Matching and sorting in online dating. mimeo.

Julien, B., Kennes, J., and King, I. (2000). Bidding for labor. *Review of Economic Dynamics*, 3:619–649.

Julien, B., Kennes, J., and King, I. (2006). Residual wage disparity and coordination unemployment. *International Economic Review*, 47(3):961–989.

Kaas, L. (2007). Variable search intensity in an economy with coordination unemployment. mimeo.

Karlin, S. (1962). *Studies in Applied Probability and Management Science*, chapter Stochastic Models and Optimal Policy for Selling an Asset, pages 148–158. Stanford University Press, Stanford.

Kircher, P. (2007). Efficiency of simultaneous search. mimeo.

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

Roth, A. E. (2008). Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36:537–569.

Sorensen, M. (2007). How smart is smart money? an empirical two-sided matching model of venture capital. *Journal of Finance*, 62:2725–2762.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Stigler, G. J. (1962). Information in the labor market. *Journal of Political Economy*, 70(5):94–105.