Federal Reserve Bank of Dallas

An Information-Based Theory of Monopsony Power

Anton Cheremukhin and Paulina Restrepo-Echavarria

Working Paper 2518 **May 2025**

Research Department

https://doi.org/10.24149/wp2518

An Information-Based Theory of Monopsony Power*

Anton Cheremukhin† and Paulina Restrepo-Echavarria‡

April 25, 2025

Abstract

We develop a tractable model of monopsony power based on information frictions in job search. Workers and firms choose probabilistic search strategies, with information costs limiting how precisely they can target matches. Firms post wages strategically, anticipating application behavior and exploiting a first-mover advantage. The model nests both directed and random search as limiting cases and yields a closed-form wage equation that shows the effects on wage-setting power of search frictions, labor market tightness, and sorting. Wage markdowns in equilibrium arise not only from limited labor supply elasticity but also from sorting patterns and demand-side frictions. In highly assortative environments, the absence of wage competition allows firms to capture nearly the full surplus, even when labor supply is elastic. Numerical results replicate markdowns of 30-40% and suggest that constrained-efficient wages would be approximately 20% higher. Our framework unifies the analysis of monopsony, sorting, and wage posting, and provides a computationally efficient method for evaluating directed search equilibria.

JEL Codes: C78, J42, D83

Keywords: Monopsony power, Labor market sorting, Information frictions, Directed search

---

* Declarations of interest: none. We thank David Berger, Olivier Coibion, Kyle Herkenhoff, Benjamin Lester, Guillaume Rocheteau, Esteban Rossi-Hansberg, Aysegul Sahin, Kjetil Storesletten, Gianluca Violante, as well as participants of the seminar at the Dallas Fed, St. Louis Fed, and SED meetings in Barcelona for their insightful comments. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Reserve Banks of Dallas and St. Louis or the Federal Reserve System.

†Anton Cheremukhin, Federal Reserve Bank of Dallas, 2200 N. Pearl St., Dallas, TX 75201, chertosha@gmail.com.

‡Paulina Restrepo-Echavarria, Federal Reserve Bank of St. Louis, One Federal Reserve Bank Plaza, Broadway and Locust St., St. Louis, MO 63166, paulinares@me.com.

1 Introduction

A central insight of models of monopsony is that firms set wages below marginal product, either because workers face search frictions or because jobs are differentiated (Card, 2022). Yet canonical models typically treat these sources in isolation. In this paper, we develop an information-based theory of monopsony power that unifies the search and job differentiation perspectives in a single framework. Our model endogenizes both the targeting precision of job search and the elasticity of firm access to labor, generating new insights into the determinants of wage-setting power.

We build on Cheremukhin, Restrepo-Echavarria, and Tutino (2020), who introduced a model of simultaneous targeted search under information constraints, more applicable in the context of marriage markets. Here, we adapt and extend that framework to the labor market, introducing sequential wage posting and strategic firm behavior. Workers and firms choose probabilistic search strategies, deciding how precisely to target a partner type. These decisions are subject to endogenous information frictions, modeled as relative entropy (Kullback-Leibler divergence) between a uniform prior and the chosen strategy. In this rational inattention specification, frictions arise endogenously from agents' limited ability to identify their preferred match and provide a microfoundation for the multinomial logit (MNL) discrete choice structure frequently assumed in the literature on job differetiation.

Our framework nests both directed and random search as limiting cases. When information costs vanish, workers can fully target their desired firms types, replicating the classic directed search environment. When costs are infinite, search is random and matches occur with uniform probability. For intermediate levels of costs, agents choose imperfectly targeted strategies, and wages posted by firms reflect their expectations about the number and composition of applicants, as well as the costliness of screening. The model preserves the key features of directed search: sequential wage posting, strategic attraction of worker types, and matching via submarket-specific technologies, but introduces information frictions which span the continuum between directed and random search.

We derive a closed-form expression for the equilibrium wage that isolates the fundamental drivers of monopsony power. In our model, monopsony arises from five distinct sources: (i) worker-side search costs affect the precision with which wage compensation attracts workers and determine the elasticity of labor supply; (ii) firm-side screening costs shape firms' willingness to screen applicants and determine the elasticity of labor demand; (iii) labor market tightness amplifies both elasticities through numerical imbalances; (iv) equilibrium sorting, where positive assortative matching reduces wage competition across firms and increases firm surplus shares; and (v) the sequential nature of search, which gives firms a first-mover advantage by letting them strategically commit to wage menus. Contrary to canonical models, we show that the relationship between wage markdowns and the labor supply elasticity is not one-to-one: demand-side factors and sorting patterns significantly affect equilibrium wages.

We calibrate the model to match empirically plausible parameters and show that moderate information frictions, characterized by search cost parameters in the range 0.1-0.5, generate wage markdowns of 30-40%, consistent with recent empirical estimates. However, these outcomes are socially suboptimal. A constrained social planner, internalizing congestion externalities, would recommend a combination of reduced search effort and higher wages, yielding markdowns of only 10-15%. This implies that optimal policy could raise wages by approximately 20% without reducing the number of matches or welfare.

Our framework also provides new insights into the properties of directed search equilibria. In the zero-cost limit, multiple equilibria arise: positively assortative (PAM), negatively assortative (NAM), and mixed-strategy types. Only some of these equilibria implement the planner's solution; others generate inefficient allocations or extreme wage dispersion. We show how the strength of productive complementarities determines whether assortative or mixed matching is efficient and how assortativeness leads to excessive monopsony power. The literature often focuses on PAM equilibria, particularly when productive complementarities are strong, since they tend to maximize output. However, we show that assortative matching systematically increases monopsony power by suppressing wage competition and allowing firms to capture most of the surplus. In contrast, mixing equilibria generate more competitive wage-setting and a fairer surplus split but may be inefficient, even when they yield higher welfare than other equilibria. Our numerical analysis highlights how the strength of productive complementarities (i.e. the shape of the surplus) governs both the efficiency and distributional properties of equilibria, and shows that even the welfare-dominant directed search equilibria may be unstable, failing to survive under small information frictions.

Our method introduces a novel approach for characterizing directed search equilibria by convexifying the discrete strategic problem using an entropic regularization. This connects our paper to recent developments in convex approximations of combinatorial problems (e.g., Cuturi, 2013; Oberfield et al., 2024). We show that our information-based approach not only yields a continuous approximation to discrete directed search equilibria but also allows efficient computation and robust comparative statics, offering a diagnostic tool for identifying equilibrium multiplicity, inefficiency, and non-survivability under frictions.

Thus, our contributions are threefold. First, we propose a unified model of monopsony that spans directed and random search through endogenous information frictions. Second, we derive a closed-form wage equation identifying five distinct channels of monopsony power, including both supply- and demand-side frictions. Third, we use the model to clarify the efficiency and wage-determination properties of directed search equilibria, providing both theoretical results and numerical characterizations. These contributions shed light on the positive and normative implications of monopsony power in modern labor markets, and respond to recent calls for more structurally grounded models of monopsony that integrate strategic interactions, sorting, and information frictions (Card, 2022; Azar and Marinescu, 2024b). By endogenizing targeting behavior and wage setting, our framework offers a unified and tractable alternative to reduced-form elasticity-based approaches. In particular, it shows that wage-setting power arises not only from limited labor supply elasticity, but also from equilibrium sorting and market tightness, which together can generate large markdowns even when labor supply is highly elastic.

The paper proceeds as follows. Section 2 presents the model of sequential targeted search with endogenous information costs and characterizes the competitive equilibria and the planner's solution. Section 3 connects our framework to the search literature by exploring the properties of the model in the directed search and random search limits. Section 4 analyzes the sources and consequences of monopsony power, derives the wage markdown equation, and provides quantitative results for wages, markdowns, and efficiency. Section 5 concludes.

Related literature

This paper relates to several strands of the economic literature. First, it contributes to the directed search literature, which studies markets where firms post wages and workers choose where to apply (e.g., Moen, 1997; Burdett, Shi, and Wright, 2001; Shi, 2001). Standard directed search models assume perfect information and costless targeting, leading to submarket-based equilibria that are often efficient and feature assortative matching (see Eeckhout and Kircher, 2010; Guerrieri, Julien, Kircher, and Wright, 2021). Our framework generalizes this class of models by introducing information costs that determine how precisely workers and firms can target one another. This friction generates a continuum between directed and random search and allows us to recover standard directed search as a limiting case. A broader perspective on these modeling choices and their implications for sorting, matching technologies, and equilibrium efficiency is provided in Chade, Eeckhout, and Smith (2017).

A key finding of our paper is that directed search models generically admit multiple equilibria. While positive assortative matching is efficient when productive complementarities are strong, we show that it systematically increases firms' monopsony power by eliminating wage competition. In contrast, mixing equilibria feature fairer surplus splits, but may be socially suboptimal even when productive complementarities are weak. Our framework provides a unified tool for analyzing efficiency and surplus distribution across these equilibria, as well as their stability under frictions.

Second, we contribute to the monopsony literature. Recent empirical studies document substantial wage markdowns across a wide range of labor markets (e.g., Staiger, Spetz, and Phibbs, 2010; Azar, Marinescu, and Steinbaum, 2019; Lamadon, Mogstad, and Setzler, 2022; Dube, Jacobs, and Naidu, 2022), often attributing them to limited labor supply elasticity. Theoretical models typically attribute monopsony power to search frictions (e.g., Burdett and Mortensen, 1998), job differentiation (Manning, 2003), or oligopsony (Robinson, 1933; Boal and Ransom, 1997) and quantify the aggregate welfare losses due to labor market power as large (Berger, Herkenhoff, and Mongey, 2022). We provide a tractable framework that unifies these mechanisms and yields a closed-form wage equation decomposing the sources of monopsony power into five channels: information frictions on both sides, labor market tightness, equilibrium sorting, and wage-setting timing. Unlike models that focus solely on labor supply elasticities, we show that markdowns emerge endogenously from strategic interactions between firms and workers, and are shaped by the structure of sorting and information in equilibrium.

Several recent papers attempt to integrate key elements of monopsony: search frictions, heterogeneity, and concentration — into unified models (e.g., Lamadon, Mogstad, and Setzler, 2022; Berger, Herkenhoff, Kostol, and Mongey, 2023). These frameworks combine existing modeling blocks, such as random search with bargaining or discrete choice with rent-sharing, to estimate wage markdowns and labor market misallocation. However, they treat the relationship between sorting, elasticities, and wage setting as largely mechanical: labor supply elasticity governs wages, while sorting is determined separately by preferences or technology. In contrast, our model captures the full strategic interaction between search precision, wage posting, and equilibrium sorting. Firms anticipate how targeted or diffuse applications will be, and this shapes both wages and sorting patterns. As a result, monopsony power depends jointly on concentrations, labor elasticities, and sorting. In assortative environments with strong complementarities, wages collapse toward the outside option, and firms retain nearly the entire surplus, even when labor supply is highly elastic. Sorting becomes a central determinant of wage markdowns, not just match efficiency.

Several papers have attempted to nest directed and random search to generate intermediate degrees of randomness in matching (e.g., Menzio, 2007; Lester, 2011; Lentz, Maibom and Moen, 2022) and study the efficiency consequences (Rabinovich, Wolthoff, 2022). Unlike these models, which impose hybrid structures, our framework endogenizes the degree of search precision via information costs. This yields a continuous spectrum between uniform random matching and directed search, within a unified model. Relatedly, Wu (2024) studies a version of our targeted search model in a labor market setting, but does not characterize the structure of monopsony power or properties of directed search equilibria.

Methodologically, our approach connects to recent work that convexifies discrete assignment or search problems using regularization techniques. Our use of information costs mirrors entropic regularization in optimal transport (Cuturi, 2013) and convex approximations in spatial assignment models (Oberfield, Rossi-Hansberg, Sarte, and Trachter 2024). This allows us to efficiently compute equilibria in environments where standard directed search models exhibit indeterminacy or are difficult to solve.

Finally, we build on our earlier work in Cheremukhin, Restrepo-Echavarria, and Tutino (2020), which studied targeted search in a simultaneous-move setting with bargaining. The current paper shifts to sequential search with wage posting, focuses on the labor market rather than the marriage market, and introduces closed-form analytical characterizations of wages and sorting. In doing so, we reconcile the tractability of multinomial logit-type models with the equilibrium structure of directed search and monopsony.

2 Model

In this section, we present a model where firms are looking to fill a vacancy, and workers — who are either employed or unemployed — are looking to find a job. Each agent chooses a probabilistic search strategy that can be interpreted as a search intensity over types, where each element of the distribution reflects the likelihood of contacting a particular agent on the other side. A more targeted search, or a probability distribution that is more concentrated on a particular group of agents (or agent), is associated with a higher cost, as the agent needs to exert more effort to locate a particular potential match more accurately.

The economy contains a large, finite number of individual agents: workers whose types are indexed by $x \in \{1, ..., W\}$ and firms whose types are indexed by $y \in \{1, ..., F\}$. We denote by μ_x the number of workers of type x and by μ_y the number of firms of type y. We think of workers and firms characterized by a multidimensional set of attributes. Types x and y are unranked indices that aggregate all attributes.

A match between any worker of type x and any firm of type y generates a payoff (surplus) f_{xy}. We do not place any restrictions on the shape of the payoff function, and we normalize the outside option of both the worker and the firm to zero. We denote the payoff (wage) appropriated by the worker ω_{xy} and the payoff appropriated by the firm η_{xy} such that $\eta_{xy} = f_{xy} - \omega_{xy}$.

Agents form a match if they meet, and each agent (weakly) benefits from forming a match; i.e., each agent's payoff is non-negative. Since a negative payoff corresponds to absence of a match, we make the following assumption on the payoffs:

Assumption 1. The payoffs are non-negative:

$$f_{xy} \geq \omega_{xy} \geq 0.$$

When seeking to form a match, both workers and firms know the number of agents of each type and the characteristics of their preferred types on the other side of the market. They face a noisy search process where they are uncertain about how to locate their preferred match. In this environment, each agent's action is a probability distribution over agents on the other side of the market. Since the number of potential matches is finite, the strategy of each agent is a discrete probability distribution. Let $\bar{p}_x(y)$ be the probability that a worker of type x targets or sends an application to a firm of type y. Similarly, we denote by $\bar{q}_y(x)$ the probability that a firm of type y targets or considers the application of a worker of type x.

Reducing the noise to locate a potential match more accurately is costly: It involves a careful analysis of the profiles of potential matches, with considerable effort in sorting through the multifaceted attributes of each firm and candidate. When seeking to form a match, agents rationally weigh costs and benefits of targeting the type of characteristics that result in a suitable match. A worker rationally chooses their strategy $\bar{p}_x(y)$ by balancing the costs and benefits of targeting a given firm. A strategy $\bar{p}_x(y)$ that is more concentrated on a particular firm of type y affords them a higher probability to be matched with their preferred firm. However, it requires more effort to sort through profiles of all the firms in the market to locate their desired match and exclude the others. So locating a particular firm or worker more accurately requires exerting more search effort, and it is costlier.

We assume that agents enter the search process with a uniform prior of whom to target, $\tilde{p}_x(y)$ and $\tilde{q}_y(x)$. Choosing a more targeted strategy implies a larger distance between the chosen strategy and the uniform prior and is associated with a higher search effort. A natural way to introduce this feature into our model is the Kullback-Leibler divergence (relative entropy),[1] which provides a convenient way of quantifying the distance between any two distributions, including discrete distributions as in our

[1] In the model of information frictions used in the rational inattention literature, κ_x represents the relative entropy between a uniform prior and the posterior strategy. This definition is a special case of Shannon's channel capacity, where information structure is the only choice variable (See Thomas and Cover (1991), Chapter 2). See also Cheremukhin, Popova, and Tutino (2015) and Matejka and McKay (2015) for applications to stochastic discrete choice with information costs.

model. We assume that the search effort of worker i of type x is defined as follows:

$$\kappa_x = \sum_{y=1}^{F} \mu_y \bar{p}_x(y) \ln \frac{\bar{p}_x(y)}{\tilde{p}_x(y)}. \tag{2.1}$$

We assume that the search costs $c_x(\kappa_x)$ are a function of the search effort κ_x. Note that κ_x is increasing in the distance between a uniform distribution over firms and the chosen strategy, $\bar{p}_x(y)$. If an agent does not want to exert any search effort, she can choose a uniform distribution over types and meet firms randomly. As she chooses a more targeted strategy, the distance between the uniform distribution and her strategy $\bar{p}_x(y)$ grows, increasing search effort κ_x and the overall cost of search. By increasing the search effort, agents bring down uncertainty about locating a prospective match, which allows them to target their better matches more accurately.

Likewise, a firm's cost of search $c_y(\kappa_y)$ is a function of the search effort defined as:

$$\kappa_y = \sum_{x=1}^{F} \mu_x \bar{q}_y(x) \ln \frac{\bar{q}_y(x)}{\tilde{q}_y(x)}. \tag{2.2}$$

Furthermore, we assume the following:

Assumption 2. The search costs of agents $c_x(\kappa)$ and $c_y(\kappa)$ are strictly increasing, twice continuously differentiable and (weakly) convex functions of search effort.

As a special case, we consider a linear cost of search. Then, the total costs of search for a worker of type x are given by $c_x = \theta_x \kappa_x$ and for a firm of type y by $c_y = \theta_y \kappa_y$, where $\theta_x \geq 0$ and $\theta_y \geq 0$ are the marginal costs of search.

For convenience, we introduce a new notation for the strategies of workers and firms. We define the workers' and firms' search intensities as the ratios of their posterior and prior: $p_x(y) = \frac{\bar{p}_x(y)}{\tilde{p}_x(y)}$ and $q_y(x) = \frac{\bar{q}_y(x)}{\tilde{q}_y(x)}$, respectively.

The meeting rate depends on the strategies of each agent, $p_x(y)$ and $q_y(x)$, and a congestion function $\phi(p_x(y), q_y(x), \mu_x, \mu_y)$, which depends in some general way on the strategies of all other agents as well as the number of agents of each type. Given this, the total number of matches formed between workers of type x and firms of type y is given by

$$M_{x,y} = \mu_x \mu_y p_x(y) q_y(x) \phi(p_x(y), q_y(x), \mu_x, \mu_y).$$

Assumption 3. The congestion function is twice continuously differentiable in each p and q.

We introduce this congestion function following Shimer and Smith (2001) and Mortensen (1982), who assume a linear search technology. Note that if $\phi(...) = 1$, then a match takes place if and only if there is mutual coincidence of interests; i.e., both agents draw each other out of their respective distribution of interests. By introducing a congestion function we are allowing for matches to depend in some general way on both an agent's search intensity[2] for a specific agent (p and q) and on the number of agents taking part. We think of this assumption as representing the matching technology in a separate submarket for each combination of x and y.

Note that when setting up the congestion function we implicitly assume that there are no direct inter-type congestion externalities. However, our model still features strong indirect equilibrium interactions between the strategies of agents that work akin to inter-type congestion by attracting or deterring agents.

2.1 Sequential targeted search

To initiate the search and matching process, firms start by posting vacancies. Each posted vacancy includes a wage menu, and the firm commits to paying a type-dependent wage in the case of matching. After the vacancies are posted, and because workers cannot perfectly distinguish which firm is of which type despite learning the wage menus of each firm, they choose a distribution that determines the likelihood of contacting a particular firm and choose one firm from this distribution to send an application. Finally, once firms have received worker's applications, each firm chooses the worker to which it will extend a job offer from the set of workers that applied to that particular firm.

When workers decide where to send their applications, they take as given the posted wages of firms, such that the set of strategies of workers $p_x(y) \in S_x$ is given by:

$$S_x = \left\{ p_x(y) \in R_+^F : \sum_{y=1}^{F} \frac{\mu_y}{\delta_x} p_x(y) \leq 1 \right\},$$

[2]Note that here, search intensity refers to how concentrated the distribution of interests of an agent is. A higher search intensity results in assigning higher probability to one or several agents within an agent's distribution of interests.

where $p_x(y) = \frac{\bar{p}_x(y)}{\tilde{p}_x(y)}$, and $\tilde{p}_x(y) = 1/\sum_{y=1}^{F}\mu_y = 1/\delta_x$ is the worker's uniform prior over the whole set of firms $\left(\delta_x = \sum_{y=1}^{F}\mu_y\right)$.

The firms will strategically choose a wage menu $\omega_{x,y}$ and screening strategy $q_y(x)$. The other difference between the problem of workers and firms is that firms do not sort through all the workers that are looking for a job; they only sort through those that send an application to their firm, and when doing so, firms do not know the types of the workers that applied, but they know the length and expected composition of the queue. In expectation, the queue of firm y contains $\mu_x p_x(y)\,\delta_x/\mu_y$ workers of type x.

We define the set of strategies available for firms as:

$$S_y = \left\{ q_y(x), \omega_{xy} \in R_+^W : \sum_{x=1}^{W} a_{xy} q_y(x) \leq 1, \omega_{xy} \leq f_{xy} \right\}.$$

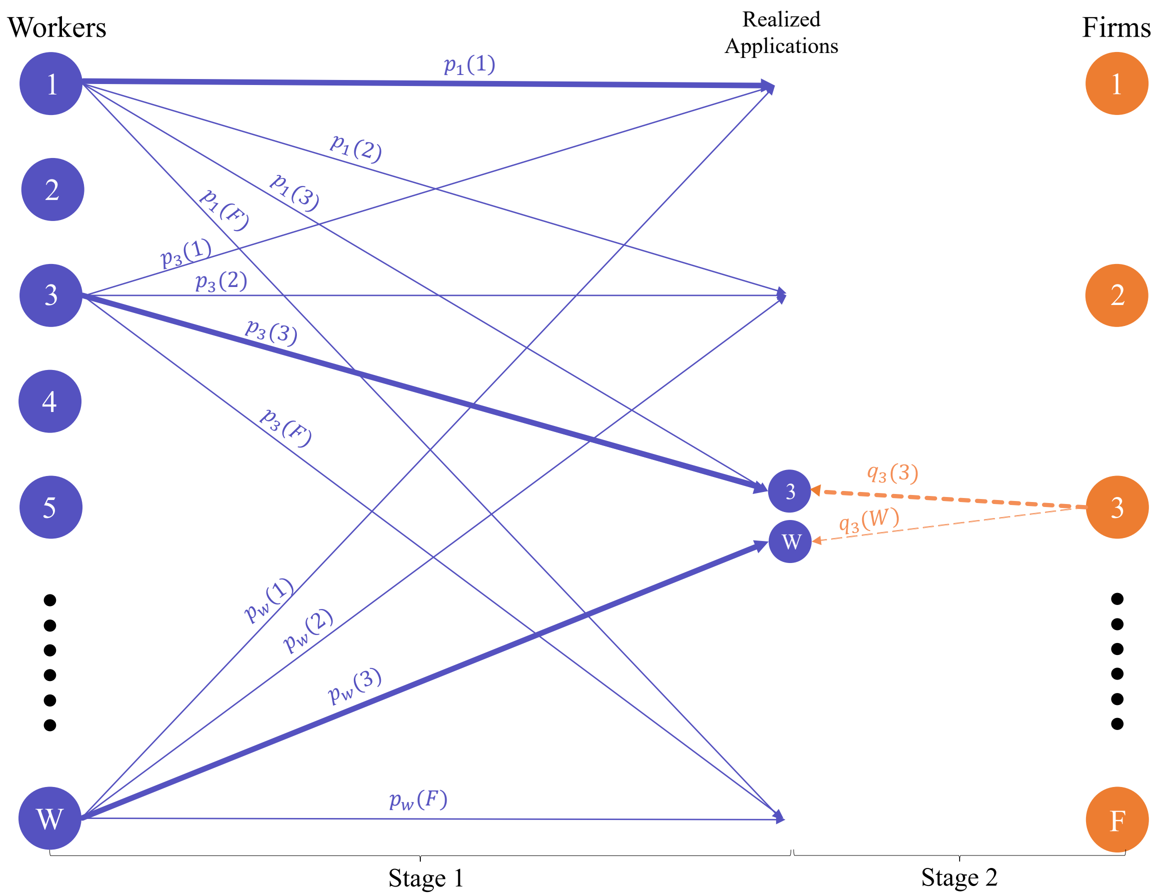
where $q_y(x) = \frac{\bar{p}_y(x)}{\tilde{p}_y(x)}$, and $\tilde{q}_y(x) = 1/\sum_{x=1}^{W}(\mu_x p_x(y)\,\delta_x/\mu_y)$ is the firm's uniform prior over their own queue. Here we define new variables for queue weights $a_{xy} = \frac{\mu_x p_x(y)}{\Sigma_{x=1}^{W}\mu_x p_x(y)}$, and queue length $\delta_y = \Sigma_{x=1}^{W}\mu_x p_x(y)$.

The set of actions $s \in S$ is given by the cartesian product of the sets of strategies of workers $s_x \in S_x$ and firms $s_y \in S_y$.

Figure 2.1 illustrates the interactions and search strategies of workers and firms. The solid arrows show the intensity $p_x(y)$ that a worker of type x assigns to targeting a firm of type y. Similarly, dashed arrows show the intensity $q_y(x)$ that a firm of type y assigns to targeting a worker of type x. Once these are selected, both workers and firms make one draw from their respective distributions to determine where to send an application and which applications to inspect (denoted by bold arrows).

Although applications and/or job offers are not lost in the mail, there is still a coordination problem: $\mu_x p_x(y)$ workers applied to type y firms, and firms sent $\mu_y q_y(x)\,\mu_x p_x(y)$ job offers, but they did not necessarily send all of those to different workers. Several firms might contact the same worker, and some workers may not get any offers. We assume that $\mu_x p_x \mu_y q_y \phi_{xy}$ matches are created, where the coordination problem between type x workers and type y firms is captured by the congestion

Figure 2.1: Strategies of Workers and Firms under Sequential Targeted Search



function/meeting technology $\phi_{xy}(p_x, q_y, \mu_x, \mu_y)$ described earlier.

Both firms and workers choose their optimal strategies, and if a firm and a worker match, the payoff f_{xy} is split between them according to the commitment whereby firms posted type-dependent wage menus in the first stage of the game.

The game is sequential as in Stackelberg in that when firms post wages and choose their search effort, they internalize the best response strategies of workers. Firms behave like leaders and workers behave like followers. However, consistent with the assumptions of the simultaneous model (see Cheremukhin, Restrepo-Echavarria, and Tutino (2020)), neither the workers nor the firms internalize the effects of their strategies on the congestion function and take matching rates in each submarket as given. This is because there are a large number of individuals of each type, so a change in an individual firm's or worker's strategy will not have a noticeable aggregate effect on the number of matches. This assumption of large number of identical agents of each type which all play identical strategies is reminiscent of "competitive" search.

Assumption 4. Agents take the meeting rates they face as given, disregarding the dependence of the congestion function on agents' own search intensities.

Definition. A *matching equilibrium* is a set of admissible strategies for workers $s_x \in S_x$, firms $s_y \in S_y$, and meeting rates, such that the strategies solve the problems for each individual firm and worker given the meeting rates, which are consistent with the strategies of the agents.

2.2 The problem of the worker

We start by describing the problem of the worker. Workers take as given $q_y(x)\phi_{xy}$—the probability of forming a match with type y firms. The worker receives a wage ω_{xy} in the case of matching and bears a linear cost of search $\theta_x \kappa_x(p_x(y))$. The goal of type x workers is to maximize surplus subject to a constraint on strategies (with renormalized Lagrange multiplier λ_x):

$$Y_x = \Sigma_{y=1}^{F} \mu_y q_y(x)\phi_{xy}\omega_{xy} p_x(y) - \theta_x \Sigma_{y=1}^{F} \frac{\mu_y}{\delta_x} p_x(y) \ln p_x(y) + \theta_x \lambda_x \left(1 - \Sigma_{y=1}^{F} \frac{\mu_y}{\delta_x} p_x(y)\right)$$

Since the objective function of workers is twice continuously differentiable and concave in their own strategies, first-order conditions are necessary and sufficient conditions for equilibrium. Using the necessary first-order condition,

$$\frac{\theta_x}{\delta_x}\left(\ln p_x\left(y\right)+\lambda_x+1\right)=q_y\left(x\right)\phi_{xy}f_{xy}, \tag{2.3}$$

we can derive a closed-form solution for the optimal strategy of workers:

$$p_x^*\left(y\right)=\frac{\exp\left(\frac{q_y(x)\phi_{xy}\omega_{xy}}{\theta_x/\delta_x}\right)}{\Sigma_{y'=1}^{F}\frac{\mu_{y'}}{\delta_x}\exp\left(\frac{q_{y'}(x)\phi_{xy'}\omega_{xy'}}{\theta_x/\delta_x}\right)}. \tag{2.4}$$

2.3 The problem of the firm

The goal of type y firms is to choose wages and search intensities over their queue of workers to maximize their expected match payoffs $f_{xy}-\omega_{xy}$, net of linear search costs $\theta_y\kappa_y\left(q_y\left(x\right)\right)$ and subject to a constraint on strategies (with renormalized Lagrange multiplier λ_y):

$$\begin{aligned}Y_y &= \Sigma_{x=1}^{W}\mu_xp_x\left(y\right)\phi_{xy}q_y\left(x\right)\left(f_{xy}-\omega_{xy}\right)-\theta_y\Sigma_{x=1}^{W}\frac{\mu_xp_x\left(y\right)}{\Sigma_{x=1}^{W}\mu_xp_x\left(y\right)}q_y\left(x\right)\ln q_y\left(x\right)\\ &\quad+\theta_y\lambda_y\left(1-\Sigma_{x=1}^{W}\frac{\mu_xp_x\left(y\right)}{\Sigma_{x=1}^{W}\mu_xp_x\left(y\right)}q_y\left(x\right)\right).\end{aligned}$$

The firm internalizes the best responses of the workers (Equation 2.4). To internalize the responses, we need to take derivatives of $p_x\left(y\right)$ with respect to the wage ω_{xy} set by the firm and with respect to the firm's search strategy $q_y\left(x\right)$. If we introduce new notation $z_{xy}=\frac{\phi_{xy}q_y(x)}{\theta_x/\delta_x}\left(1-\frac{\mu_y}{\delta_x}p_x\left(y\right)\right)$, then the partial derivatives of (2.4) are conveniently given by: $\frac{\partial p_x(y)}{\partial q_y(x)}\frac{q_y(x)}{p_x(y)}=\omega_{xy}z_{xy}$ and $\frac{\partial p_x(y)}{\partial\omega_{xy}}\frac{1}{p_x(y)}=z_{xy}$. In addition, note that the derivatives of queue weights $a_{xy}=\frac{\mu_xp_x(y)}{\Sigma_{x=1}^{W}\mu_xp_x(y)}$ can be computed as $\frac{\partial a_{xy}}{\partial X}=a_{xy}\left(1-a_{xy}\right)\frac{\partial p_x(y)}{\partial X}\frac{1}{p_x(y)}$.

The problem can be rewritten as:

$$Y_y=\Sigma_{x=1}^{W}\mu_xp_x\left(y\right)\phi_{xy}q_y\left(x\right)\left(f_{xy}-\omega_{xy}\right)-\theta_y\Sigma_{x=1}^{W}a_{xy}q_y\left(x\right)\left(\ln q_y\left(x\right)+\lambda_y\right)+\theta_y\lambda_y,$$

and we can write the first-order condition of the firm with respect to search intensities as follows:

$$\frac{\partial Y_y}{\partial q_y} = \mu_x p_x\left(y\right)\frac{\theta_y}{\delta_y}\left[\begin{array}{c}\frac{\phi_{xy}}{\theta_y/\delta_y}\left(f_{xy}-\omega_{xy}\right)\left(1+z_{xy}\omega_{xy}\right)-1\\ -\left(\ln q_y\left(x\right)+\lambda_y\right)\left(1+\left(1-a_{xy}\right)z_{xy}\omega_{xy}\right)\end{array}\right]=0. \tag{2.5}$$

Strategies of firms then satisfy:

$$\ln q_y\left(x\right)+\lambda_y = \left(\frac{\phi_{xy}}{\theta_y/\delta_y}\left(f_{xy}-\omega_{xy}\right)\left(1+z_{xy}\omega_{xy}\right)-1\right)/\left(1+\left(1-a_{xy}\right)z_{xy}\omega_{xy}\right).$$

Firms' strategies must therefore satisfy the following necessary condition for equilibrium:

$$q_y^*\left(x\right) = \frac{\exp\left(\frac{\frac{\phi_{xy}}{\theta_y/\delta_y}\left(f_{xy}-\omega_{xy}\right)\left(1+z_{xy}\omega_{xy}\right)-1}{1+\left(1-a_{xy}\right)z_{xy}\omega_{xy}}\right)}{\Sigma_{x'=1}^{W}a_{x'y}\exp\left(\frac{\frac{\phi_{x'y}}{\theta_y/\delta_y}\left(f_{x'y}-\omega_{x'y}\right)\left(1+z_{x'y}\omega_{x'y}\right)-1}{1+\left(1-a_{x'y}\right)z_{x'y}\omega_{x'y}}\right)}. \tag{2.6}$$

Firms also optimally choose wage menus in the first stage. We can write the first-order condition with respect to wages as follows:

$$\frac{\partial Y_x}{\partial \omega_{xy}} = \mu_x p_x\left(y\right)q_y\left(x\right)\frac{\theta_y}{\delta_y}\left[\begin{array}{c}\frac{\phi_{xy}}{\theta_y/\delta_y}\left(\left(f_{xy}-\omega_{xy}\right)z_{xy}-1\right)\\ -\left(\ln q_y\left(x\right)+\lambda_y\right)\left(1-a_{xy}\right)z_{xy}\end{array}\right]=0, \tag{2.7}$$

and the second-order derivatives as:

$$\frac{\partial^2 Y_x}{\partial q_{xy}^2} = -\frac{1}{q_y\left(x\right)}, \qquad \frac{\partial^2 Y_x}{\partial \omega_{xy}^2} = -\frac{\phi_{xy}}{\theta_y/\delta_y}z_{xy}.$$

Since the objective function of firms is twice continuously differentiable and strictly concave with respect to their own strategies, the first-order conditions are necessary and sufficient conditions for equilibrium. Furthermore, we can combine the two optimality conditions (2.5) and (2.7) to eliminate $q_y\left(x\right)$ and λ_y to obtain a simple expression for an interior solution $0 \leq \omega_{xy} \leq f_{xy}$ for the wage:

$$\omega_{xy}^* = \left[a_{xy}f_{xy}+\left(1-a_x\right)\frac{\theta_y/\delta_y}{\phi_{xy}}-\frac{1}{z_{xy}}\right]_0^{f_{xy}}. \tag{2.8}$$

Wages stay at the limits because beyond the limits there is no match and the decision-maker is strictly worse off (as reflected in the constraints on the strategy space). In this case we can also substitute the (interior) optimal wage to obtain optimal search intensities of firms:

$$q_y^*(x) = \frac{\exp\left(\frac{\phi_{xy}}{\theta_y/\delta_y} f_{xy}\right)}{\Sigma_{x'=1}^{W} a_{x'y} \exp\left(\frac{\phi_{x'y}}{\theta_y/\delta_y} f_{x'y}\right)}.$$

The properties of the equilibrium, fully characterized by necessary conditions (2.4), (2.6) and (2.8) critically depend on the assumptions regarding the congestion function, in other words, the matching technology.

The matching technology we introduce is a standard symmetric constant returns to scale matching technology that combines the number of participants in each submarket. The number of agents entering each submarket (x, y) are $c_{x,y} = \mu_x p_x(y) \frac{\mu_y}{\delta_x}$ and $d_{x,y} = \mu_y q_y(x) a_{x,y}$. We assume that the matching technology is described by a symmetric CES function $M(c,d) = \left(\frac{1}{2}c^{\frac{\sigma-1}{\sigma}} + \frac{1}{2}d^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$, with $\sigma > 0, \sigma \neq 1$, with special cases for Cobb-Douglas when $\sigma = 1$ and Leontieff when $\sigma = 0$. In this case, the congestion function is defined as $\phi_{x,y} = M(c_{x,y}, d_{x,y}) / \mu_x \mu_y p_x(y) q_y(x)$. This assumption for various parameter choices encompasses most of the interesting cases studied in the literature. It is also directly comparable to our simultaneous targeted search model as it gives the same first best allocation when search costs approach zero.

Proposition 1. *Under assumptions 1- 4, there exists $\underline{\theta}$ such that for high enough costs relative to the number of agents $\left(\frac{\theta_x}{\delta_x}, \frac{\theta_y}{\delta_y}\right) > \underline{\theta}$ a matching equilibrium exists and is unique.*

Proof. The equilibrium of the matching model can be interpreted as a pure-strategy Nash equilibrium of a strategic form game among first-stage decisions of firms. Since the strategy space is a simplex and, hence, a non-empty, convex, compact set, sufficient conditions for the existence of the equilibrium require us to check whether the payoff functions are super-modular on the whole strategy space as in Tarski (1955). Supermodularity can be proven by showing negativity of diagonal elements and non-negativity of the off-diagonal elements of the Hessian matrix.

Let $J_y = \begin{bmatrix} \frac{\partial Y_y}{\partial q_{yx}} & \frac{\partial Y_y}{\partial \omega_{xy}} \end{bmatrix}$ be the Jacobian matrix collecting the set of first-order conditions for all firms $y \in \{1, ..., M\}$, and let H be the corresponding Hessian matrix.

To derive the Hessian matrix, note that under A.1, strategies of each firm are non-cooperative, i.e., independent of the strategies of other types as well as the strategies of the other agents of their own type. Note also that we have assumed no direct inter-type congestion externalities. These assumptions produce a Hessian matrix with a block-diagonal structure, which greatly simplifies the analysis. The Hessian consists of 2x2 blocks along the diagonal of the form:

$$H_{xy} = \begin{bmatrix} \frac{\partial^2 Y_y}{\partial q_{yx} \partial q_{yx}} & \frac{\partial^2 Y_y}{\partial \omega_{xy} \partial q_{yx}} \\ \frac{\partial^2 Y_y}{\partial q_{yx} \partial \omega_{xy}} & \frac{\partial^2 Y_y}{\partial \omega_{xy} \partial \omega_{xy}} \end{bmatrix}.$$

All the remaining off-diagonal elements are zero. The derivatives of interest are quite cumbersome to compute. However, we can express the elements of the Hessian as follows (where F and G are some positive functions):

$$\frac{\partial^2 Y_y}{\partial q_{yx} \partial q_{yx}} = -\frac{1}{q_{xy}} + \frac{\delta_x \delta_y}{\theta_x \theta_y} F\left(f_{xy}, \omega_{xy}, q_{xy}, a_{xy}\right) \leq 0,$$

$$\frac{\partial^2 Y_y}{\partial q_{yx} \partial \omega_{xy}} = \frac{\delta_x \delta_y}{\theta_x \theta_y} G\left(f_{xy}, \omega_{xy}, q_{xy}, a_{xy}\right) \geq 0,$$

$$\frac{\partial^2 Y_y}{\partial \omega_{xy} \partial \omega_{xy}} = -\frac{\delta_x \delta_y}{\theta_x \theta_y} \phi_{xy} \phi_{xy} q_{yx} \leq 0.$$

From this structure, it is clear that if costs of search are large enough (separately or in combination) relative to the number of agents, then all of these inequalities hold, while if costs are very small (or number of agents large) the first inequality is violated. For uniqueness, we need diagonal dominance of the form:

$$\left| \frac{\partial^2 Y_y}{\partial \omega_{xy} \partial \omega_{xy}} \right| \left| \frac{\partial^2 Y_y}{\partial q_{yx} \partial q_{yx}} \right| > \left(\frac{\partial^2 Y_y}{\partial q_{yx} \partial \omega_{xy}} \right)^2.$$

If costs are large enough (or number of agents small enough), then the diagonal terms dominate the off-diagonal terms. On the contrary, when costs are small (or numbers of agents large), then diagonal dominance may well be violated. We observe important cases of multiplicity numerically and discuss these in Section 3. □

In Section 4, we find that the threshold $\underline{\theta}$ is quite low, implying that for most parameterizations relevant for practical analysis the equilibrium is unique. Multiplicity of equilibria for low search costs is an important and robust finding which relates to the directed search literature, as we discuss in Section 3. The computation of all the equilibria of the model is possible using the necessary first-order conditions of the model for an exhaustively wide range of costs.

2.4 Social planner's solution

We solve the social planner's problem for the sequential model assuming an utilitarian welfare function. Consequently, the wage decision disappears from the social planner's problem altogether. We can write social welfare as the sum of objective functions of all the agents in the model, as the planner takes into account all the same benefits and costs of the matching process as the agents, subject to the same constraints on search intensities as individual agents. The social welfare function is then:

$$\Omega = \Sigma_{x=1}^{W}\mu_x Y_x + \Sigma_{y=1}^{F}\mu_y Y_y = \Sigma_{x=1}^{W}\mu_x\theta_x\lambda_x + \Sigma_{y=1}^{F}\mu_y\theta_y\lambda_y$$
$$+\Sigma_{x=1}^{W}\Sigma_{y=1}^{F}\mu_x\mu_y p_x(y)\left(q_y(x)\,\phi_{xy}f_{xy} - \tfrac{\theta_x}{\delta_x}\left(\ln p_x(y)+\lambda_x\right) - \tfrac{\theta_y}{\delta_y}q_y(x)\left(\ln q_y(x)+\lambda_y\right)\right).$$

The wages cancel out from the problem, and hence the planner's solution only describes allocations of search effort, but does not place restrictions on wage determination. The first-order conditions for the planner's problem can be written as follows:

$$\frac{\partial\Omega}{\partial p_x(y)} = \mu_x\mu_y\begin{pmatrix} q_y(x)\,f_{xy}\phi_{xy}\left(1+\varepsilon_{\phi,p}\right) - \frac{\theta_x}{\delta_x}\left(\ln p_x(y)+\lambda_x+1\right) \\ -\frac{\theta_y}{\delta_y}\left(1-a_{xy}\right)q_y(x)\left(\ln q_y(x)+\lambda_y\right)\end{pmatrix} = 0, \qquad (2.9)$$

$$\frac{\partial\Omega}{\partial q_y(x)} = \mu_x\mu_y p_x(y)\left(f_{xy}\phi_{xy}\left(1+\varepsilon_{\phi,q}\right) - \frac{\theta_y}{\delta_y} - \frac{\theta_y}{\delta_y}\left(\ln q_y(x)+\lambda_y\right)\right) = 0, \qquad (2.10)$$

where we denote $\varepsilon_{\phi,q} = \frac{\partial\phi_{xy}}{\partial q_y(x)}\frac{q_y(x)}{\phi_{xy}}$ and $\varepsilon_{\phi,p} = \frac{\partial\phi_{xy}}{\partial p_x(y)}\frac{p_x(y)}{\phi_{xy}}$. We can deduce from equations (2.9) and (2.10) that the search intensities prescribed by the planner satisfy:

$$\frac{\theta_y}{\delta_y}\left(\ln q_y(x)+\lambda_y\right) = \phi_{xy}f_{xy}\left(1+\varepsilon_{\phi,q}\right) - \frac{\theta_y}{\delta_y},$$

$$\frac{\theta_x}{\delta_x}\left(\ln p_x\left(y\right)+\lambda_x+1\right)=q_y\left(x\right)\left(f_{xy}\phi_{xy}\left[\left(1+\varepsilon_{\phi,p}\right)-\left(1-a_{xy}\right)\left(1+\varepsilon_{\phi,q}\right)\right]+\left(1-a_{xy}\right)\frac{\theta_y}{\delta_y}\right).$$

From our definition of the function ϕ_{xy} and its relationship to the matching function one can show that the elasticities satisfy: $1+\varepsilon_{\phi,q}=\varepsilon_{m,d}$ and $1+\varepsilon_{\phi,p}=\varepsilon_{m,c}+\left(1-a_{xy}\right)\varepsilon_{m,d}$. We can compare planner's optimality conditions with those of the competitive equilibrium described in (2.3) and (2.5), which we repeat here:

$$\frac{\theta_y}{\delta_y}\left(\ln q_y\left(x\right)+\lambda_y\right)=\left(\phi_{xy}\left(f_{xy}-\omega_{xy}\right)\left(1+z_{xy}\omega_{xy}\right)-\frac{\theta_y}{\delta_y}\right)/\left(1+\left(1-a_{xy}\right)z_{xy}\omega_{xy}\right),$$

$$\frac{\theta_x}{\delta_x}\left(\ln p_x\left(y\right)+\lambda_x+1\right)=q_y\left(x\right)\phi_{xy}\omega_{xy}.$$

We can derive expressions for wages in competitive equilibrium which would give efficient allocations by finding wages which would make the conditions for the workers equivalent, and wages which would make the conditions for the firms equivalent. To implement the strategies proposed by the social planner, workers should be promised a wage:

$$\omega_{xy}^{PO,W}=f_{xy}\varepsilon_{m,c}+\left(1-a_{xy}\right)\frac{\theta_x/\delta_x}{\phi_{xy}}.$$

Under the symmetric Cobb-Douglas calibration of the congestion function, $\varepsilon_{m,c}=\varepsilon_{m,d}=\frac{1}{2}$. The planner promises the worker half the surplus plus a positive term which vanishes as workers' search costs approach 0. In the limit, workers should receive exactly half the surplus. Comparing with the wage prevailing in competitive equilibrium given by (2.8), we observe that workers are promised a fraction a_{xy} of the surplus instead of half, and that firms charge an additional monopsony discount $1/z_{xy}$ reflecting their first mover advantage. Comparing the conditions for the firms, to implement the socially optimal strategies, firms should be promised a wage that satisfies:

$$\left(\phi_{xy}\left(f_{xy}-\omega_{xy}\right)\left(1+z_{xy}\omega_{xy}\right)-\frac{\theta_y}{\delta_y}\right)=\left(\phi_{xy}f_{xy}\varepsilon_{m,d}-\frac{\theta_y}{\delta_y}\right)\left(1+\left(1-a_{xy}\right)z_{xy}\omega_{xy}\right),$$

which boils down to a quadratic equation with respect to wages with one positive solution

$$\omega_{xy}^{PO,F} = \frac{A}{2} + \sqrt{\frac{A^2}{4} + \frac{1}{z_{xy}} f_{xy} \left(1 - \varepsilon_{m,d}\right)},$$

where we denote $A = \left(1 - \left(1 - a_{xy}\right) \varepsilon_{m,d}\right) f_{xy} + \left(1 - a_{xy}\right) \frac{\theta_y/\delta_y}{\phi_{xy}} - \frac{1}{z_{xy}}$. The wage prescribed to the firm by the planner and the equilibrium wage coincide when $\varepsilon_{m,d}$ equals 1. Otherwise, the wage prescribed by the planner is higher. In the symmetric case, one can show that as long as $a_{xy} f_{xy} > \frac{1}{z_{xy}}$ and $\frac{1}{2} f_{xy} > \left(1 - a_{xy}\right) \frac{\theta_x/\delta_x}{\phi_{xy}}$, competitive wages are lower than optimal wages: $\omega_{xy}^{CE} < \omega_{xy}^{PO,W} < \omega_{xy}^{PO,F}$.

For most parameters of interest, for low values of search costs, the firms should give away noticeably more than half of the surplus, leaving less than half for themselves, while workers should be getting exactly half. This demonstrates the fact that in the presence of strong negative congestion externalities, workers and firms jointly oversupply search effort in equilibrium. For the most part, workers need to be incentivized and firms need to be dis-incentivized from putting inefficiently high search effort by the planner promising lower payoffs in the case of matching. Implementation of this solution looks very much like a tax scheme that benefits workers and hurts firms yet obtains a better matching outcome at a lower search cost and generates on net extra revenue for society.

3 Connections to the search literature

This section examines how our model connects to and extends the search literature. We take a deliberately staged approach: we first study the model in the zero-cost limit and the infinite-cost limit, which correspond to classical directed and random search environments, respectively. We then explore how equilibrium outcomes evolve in the interior, where information frictions are small but non-negligible, and demonstrate how sorting and wage compression interact to influence monopsony power. This section highlights how canonical results from the literature emerge as special cases of our model, and the next section explores how our framework interpolates between them to uncover new results.

We begin by connecting our model to the directed search literature. In Appendix A, we show the full derivation of competitive equilibria and the planner's solution for a stylized 2-worker 2-firm heterogeneous agents directed search model. This model replicates as special cases under specific parameteric assumptions many standard directed search models, such as e.g. Burdett, Shi and Wright (2001), Moen (1997), Shi (2001), Eeckhout and Kircher (2010), and many others described in the survey by Guerrieri et al (2021). We demonstrate that the commonly held perception — that directed search models naturally lead to a unique assortative equilibrium with an efficient surplus division which decentralizes the socially efficient allocation — is only partially correct. As we shall see, there is truth to each of these perceptions separately, but they tend not to hold all at the same time.

In the limit as search costs approach zero, our sequential wage-posting targeted search model replicates all of the core features and agents' problems of directed search: strategic wage posting by firms, optimal worker application strategies, optimal firm screening strategies, and equilibrium assignment through submarkets. As comparison with the model in the Appendix reveals, the only additional assumption of our targeted search model is the positive costs of search linked to the precision of mixed strategies. Clearly, when search costs approach infinity, all the strategies will be optimally uniformly flat, delivering uniform random matching in equilibrium. Thus, for intermediate values of costs, our model fills in the middle ground between directed search (as costs approach zero) and random search with wage posting (as costs approach infinity).

To evaluate the full equilibrium structure of the model, we calibrate it in a two-worker, two-firm setting with a unit mass of workers and firms of each type, $\mu_x = \mu_y = 1$. We assume a Cobb-Douglas matching function $M(c, d) = \sqrt{c \cdot d}$, which satisfies constant returns and allows for analytical tractability. We consider four canonical surplus matrices:

1. Homogeneous: $f_{xy} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$, representing identical workers and identical firms.
2. One-sided heterogeneity: $f_{xy} = \left[\begin{smallmatrix} 2 & 1 \\ 2 & 1 \end{smallmatrix}\right]$, where firms differ, but workers are identical.
3. Horizontal surplus: $f_{xy} = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right]$, with strong productive complementarity.
4. Vertical surplus: $f_{xy} = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 0.4 \end{smallmatrix}\right]$, where one worker is better than the other, and one firm is better than the other.

For each of the four specifications, we assume symmetric costs of search $\theta_x = \theta_y = \theta$ and vary the cost parameter θ in a wide range from 0.0001 to 100. For each value of the cost parameter, we compute all the competitive equilibria and the socially optimal allocation using the necessary and sufficient conditions derived in the previous section. We compute each solution by making an initial guess for the strategies of the workers and firms, computing the equilibrium posted wage, and then checking if the optimality conditions for the remaining strategies are satisfied. We vary the vector of strategies until we find a fixed point.

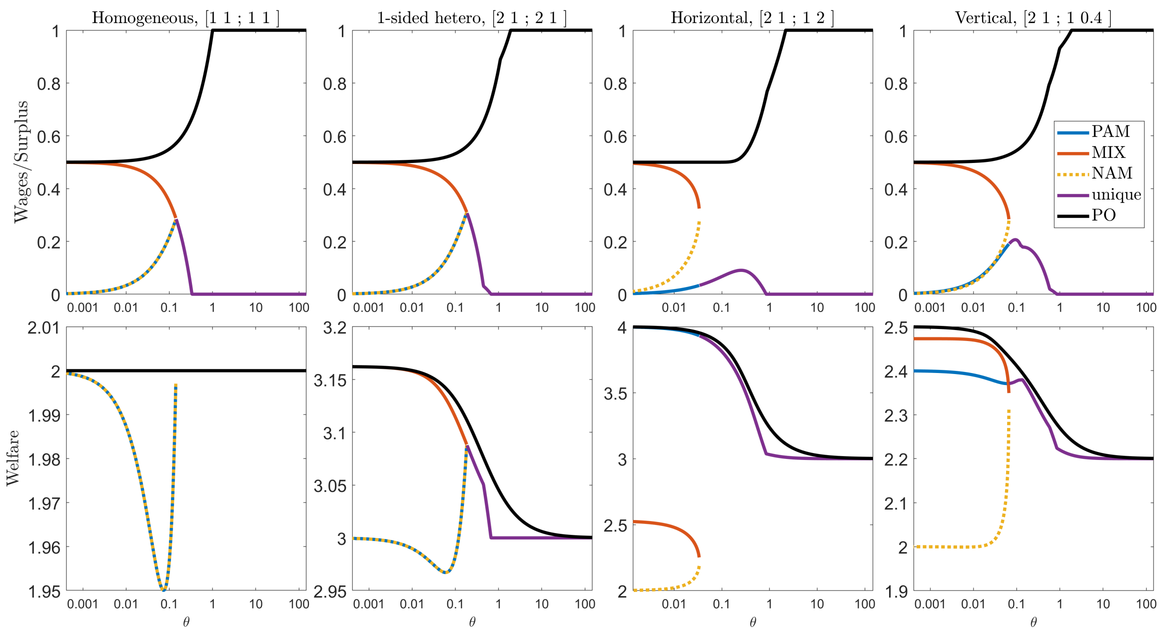
Each column of Figure 3.1 shows how the average posted wages as a fraction of the surplus and total welfare vary with search costs for various equilibria, and for the social planner's solution, for each of the four surplus specifications. The first specification replicates the findings of Burdett, Shi and Wright (2001).[3] There are two pure-strategy equilibria supported by wages arbitrarily close to zero, and one mixed-strategy equilibrium which splits the surplus equally. This case also illustrates how splitting identical types in the textbook homogeneous-agent directed search model produces multiplicity of equilibria.

The second specification replicates the findings in Moen (1997). The mixing equilibrium in this case reproduces the near equalization of expected values $q_y(x)\,\phi_{xy} \cdot \omega_{xy}$ (queue length times the wage) for each worker across locations, a feature commonly used in the directed/competitive search literature and in the subsequent block-recursive specifications (see Menzio and Shi, 2010). In this case, the mixing equilibrium also implements the socially optimal allocation with wages that fairly split the surplus - the celebrated properties of directed search that have come to be treated as automatic. The key assumption that delivers these results is the ex ante homogeneity of workers: small deviations towards two-sided heterogeneity can lead to very different predictions, as we see in the remaining two specifications. Compared with Moen's (1997) original model, the simple relabeling of homogeneous workers from one type into two identical types leads to multiplicity of equilibria.

In contrast, the horizontal and vertical cases produce much richer dynamics. In the horizontal case, the PAM equilibrium emerges as the best equilibrium in the zero-cost limit. This equilibrium maximizes welfare but results in extreme monopsony power: workers apply only to one firm type, and firms face no wage competition. Wages collapse to the outside option, and firms extract the full surplus. In the vertical surplus case, the mixed equilibrium gives the highest welfare, with posted wages delivering an efficient split of the surplus, yet it yields a socially suboptimal allocation even in the zero-cost limit. These results are consistent with the predictions of the directed search model

[3] The only difference from Burdett et al (2001) is that their model assumes increasing returns to scale ($M(c, d) = c \cdot d$), which lead to social inefficiency of all the equilibria.

Figure 3.1: Model comparison



we describe in the Appendix. One new insight from our targeted search framework is that one of the mixing equilibria of the directed search framework which implements the socially optimal allocation does not survive for (arbitrarily small) positive costs of search.

The three outcomes of the heterogeneous-agents cases are consistent with the derivation of Eeckhout and Kircher (2010) whereby prevalence of sorting depends on the interplay of production complementarity and matching complementarity. They predict that for positive assortative matching to prevail, the strength of supermodularity of production needs to exceed the strength of matching complementarity. In the cases we consider, the strength of matching complementarity equals 1. The production complementarity index $\frac{f_{xy}f}{f_x f_y}$ equals 1 for the two homogeneous cases, is between 0 and 1 for the vertical surplus case, and exceeds 1 for the horizontal surplus case. Indeed, we find that when production supermodularity dominates, PAM is efficient, otherwise mixing equilibria are best, but not necessarily efficient.

Similarly to Burdett et al (2001), in all cases, for low values of costs there are exactly three equilibria: two assortative equilibria (PAM and NAM) and one mixing equilibrium. While we get three equilibria for low search costs, as predicted by our propo-

Figure 3.2: Equilibrium strategies and wages for horizontal case

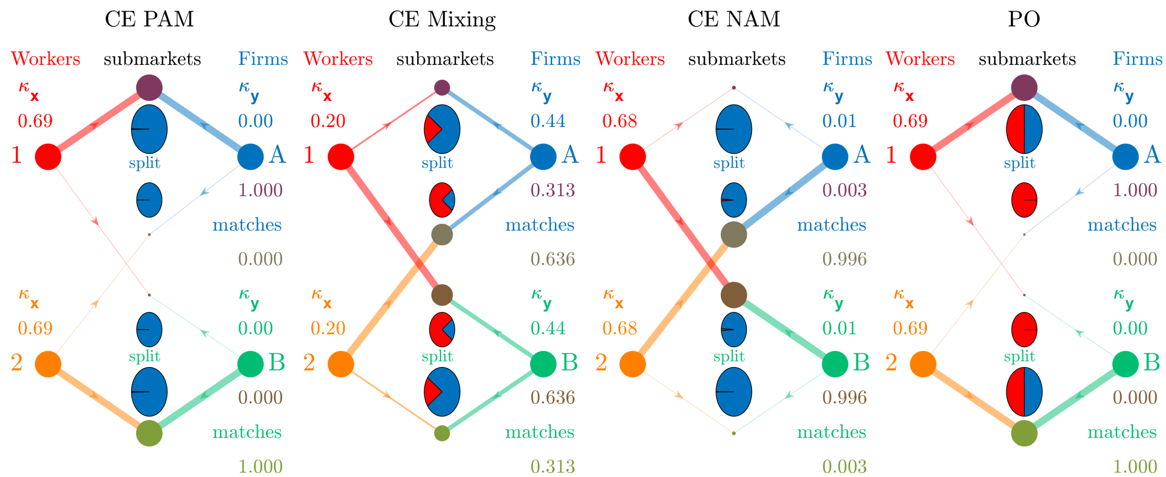
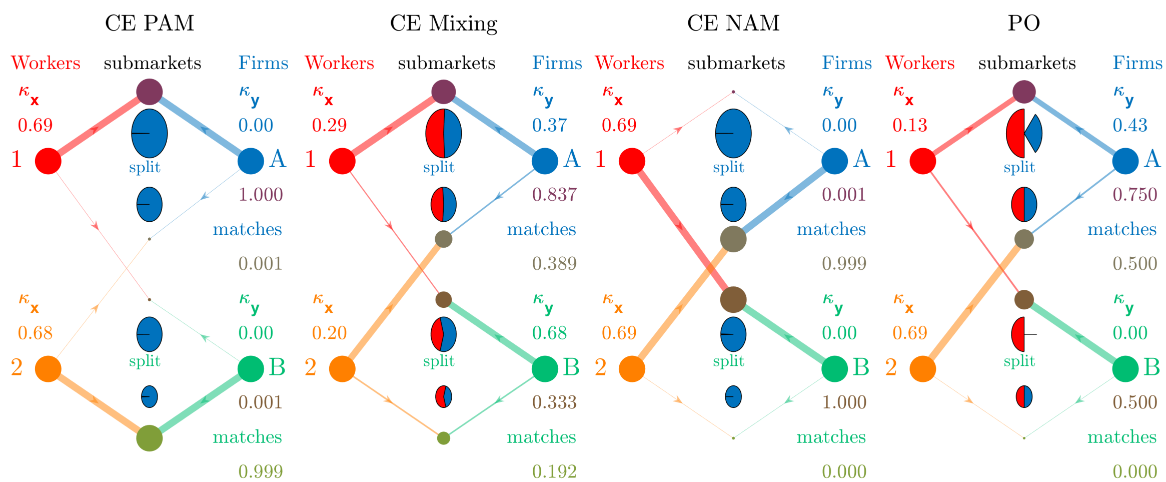


Figure 3.3: Equilibrium strategies and wages for vertical case



sition in section 2, for costs above a certain threshold a unique equilibrium prevails. The unique equilibrium often has the structure of the best of the multiple equilibria, specifically when the PAM equilibrium is best, as well as for one-sided heterogeneity, but not necessarily when the mixing equilibrium is best.

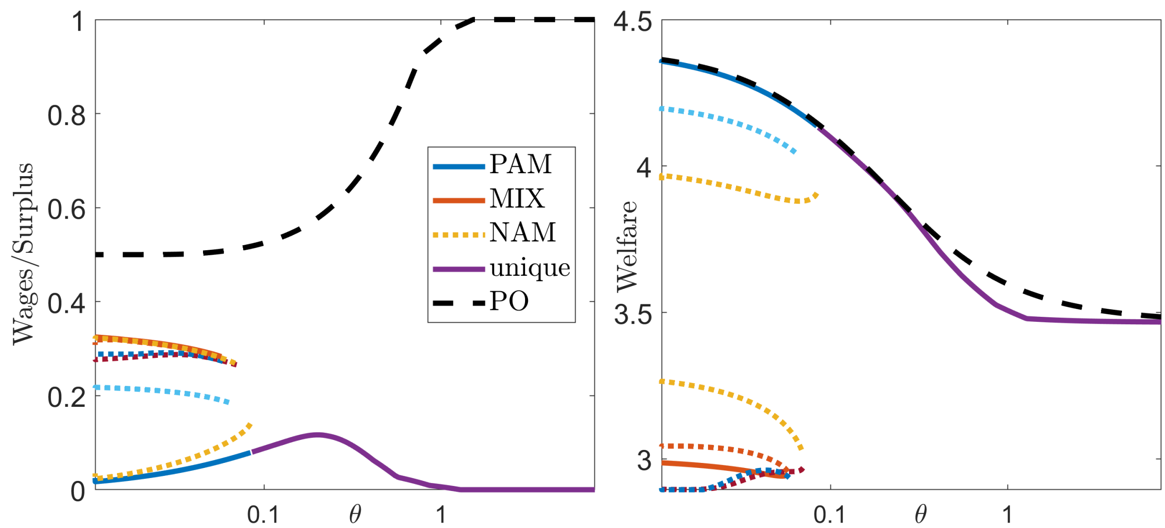
Figures 3.2 and 3.3 illustrate the equilibrium strategies of workers and firms, the surplus split and numbers of matches — for three equilibria (PAM, Mixing, NAM) and for the planner's solution (PO) — allowing us to compare them for horizontal and vertical structure of the surplus close to the zero-cost limit. In the horizontal case, shown in Figure 3.2, the positively assortative equilibrium comes qualitatively closest to the planner's solution, but it promises workers extremely low wages, while the planner still promises workers half the surplus. In the vertical case, shown in Figure 3.3, the mixing equilibrium is qualitatively the closest to the planner's solution, although it cannot fully achieve it as that would require firms to get substantially less than half the surplus, while workers would still get half. Notably, in both cases, in each assortative equilibrium, positive or negative, workers are promised extremely low wages. We observe that assortativeness reduces competition for workers and substantially increases the firms' monopsony power. Mixing equilibria imply a fairer split of the surplus, yet may lead to welfare losses.

In light of these results consistent with the derivations of the corresponding directed search model, we believe that our sequential targeted search model could serve as an effective diagnostic tool for studying properties of equilibria in directed search models. It is much easier to set up the model and find numerically all of its equilibria when costs approach zero, than considering manually the simplex of combinations of equilibrium equalities and inequalities when costs are equal to zero. A related approach has been proposed for convexifying combinatorial optimal transport problems — using an entropic regularization term disappearing in the limit — by Cuturi (2013) and in economics for convexifying the problem of plant location in space by Oberfield et al (2024). Our approach not only significantly speeds up the analysis, but also allows to quickly identify cases of inefficient directed search equilibria — by excluding the directed search equilibria which are unlikely to survive in practice.

To illustrate the power of the targeted search model, Figure 3.4 shows the paths of wages and welfare in the unique equilibrium (for high search costs) and in 8 equilibria (for low search costs) for a model with 3 worker and 3 firm types with surplus shape $f_{xy} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1.5 & 1 \\ 1 & 1 & 0.9 \end{bmatrix}$. As the surplus is relatively vertical, the unique equilibrium and then the best equilibrium are positively assortative (PAM), with wages approaching zero. There is also a negatively assortative (NAM) equilibrium, and a large number of mixing equilibria, one of which gives the highest surplus share to the workers but delivers an inferior level of welfare.

To summarize, the wage-posting competitive equilibrium in the sequential targeted search model fills in the continuum between random matching (when $\theta \to \infty$) and

Figure 3.4: Equilibrium wages and welfare for a 3-worker 3-firm model



directed search (when $\theta \to 0$). This model features all the defining assumptions of directed search models: a) search is sequential: firms post wages, workers apply, firms choose among those that applied; b) firms post type-specific wages strategically such that they attract specific kinds of workers; c) after deciding direction of search, workers and firms meet in submarkets each featuring a matching technology which determines the number of matches. The novel feature of our model is that we fill in this continuum by varying the degree to which firms and workers are able to inform themselves about the available options.

In the zero-cost limit our framework reduces to a standard directed search model. At the other extreme, in the infinite-cost limit, it converges to a random search model with wage posting, where like in Burdett and Mortensen (1998) wage dispersion coexists with random search. Our framework can generate outcomes ranging from complete monopsony, where the whole surplus goes to the firm, to perfect competition, where the whole surplus goes to the worker. As we illustrate in the next section, the outcome depends on labor market tightness — the ratio of firms to workers. As tightness rises, the effective offer arrival rate (in the sense of Burdett and Mortensen) increases, shifting the equilibrium from monopsony toward perfect competition. Thus, our model nicely captures for intermediate level of search costs a hybrid model having features of both directed search and random search a la Burdett and Mortensen.

We described several novel findings that shed light on properties of directed search models. First, for low enough levels of search costs, there are always multiple equilibria. These are typically overlooked or assumed away in the literature either through focusing only on specific types of equilibria, or through assuming that identical agents must play identical strategies, or due to incomplete consideration of the combinations of equilibrium conditions, some of which may hold as equalities and others as inequalities. In Appendix A, we consider a standard directed search model with 2 heterogeneous types of workers and firms. We fully derive the conditions for equilibrium and consider their properties and combinations. We show that there are always two pure-strategy equilibria (PAM and NAM) and there can be also mixed-strategy equilibria. We show that these directed search equilibria are the limiting cases to which the equilibria in our numerical examples converge as search costs approach zero. However, some of the directed search equilibria may not survive for any small positive costs.

The second, striking finding (novel, although reminiscent of BSW) is that the commonly assumed efficient split of the surplus between the worker and the firm based on the elasticities of the matching technology only holds for wages of workers which play mixed strategies and must be indifferent in equilibrium. In pure-strategy equilibria, characterized by assortativeness, firms post wages arbitrarily close to zero. This result is in striking contrast with the directed search literature which is used to assuming that the efficient split of the surplus is a uniform feature of directed search models.

Third, while we find competitive equilibria to be generically constrained-inefficient, in the zero-cost limit, one of the directed search equilibria may implement the planner's allocation. If the planner's allocation is assortative, then one of the assortative equilibria implements it with wages set close to zero. If the planner's allocation is in semi-mixed strategies, it may not survive for positive costs. In this case, all of the competitive equilibria may be socially suboptimal in the zero-cost limit.

Fourth, whether the best competitive equilibrium exhibits positive assortative matching depends on whether the strength of production complementarities exceeds matching complementarities, consistent with Eeckhout and Kircher (2010). In the numerical case we considered, this condition is $f_{1,1}f_{2,2} > f_{1,2}f_{2,1}$. When production complementarities are strong, competitive equilibria will often exhibit assortativeness, implying that firms face little competition for workers, and therefore gain monopsony power and use it to reduce promised wages. When production complementarities are weak, the best competitive equilibrium tends to exhibit a mixed sorting pattern, and is characterized by intense competition for workers which leads to a more even split of the surplus.

These findings challenge the view that directed search naturally leads to fair or efficient outcomes. In the next section, we shift focus to the middle ground where search frictions are small but finite, and study how wage-setting power evolves continuously across this space. We quantify how the interaction of frictions and sorting continuously affects wage markdowns and welfare.

4 Properties of wages and monopsony power

The key questions we seek to answer next are: (i) How do search costs shape equilibrium wages? (ii) What are the sources of monopsony power in this framework? (iii) How does labor market tightness interact with firm wage-setting strategies? This section presents analytical results that shed light on these questions, followed by numerical simulations that illustrate and validate our findings.

We start the discussion of properties of equilibrium from equation (2.8) which describes how the equilibrium posted wage is determined. To aid the following discussion, we substitute the symmetric Cobb-Douglas congestion function into the wage equation:

$$\omega_{xy}^{*} = \left[\underbrace{a_{xy} f_{xy}}_{\text{Marginal product}} - \underbrace{\frac{\theta_x}{\sqrt{q_y(x)\frac{\delta_x}{\delta_y}}} \frac{1}{\left(1 - \frac{\mu_y}{n_f} p_x(y)\right)}}_{\text{Monopsony discount, } 1/z_{xy}} + \underbrace{(1 - a_{xy}) \frac{\theta_y}{\sqrt{\frac{1}{q_y(x)} \frac{\delta_y}{\delta_x}}}}_{\text{Competitive premium}} \right]_0^{f_{xy}}$$

Let us unpack each of these terms. The first term promises the workers of type x a share of the suplus $a_{xy} = \frac{\mu_x p_x(y)}{\Sigma_{x=1}^{W} \mu_x p_x(y)}$ equal to their fraction in the queue of workers applying to positions at firms of type y. This term reflects two mechanisms. The workers are given an incentive to search harder so that if they are able to better self-select into this type of job, they will get a higher wage. This term also reflects the fact that if other types of workers do not apply to this job, then this type of workers faces less direct wage competition from other types of workers and can expect a higher wage.

The second term represents the firms' monopsony discount (we deliberately use a different word to distinguish it from the wage markdown). This term is reminiscent of the literature on monopsony power where the wage markdown is often shown to be proportional to the inverse of the labor supply elasticity. Recall that the variable z_{xy} was introduced as the semi-elasticity measuring the effect of an increase in the wage on the number of workers applying to the firm: $z_{xy} = \frac{\partial p_x(y)}{\partial \omega_{xy}} \frac{1}{p_x(y)}$. The more conventional labor supply elasticity in this case can be computed as follows: $\epsilon_{p,\omega} = \frac{\partial p_x(y)}{\partial \omega_{xy}} \frac{\omega_{xy}}{p_x(y)} = \omega_{xy} z_{xy}$.

The monopsony discount is driven by competition with other firms. It works towards equalizing the expected payoffs faced by the worker from this type of firm compared with other types of firms. In the formula above, we substituted $z_{xy} = \frac{\phi_{xy} q_y(x)}{\theta_x/\delta_x} \left(1 - \frac{\mu_y}{\delta_x} p_x\left(y\right)\right)$, derived from the strategies of the worker, and then substituted the congestion function. The interpretation of monopsony discount is similar to that in the literature deriving a multinomial logit form for the firms' labor supply from a distribution of idiosyncratic tastes. In our specification, labor supply also has the multinomial logit (MNL) form (2.4) derived from a micro-founded information search friction. The elasticity of labor supply in this MNL form is inversely proportional to the marginal search costs faced by workers, θ_x. As it gets harder for workers to distinguish between firms, labor supply becomes less elastic which lowers the equilibrium wage. The elasticity also depends on the ratio of the total number of firms faced by each worker of type x ($\delta_x = \Sigma_y \mu_y$) and the total number of workers that apply to each firm of type y ($\delta_y = \Sigma_x \mu_x p_x(y)$). As the firms' relative numerical disadvantage increases, the elasticity of labor supply decreases, the monopsony discount increases, and the equilibrium posted wage decreases.

The third term adds on top a premium proportional to the marginal search cost faced by the firms. As it gets harder for the firm to screen workers in their queue, they prefer to delegate some of that self-selection to the workers by promising a higher wage. The firms' marginal search cost acts as the inverse of the elasticity of labor demand in conventional models, analogous to the elasticity of labor supply determining the monopsony discount. The competitive premium also depends positively on the ratio of the overall number of firms considered by workers (δ_x) to the total number of workers that apply to firms (δ_y). If firms are at a numerical disadvantage (δ_x/δ_y is small), this increases the firms' monopsony power and decreases the equilibrium wage.

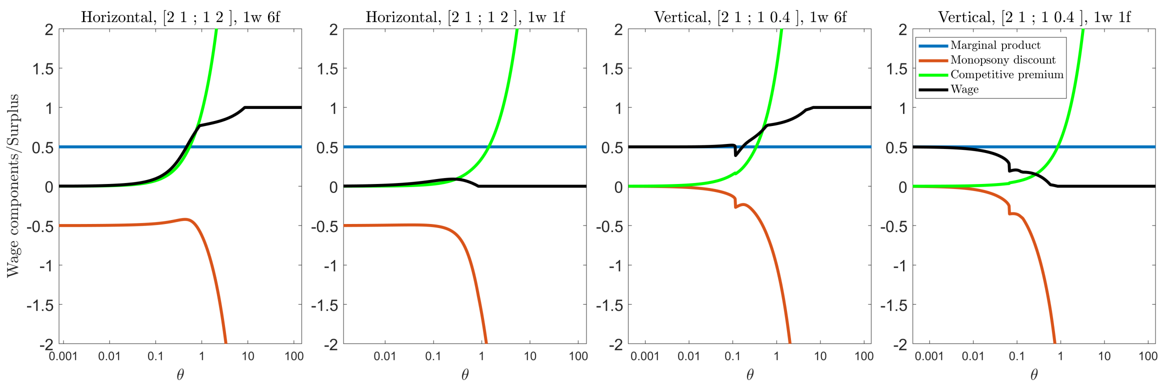
Equation 2.8 is analogous to the standard oligopsony model. Here we show the analogy following the notation of Boal and Ransom (1997). If the firm is facing an upward sloping labor supply curve, described by $w(L)$, and the firm maximizes expected profit, computed as revenue net of the wage bill ($\pi = R(L) - w(L) \cdot L$), then profit maximization pins down the wage as the marginal product minus the inverse semi-elasticity of labor supply: $w = R' - w'(L) \cdot L$. In our model, the first term $a_{xy} f_{xy}$ represents the marginal product adjusted for strategic competition of firms for workers' attention, captured by the queue composition a_{xy}, and the second term is the inverse of the firm-level labor supply semi-elasticity z_{xy}. The labor supply elasticity is derived from the (multinomial logit) discrete choice decision of the workers, which reflects endogenous responses to worker-job match heterogeneity and the information friction. The third term in our model represents the labor demand semi-elasticity which arises in the oligopsony model when e.g. revenue directly depends on the wage. This is true in our model because a higher wage not only incentivizes workers to search harder, but also incentivizes firms to screen less thoroughly, which also affects the matching rates. While the labor supply elasticity is determined by search costs of workers, the labor demand elasticity is deteremined by the search costs of firms, and both are modulated by labor market tightness.

Thus, our model is a direct next step from the existing literature - putting together in a single equation the models of oligopsony, models of job differentiation and search-and-matching models. While the elements of our model have familiar interpretation, they are derived from fundamentals — match preferences, search technology, endowments of agents — in an internally consistent way and are therefore suitable for welfare and policy analysis.

How do these three components evolve and interact to give the wage rates we saw in Figure 3.1? In panels 2 and 4 of Figure 4.1 we show the three components and the resulting wage for the best/unique equilibrium for horizontal and vertical shapes of the surplus we saw before. Each component here is averaged across submarkets. The marginal product averages to half of the surplus even though the components of the queue composition a_{xy} vary depending on the shape of the surplus and the implied sorting pattern. For the PAM equilibrium in the horizontal case the queue composition matrix is diagonal with ones on the diagonal and zeros otherwise, while for the mixing equilibrium in the vertical case the queue composition matrix is closer to uniform with all elements close to one half.

The monopsony discount is negative and grows with the increase in search costs of workers. We would have expected the monopsony discount to approach zero in the

Figure 4.1: Equilibrium wage components for horizontal and vertical surplus



zero-cost limit, and if that was the case all of the equilibria would give a fair split of the surplus in all of the directed search equilibria. Interestingly, in the zero-cost limit the monopsony discount approaches zero for mixed equilibria, but remains sizeable for pure-strategy equilibria. This is because in pure-strategy (assortative) equilibria the workers put all of their attention towards one firm, and therefore not only $\theta_x \to 0$ but also $\left(1 - \frac{\mu_y}{n_f} p_x(y)\right) \to 0$. As these two terms in the numerator and denominator approach zero at the same speed, the monopsony discount has a finite non-zero limit as search costs tend to zero for assortative equilibria. For mixing equilibria, the limit is zero.

The competitive premium is positive and grows with the increase in search costs of firms. This component always approaches zero as search costs of firms tend to zero. Thinking about the interaction of the three wage components, for low values of costs, the equilibrium wage is determined by the sorting pattern that determines how much the monopsony discount subtracts from the marginal product. As search costs increase, both the monopsony discount and the competitive premium grow without bound. Depending on which of the two grows faster, the equilibrium wage, confound to stay between the outside option and the surplus, reaches either the upper bound or the lower bound as costs tend to infinity. As we can see from the wage equation, which of the two forces dominates is determined by the relative sizes of search costs, and by the labor market tightness. To illustrate this dependence on labor market tightness, in panels 1 and 3 of the same Figure we plot the wage decomposition for the same setups but with a different level of labor market tightness: for 1 worker of each type there are 6 firms of each type. In this case, as search costs grow, the competitive premium increases faster than the monopsony discount, and the equilibrium wage quickly approaches the upper bound.

For further comparison with the literature, we can also derive the equilibrium markdown. Here we need to note that we have assumed, without loss of generality, that the outside option of workers is normalized to 0. For the definition and appropriate computation of the markdown (and the labor supply elasticity) we need to calibrate the size of the outside option b relative to the size of the surplus f_{xy}. Taking into account the outside option, the equilibrium markdown can be computed as:

$$\frac{f_{xy} + b - \omega_{xy} - b}{\omega_{xy} + b} = \frac{1 + \frac{b}{f_{xy}}}{\frac{\omega_{xy}}{f_{xy}} + \frac{b}{f_{xy}}} - 1 = \frac{1 + \frac{b}{f_{xy}}}{a_{xy} + \frac{(1-a_x)}{\phi_{xy}\delta_y}\frac{\theta_y}{f_{xy}} - \frac{1}{z_{xy}}\frac{1}{f_{xy}} + \frac{b}{f_{xy}}} - 1$$

In standard models, wage markdowns are often treated as a direct function of labor supply elasticity. In our framework, however, this relationship is mediated by equilibrium sorting and strategic wage posting. Although a semi-elasticity term $(\omega_{xy} + b)\, z_{xy}$ enters the wage equation, there is no one-to-one mapping between this elasticity and the resulting markdown. Instead, equilibrium wages and markdowns are shaped by five interacting factors:

1. The workers' marginal costs of search determine the labor supply elasticity and the monopsony discount.

2. The firms' marginal costs of search determine the labor demand elasticity and the competitive premium.

3. The numerical (dis)advantage each firm and worker face in the market (labor market tightness) attenuate both elasticities.

4. The strength and pattern of equilibrium sorting determines the starting point for the equilibrium wage and markdown.

5. The first-mover advantage (the ability of firms to announce and pre-commit to submarket-specific wage) gives the firms the ability to strategically manipulate wages and extract the monopsony discount. Unlike this model, in a model of simultaneous search with bargaining the competitive equilibrium implements the constrained social optimum.

This theoretical analysis highlights several key drivers of monopsony power, including search costs, labor market tightness, and equilibrium sorting patterns. However, the strength of these effects in realistic labor markets remains an open question. To quantify these mechanisms and compare them to empirical benchmarks, we now turn to numerical simulations using calibrated parameter values.

We calibrate the outside option of the workers b to approximately 70% of their product of labor $f_{xy}+b$, in the middle of the range of calibrations of DMP models. This implies a value of $b/f_{xy} = 2.5$ and a maximum markdown of 40%. For consistency, we return to the horizontal and vertical surplus shapes considered in the previous section. We consider a two-dimensional set of all combinations of costs $\theta_x = \theta_y = \theta$ and the ratio of the number of jobs to workers μ_y/μ_x, a variable also known as labor market tightness. For each combination of $\{\theta, \mu_y/\mu_x\}$, we compute all the equilibria and the socially optimal allocation.

In Figures 4.2 and 4.3 we show in six panels how the number of equilibria, monopsony power of (share of the surplus going to) the firm, the elasticity of labor supply with respect to the wage, the average equilibrium markdown, the socially optimal markdown, and welfare — vary with search costs and labor market tightness. Figure 4.2 shows the results for a horizontal structure of preferences, consistent with assumptions for column 3 of Figure 3.1, which produces positive assortative matching in the best equilibrium. Figure 4.3 shows the results for a vertical shape of surplus, consistent with assumptions for column 4 of Figure 3.1, which produces mixed sorting in the best equilibrium. In both cases there is an area of low costs producing 3 equilibria: PAM, mixed, and NAM — with only the mixed equilibrium surviving for higher costs in the vertical case, and only PAM in the horizontal case.

Both Figures illustrate how the equilibria of the model gradually transform from multiple directed search equilibria (for low costs, on the left) to a unique random search wage posting equilibrium a la Burdett and Mortensen (for high costs, on the right). Both Figures also illustrate how monopsony power changes with labor market tightness, from pure monopsony and wages equal to outside options of workers (low market tightness, at the bottom) to perfect competition and wages equal to the marginal product of labor (high market tightness, at the top).

Overall, the patterns of monopsony power and markdowns are not very different between the two surplus calibrations. Outcomes corresponding to labor market tightness in the range from 0.5 to 1.5, which are routinely observed in the U.S. labor market, pro-

Figure 4.2: Equilibrium monopsony power and wage markdowns for horizontal case

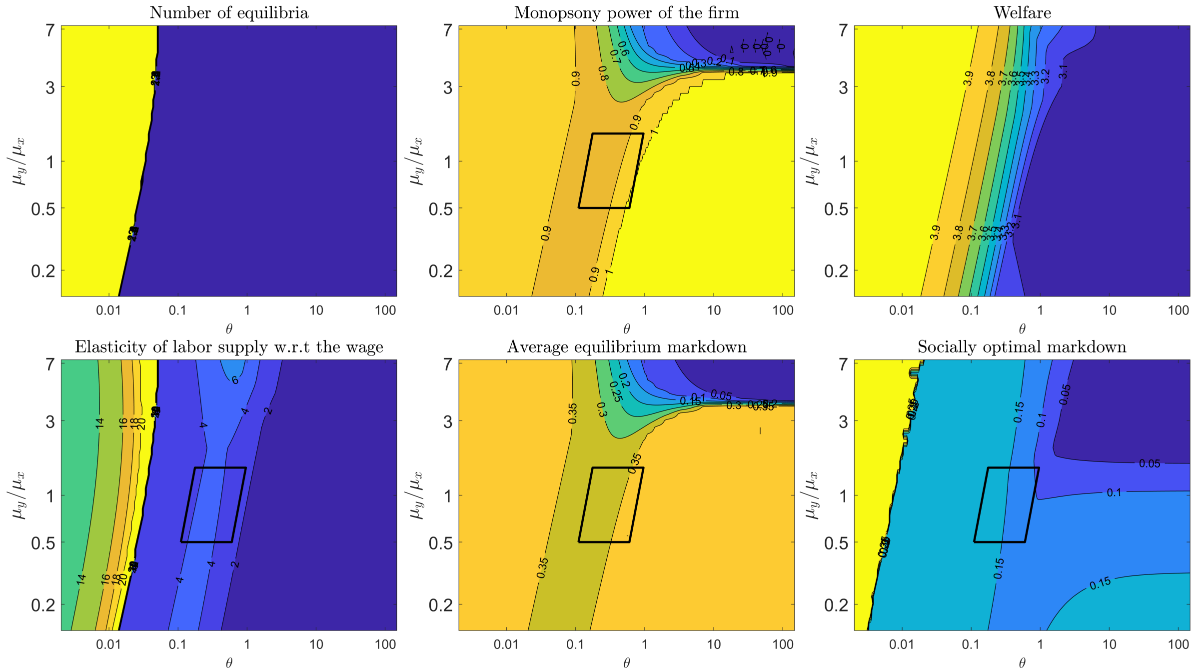
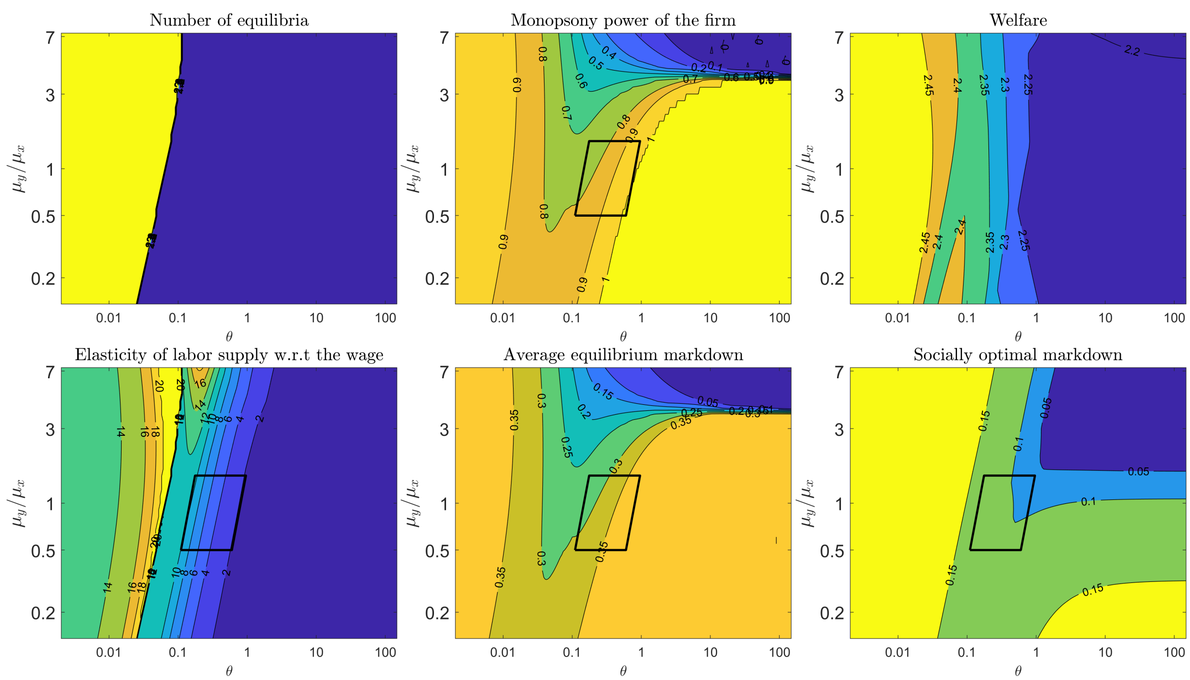


Figure 4.3: Equilibrium monopsony power and wage markdowns for vertical case



duce markdowns on the order of 30%-40%. This corresponds to extremely high levels of monopsony power on the part of firms, capturing between 70% and 100% of match surplus. Consistent with our theoretical derivations, the elasticity of labor supply is tightly linked with the inverse of marginal search costs faced by workers. Therefore, the empirical estimates of the firm-level labor supply elasticity which are usually in the 2-6 range[4], put a relatively tight bound on the value of search costs θ which should be in the 0.1-0.5 range. The bounds on the parameter range implied by empirical evidence are shown in the Figures by the black polygons. As we can see from numerical simulations, this is well above the level of costs, below which the model produces multiple equilibria. This also suggests that both wage markdowns are unfairly high and the market allocations are socially suboptimal.

What are the possible ways to correct the inefficient allocations and unfair wages? Here we only provide some suggestions and leave more detailed analysis for future research. As indicated by the top right panels of Figures 3.2, 3.3, 4.2, and 4.3, a constrained social planner would prescribe a redistribution scheme equivalent to increasing wages promised to workers by about 20%, lowering wage markdowns towards the 10-15% range. It would also suggest substantially taxing the firms' part of the surplus to reduce their incentive to screen. But this tax and redistribution scheme is unlikely to work so long as the firms retain the first-mover advantage and ability to strategically set wages. Note that the simultaneous search version of the model with ex-post bargaining, described in Cheremukhin et al (2020), for the same fundamentals delivers the socially optimal allocation of this model as the unique competitive equilibrium. Removing the sequentiality and first-mover advantage of firms could go a long way towards ensuring both efficiency and fairness. If the sequentiality of search is too deeply rooted and hard to combat, another alternative would be to combat wage-posting and incentivize ex-post bargaining with a coordinated effort to prop up the workers' bargaining positions.

As an illustration, we compute competitive equilibria of the sequential targeted search model replacing ex ante wage posting with an exogenously set ex post wage bargaining. We jointly vary bargaining weights in all submarkets seeking the bargaining weights that lead to highest welfare, which we call the optimal bargaining weight equilibrium. In Figure 4.4 we show how the wages and welfare loss depend on the

[4] See Azar and Marinescu (2024a) for an overview of the the estimation methods and empirical estimates, and Sokolova and Sorensen (2018) for a meta-analysis.

Figure 4.4: Wages and welfare loss under wage posting and optimal bargaining

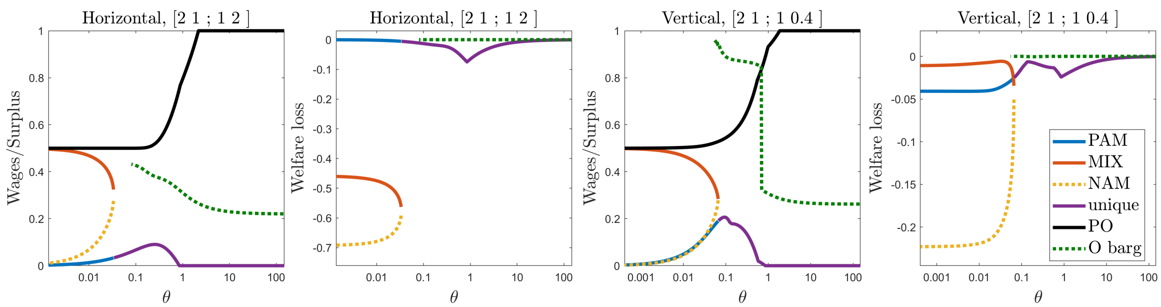
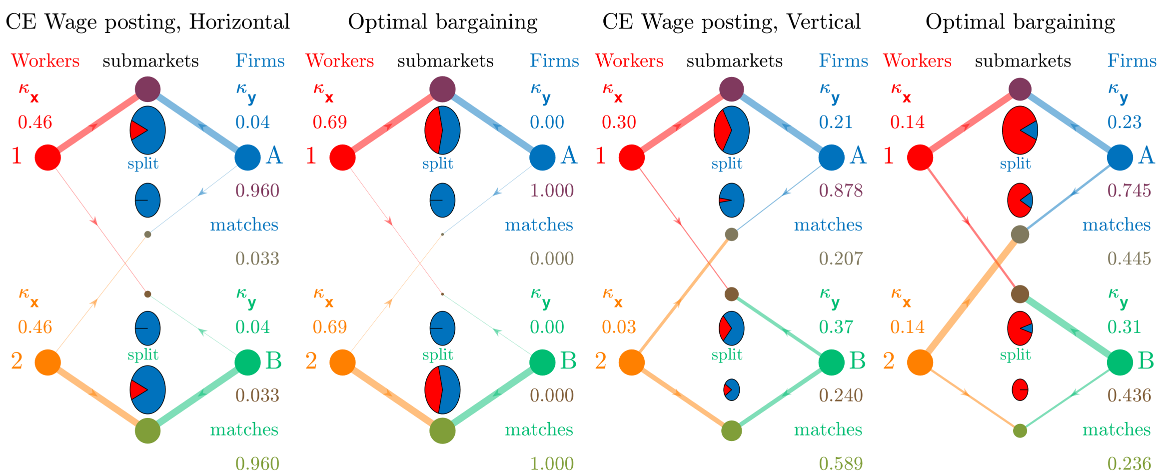


Figure 4.5: Strategies and wages under wage posting and optimal bargaining for $\theta = 0.1$



search costs for the horizontal and vertical surplus shapes. In Figure 4.5 we compare the strategies and wages prevailing under the competitive equilibria of ex ante wage posting and optimal ex post wage bargaining for an intermediate value of search costs $\theta = 0.1$ (consistent with realistic levels of labor supply elasticity and unique equilibrium) for the same horizontal and vertical surplus shapes. We find that for both shapes of the surplus switching to optimal ex post bargaining substantially improves welfare, essentially removing welfare loss, which requires substantially higher wages than those naturally prevail in wage posting equilibria. These results suggest that a much fairer surplus split enforced through bargaining could lead to substantial welfare improvements in the sequential model.

Our model can be used for all kinds of policy evaluation.[5] To study the effects of changes in the minimum wage one could increase the outside option of workers in our model and consider the effects on wages and matching rates. To study the effects of wage transparency policies one could compare the model with ex ante wage posting with a model that counterfactually enforces ex post bargaining. Since the model does not make short-cut assumptions and the user controls preferences, technologies and endowments, the model should be directly applicable to a wide range of policy questions.

5 Conclusion

This paper develops a unified theory of monopsony power based on information frictions in job search. We model both firms and workers as choosing probabilistic targeting strategies, subject to endogenous information costs. This framework nests random and directed search as limiting cases and delivers a tractable wage-posting equilibrium with a closed-form wage equation. Monopsony power in this setting arises not only from labor supply elasticity but from five structural channels: information frictions on both sides of the market, labor market tightness, equilibrium sorting, and the sequential nature of wage setting.

We show how wage markdowns emerge endogenously from the interaction of these forces and that equilibrium sorting patterns play a central role in shaping the degree of monopsony power. In highly assortative environments with strong complementarities, wages collapse toward the outside option even when labor supply is elastic, allowing firms to capture nearly the full surplus. These outcomes are inefficient: a constrained social planner would prescribe lower targeting precision and higher wages, closing roughly half of the markdown gap.

Our approach offers a tractable alternative to models that impose wage-setting rules or estimate reduced-form elasticities. By deriving wages, applications, and sorting jointly from primitives, we clarify the mechanisms behind monopsony and provide tools

[5] Azar and Marinescu (2024b) survey recent theoretical models of monopsony power and link them to a broad set of labor market policies. They emphasize the importance of developing structural models that can accommodate information frictions and strategic interactions when evaluating interventions such as minimum wages, wage transparency mandates, and non-compete regulations. Our framework is well-suited for such applications, as it nests key modeling approaches while preserving microfoundations and tractability.

for analyzing the efficiency and equity of market outcomes. This framework can be extended to study policy design, wage floors, or labor market segmentation in settings where information and strategic interaction jointly determine matching patterns and pay.

References

[1] **Azar, J. and I. Marinescu (2024a)** "Monopsony Power in the Labor Market" Chapter 10 in *Handbook of Labor Economics*, Volume 5, pp. 761-82. https://doi.org/10.1016/bs.heslab.2024.11.007

[2] **Azar, J. and I. Marinescu (2024b)** "Monopsony Power in the Labor Market: From Theory to Policy" *Annual Review of Economics*, 16, pp. 491-518. https://doi.org/10.1146/annurev-economics-072823-030431

[3] **Azar, J., Marinescu, I., and M. Steinbaum (2019)** "Measuring Labor Market Power Two Ways" *AEA Papers and Proceedings*, 109, pp. 317-321. https://doi.org/10.1257/pandp.20191068

[4] **Berger, D., Herkenhoff, K. and S. Mongey (2022)** "Labor Market Power" *American Economic Review*, 112(5), pp. 1147-1193. https://doi.org/10.1257/aer.20191521

[5] **Berger, D., Herkenhoff, K., Kostol, A. and S. Mongey (2023)** "An Anatomy of Monopsony: Search Frictions, Amenities and Bargaining in Concentrated Markets." NBER Working Paper 31149.

[6] **Boal, William M., and Michael R. Ransom (1997)** "Monopsony in the Labor Market." *Journal of Economic Literature*, 34(1), pp. 86-112. https://www.jstor.org/stable/2729694

[7] **Burdett, K. and D.T. Mortensen (1998)** "Wage differentials, employer size, and unemployment." *International Economic Review*, 39(2), pp. 257-273. https://doi.org/10.2307/2527292

[8] **Burdett, K., Shi, S., and R. Wright (2001)** "Pricing and Matching with Frictions." *Journal of Political Economy*, 109(5), pp. 1060-1085. https://doi.org/10.1086/322835

[9] **Card, D. (2022)** "Who Set Your Wage?" *American Economic Review*, 112(4), pp. 1075-1090. https://doi.org/10.1086/322835

[10] **Chade, H., Eeckhout, J. and L. Smith (2017).** "Sorting Through Search and Matching Models in Economics." *Journal of Economic Literature*, 55(2), pp. 493-544. https://doi.org/10.1257/jel.20150777

[11] **Cheremukhin, A., Popova A. and A. Tutino (2015).** "A Theory of Discrete Choice with Information Costs." *Journal of Economic Behavior and Organization*, 113, pp. 34-50. https://doi.org/10.1016/j.jebo.2015.02.022

[12] **Cheremukhin, A., Restrepo-Echavarria P., and A. Tutino (2020).** "Targeted Search in Matching Markets." *Journal of Economic Theory*, 185, January. https://doi.org/10.1016/j.jet.2019.104956

[13] **Cuturi, M. (2013)** "Sinkhorn distances: Lightspeed computation of optimal transport." In Advances in Neural Information Processing Systems, pages 2292-2300.

[14] **Dube, A., Naidu S., and A.D. Reich (2022)** "Power and dignity in the low-wage labor market: theory and evidence from Wal-Mart workers." NBER Working Paper 30441.

[15] **Eeckhout, J. and P. Kircher (2010).** "Sorting and Decentralized Price Competition." *Econometrica*, 78(2), pp. 539-574. https://doi.org/10.3982/ECTA7953

[16] **Guerrieri, V., Julien, B., Kircher, P. and R. Wright (2021)** "Directed Search and Competitive Search Equilibrium: A Guided Tour." *Journal of Economic Literature*, 59(1), pp. 90-148. https://doi.org/10.1257/jel.20191505

[17] **Lamadon, T., Mogstad, M., and B. Setzler (2022)** "Imperfect competition, compensating differentials, and rent sharing in the US labor market." *American Economic Review* 112 (1), 169-212. https://doi.org/10.1257/aer.20190790

[18] **Lentz, R., Maibom, J., and E. Moen (2022)** "Competitive or random search?" unpublished manuscript.

[19] **Lester, B. (2011).** "Information and Prices with Capacity Constraints," *American Economic Review*, 101(4), pp. 1591-1600. https://doi.org/10.1257/aer.101.4.1591

[20] **Matejka F. and A. McKay (2015).** "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, 105(1), pp. 272-298. https://doi.org/10.1257/aer.20130047

[21] **Manning, A., 2003.** "Monopsony in Motion: Imperfect Competition in Labor Markets." Princeton Univ. Press, Princeton, NJ

[22] **Menzio, G. (2007).** "A Theory of Partially Directed Search," *Journal of Political Economy*, 115(5), pp. 748-769. https://doi.org/10.1086/525544

[23] **Menzio, G., and S. Shi (2010).** "Block recursive equilibria for stochastic models of search on the job" *Journal of Economic Theory*, 145(4), pp. 1453-1494. https://doi.org/10.1016/j.jet.2009.10.016

[24] **Moen, E.R. (1997)** "Competitive search equilibrium." *Journal of Political Economy*, 105 (2), pp. 385-411. https://doi.org/10.1086/262077

[25] **Mortensen, D. T. (1982).** "Property Rights and Efficiency in Mating, Racing, and Related Games," *American Economic Review*, 72(5), pp. 968-979.

[26] **Oberfield, Ezra, Rossi-Hansberg, Esteban, Sarte, Pierre-Daniel, and Nicholas Trachter (2024)** "Plants in Space." *Journal of Political Economy*, 132 (3), pp. 867-909.

[27] **Rabinovich, S., and R. Wolthoff (2022)** "Misallocation inefficiency in partially directed search." *Journal of Economic Theory*, 206(C).

[28] **Robinson, J. (1933)** "Imperfect competition." In: Robinson, J. (1966), An Essay on Marxian Economics. Palgrave MacMillan, UK, pp. 73-81. https://doi.org/10.1007/978-1-349-15228-5_9.

[29] **Shi, S. (2001).** "Frictional Assignment. I. Efficiency." *Journal of Economic Theory*, 98, pp. 232-260. https://doi.org/10.1006/jeth.2000.2713

[30] **Shimer, R. and L. Smith (2001).** "Matching, Search and Heterogeneity." The B.E. Journal of Macroeconomics, 1(1), pp. 1-18. https://doi.org/10.2202/1534-6013.1010

[31] **Sokolova, A., and T. Sorensen (2021)** "Monopsony in labor markets: a meta-analysis." *ILR Review* 74 (1), 27-55. https://doi.org/10.1177/0019793920965562

[32] **Staiger, D. O., J. Spetz, and C. S. Phibbs (2010)** "Is there monopsony in the labor market? evidence from a natural experiment." *Journal of Labor Economics* 28(2), pp. 211-236.

[33] **Tarski, A. (1955).** "A Lattice-theoretical Fixpoint Theorem and Its Applications," *Pacific Journal of Mathematics*, 5, pp. 285-309. https://doi.org/10.2140/pjm.1955.5.285

[34] **Thomas, C. and J. Cover (1991)**. "Elements of Information Theory." John Wiley & Sons, Inc.

[35] **Wu L. (2024)** "Partially directed search in the labor market." *Journal of Economic Theory* 220(C).

Appendix A: Directed Search Model Analysis

In this appendix, we solve for directed search equilibria in a simple environment with 2 worker and 2 firm types. Following the abbreviations from Burdett Shi and Wright (BSW, 2001), who describe buyers and sellers instead of workers and firms, we denote two worker types 1 and 2, and the two firm types we denote A and B.

We extend their assumptions in several respects. Our first departure: instead of agents' strategies being probabilities with which workers and firms look at each other, we assume that there is a unit mass of agents of each type, and they decide which fraction of that unit mass to send towards which type of agent. More specifically, workers of type 1 send m_{1A} mass to meet firms of type A and the remaining $1 - m_{1A}$ are sent to meet firms of type B. Workers of type 2 send m_{2B} mass to meet firms of type B and the remaining $1 - m_{2B}$ are sent to meet firms of type A. Both the reinterpretations of probabilities into masses and the default assignment of 1 to A and 2 to B are for convenience and without loss of generality. Similarly, firms of type A send mass n_{1A} to meet workers of type 1, $1 - n_{1A}$ to meet type 2, and firms of type B send n_{2B} to meet workers of type 2 and $1 - n_{2B}$ to meet type 1.

In a second extension of BSW, and consistent with conventional assumptions in the directed search literature, we assume four meeting locations where each pair of agents can meet separately: $1A, 1B, 2A, 2B$. We also allow each firm to post differing wage rates at these four locations. In BSW, all pairwise matches were identical and produced identical surplus, therefore, the equilibrium surplus splits were assumed to be identical in the locations controlled by the same firms. Since our extension allows for heterogeneity and unrestricted surplus shapes, the increased number of locations is natural. As we shall see, under the special case considered by BSW our model produces identical results, but it is more general in allowing us to consider a wider range of possibilities.

We allow for arbitrary shapes of surplus achieved by pairwise matches and denote them $f_{1A}, f_{2A}, f_{1B}, f_{2B}$, all strictly greater than 0, and normalize the default outside option of workers to 0. In BSW all values of the surplus equal 1. We denote the wages posted by firms $w_{1A}, w_{2A}, w_{1B}, w_{2B}$ respectively. Instead of each firm posting a single wage rate, firms post different wages in worker-specific locations, or equivalently post wage contracts conditional on productivity.

The last and key element — we assume that each meeting location is characterized by a matching technology which determines how many matches are formed between mass m of workers and mass n of firms. The BSW paper makes calculations in expected value terms and assumes that the number of matches formed is the product of probabilities. It actually assumes without proof that if both workers show up at location A, their chances of being picked by the firm are equal. As we shall see, it may be optimal for the firm to choose unequal probabilities, and it is mostly a coincidence that for the specific surplus considered in the BSW paper that optimal probability is indeed one half. To replicate the predictions of the BSW model, we need to set the number of matches $M(m,n) = m \cdot n$. To represent the more conventional case of constant returns to scale (and make the model similar to our targeted search model, both the simultaneous and sequential versions) we also study the case $M(m,n) = \sqrt{m \cdot n}$. We note here that although this type of function has objectional properties, such as combinations for which the number of matches exceeds the number of searchers on one side, $M(m,n) > \min(m,n)$, this is without loss of generality since one can multiply the function by a small enough constant, such as $\frac{1}{2}$, removing this objection without affecting any of the properties of the model.

The game is sequential, and the timing works as follows. First, each firm posts wages in each of the two locations. Then each worker and each firm decides how to allocate their unit mass between the two available locations to maximize expected payoffs. The problem of worker 1 is to maximize her expected payoff:

$$Ew_1 = m_{A1} \cdot mr_{w1A} \cdot w_{A1} + (1 - m_{A1}) \cdot mr_{w1B} \cdot w_{B1} \to \max_{m_{A1}},$$

where the worker takes matching rates $mr_{w1A} = \frac{M(m_{1A},n_{1A})}{m_{1,A}}$, $mr_{w1B} = \frac{M(1-m_{1A},1-n_{B2})}{1-m_{1,A}}$ and posted wages as given. The solution to this linear problem is:

$$m_{A1} = \begin{cases} 1, & mr_{w1A} \cdot w_{A1} > mr_{w1B} \cdot w_{B1} \\ \in [0,1], & mr_{w1A} \cdot w_{A1} = mr_{w1B} \cdot w_{B1} \\ 0, & mr_{w1A} \cdot w_{A1} < mr_{w1B} \cdot w_{B1} \end{cases}$$

Similarly, the problem of worker 2 is to maximize her expected payoff:

$$Ew_2 = (1 - m_{B2}) \cdot mr_{w2A} \cdot w_{A2} + m_{B2} \cdot mr_{w2B} \cdot w_{B2} \to \max_{m_{B2}}$$

where the worker takes matching rates $mr_{w2A} = \frac{M(1-m_{B2},1-n_{A1})}{1-m_{B2}}$, $mr_{w2B} = \frac{M(m_{B2},n_{B2})}{m_{B2}}$ and posted wages as given. The solution to this linear problem is:

$$m_{B2} = \begin{cases} 1, & mr_{w2A} \cdot w_{A2} < mr_{w2B} \cdot w_{B2} \\ \in [0,1], & mr_{w2A} \cdot w_{A2} = mr_{w2B} \cdot w_{B2} \\ 0, & mr_{w2A} \cdot w_{A2} > mr_{w2B} \cdot w_{B2} \end{cases}$$

Firm A maximizes her expected payoff:

$E\pi_A = n_{A1} \cdot mr_{f1A} \cdot (f_{A1} - w_{A1}) + (1 - n_{A1}) \cdot mr_{f2A} \cdot (f_{A2} - w_{A2}) \rightarrow \max_{n_{A1}, w_{A1}, w_{A2}}$ where the firm considers its effect on matching rates $mr_{f1A} = \frac{M(m_{1A}, n_{1A})}{n_{1A}}$, $mr_{f2A} = \frac{M(1-m_{B2}, 1-n_{A1})}{1-n_{A1}}$. It also chooses wages, internalizing their interaction with the matching rate, which nevertheless needs to respect the prevailing relationship of indifference for each worker. It takes the strategies of the other firm as given, however.

Consider first wage setting for location $1A$. If worker 1 plays a pure strategy m_{1A} equal to 0 or 1, the firm can take the matching rate as given and maximize $(f_{A1} - w_{A1})$. This implies that the wage w_{A1} can be set arbitrarily close to 0, yet above 0, since a wage equal to 0 would break the optimality of the pure strategy played by the worker.

If worker 1 is indifferent and plays a mixed strategy, the firm aims to choose a wage level taking into account how that wage rate would influence the choice of the worker's strategy m_{A1}, how that would affect the matching rate mr_{w1A} as it would move (opposite) with w_{A1} to preserve the indifference: $mr_{w1A} \cdot w_{A1} = mr_{w1B} \cdot w_{B1}$. We can therefore write the problem of the firm with respect to wage w_{A1} as follows:

$\frac{M(m_{A1}, n_{A1})}{n_{A1}} \cdot (f_{A1} - w_{A1}) -> \max_{n_{A1}, w_{A1}} \quad \text{s.t.} \quad \frac{M(m_{A1}, n_{A1})}{m_{A1}} \cdot w_{A1} = mr_{w1B} \cdot w_{B1}$

$\Lambda = \frac{M(m_{A1}, n_{A1})}{n_{A1}} \cdot (f_{A1} - w_{A1}) + \lambda \left(mr_{w1B} \cdot w_{B1} - \frac{M(m_{A1}, n_{A1})}{m_{A1}} \cdot w_{A1} \right) -> \max_{n_{A1}, w_{A1}}$

$\text{FOC}_{n_{A1}}: \quad \left(\frac{\partial M(m_{A1}, n_{A1})}{\partial n_{A1}} - \frac{M(m_{A1}, n_{A1})}{n_{A1}} \right) \cdot \frac{(f_{A1} - w_{A1})}{n_{A1}} = \lambda \frac{\partial M(m_{A1}, n_{A1})}{\partial n_{A1}} \cdot \frac{w_{A1}}{m_{A1}}$

$\text{FOC}_{w_{A1}}: \quad -\frac{M(m_{A1}, n_{A1})}{n_{A1}} - \lambda \frac{M(m_{A1}, n_{A1})}{m_{A1}} = 0$

Find λ and substitute: $\quad \frac{\left(\frac{\partial M(m_{A1}, n_{A1})}{\partial n_{A1}} - \frac{M(m_{A1}, n_{A1})}{n_{A1}} \right)}{\frac{\partial M(m_{A1}, n_{A1})}{\partial n_{A1}}} \cdot \frac{(f_{A1} - w_{A1})}{w_{A1}} = -1$

$w_{A1} = \left(1 - \frac{\partial M(m_{A1}, n_{A1})}{\partial n_{A1}} \frac{n_{A1}}{M(m_{A1}, n_{A1})} \right) f_{A1} = \alpha f_{A1}$

The equilibrium posted wage is a fraction of the surplus equal to one minus the elasticify of the matching function with respect to the firm's decision. When the worker needs to remain indifferent, the optimal wage decouples from overall optimization and equals a fraction of the surplus, which depends on the curvature of the matching function. The same principle works for all four wages posted by firms. In addition, the firm must itself be indifferent between the options to optimize the wage that determines the payoff of any one of the options. Conditional on the wage decision and the strategies of workers, the problem of the firm remains how to optimally distribute its mass:

$E\pi_A = n_{A1} \cdot mr_{f1A} \cdot (f_{A1} - w_{A1}) + (1 - n_{A1}) \cdot mr_{f2A} \cdot (f_{A2} - w_{A2}) \to \max_{n_{A1}}$

$$n_{A1} = \begin{cases} 1 & mr_{f1A} \cdot (f_{A1} - w_{A1}) > mr_{f2A} \cdot (f_{A2} - w_{A2}) \\ n_{A1} & mr_{f1A} \cdot (f_{A1} - w_{A1}) = mr_{f2A} \cdot (f_{A2} - w_{A2}) \\ 0 & mr_{f1A} \cdot (f_{A1} - w_{A1}) < mr_{f2A} \cdot (f_{A2} - w_{A2}) \end{cases}$$

Similarly, for the second firm,

$E\pi_B = (1 - n_{B2}) \cdot mr_{f1B} \cdot (f_{B1} - w_{B1}) + n_{B2} \cdot mr_{f2B} \cdot (f_{B2} - w_{B2}) \to \max_{n_{B2}}$

$$n_{B2} = \begin{cases} 1 & mr_{f1B} \cdot (f_{B1} - w_{B1}) < mr_{f2B} \cdot (f_{B2} - w_{B2}) \\ n_{B2} & mr_{f1B} \cdot (f_{B1} - w_{B1}) = mr_{f2B} \cdot (f_{B2} - w_{B2}) \\ 0 & mr_{f1B} \cdot (f_{B1} - w_{B1}) > mr_{f2B} \cdot (f_{B2} - w_{B2}) \end{cases}$$

Returning to the description of wages, wage w_{A1} needs to be set optimally if either of the conditions involving it is satisfied as equality. Therefore,

$$w_{A1} = \begin{cases} \alpha f_{A1}, & mr_{w1A} \cdot w_{A1} = mr_{w1B} \cdot w_{B1} \\ \alpha f_{A1}, & mr_{f1A} \cdot (f_{A1} - w_{A1}) = mr_{f2A} \cdot (f_{A2} - w_{A2}) \\ 0 + \varepsilon, & else \end{cases}$$

$$w_{A2} = \begin{cases} \alpha f_{A2}, & mr_{w2A} \cdot w_{A2} = mr_{w2B} \cdot w_{B2} \\ \alpha f_{A2}, & mr_{f1A} \cdot (f_{A1} - w_{A1}) = mr_{f2A} \cdot (f_{A2} - w_{A2}) \\ 0 + \varepsilon, & else \end{cases},$$

$$w_{B1} = \begin{cases} \alpha f_{B1}, & mr_{w1A} \cdot w_{A1} = mr_{w1B} \cdot w_{B1} \\ \alpha f_{B1}, & mr_{f1B} \cdot (f_{B1} - w_{B1}) = mr_{f2B} \cdot (f_{B2} - w_{B2}) \\ 0 + \varepsilon, & else \end{cases},$$

$$w_{B2} = \begin{cases} \alpha f_{B2}, & mr_{w2A} \cdot w_{A2} = mr_{w2B} \cdot w_{B2} \\ \alpha f_{B2}, & mr_{f1B} \cdot (f_{B1} - w_{B1}) = mr_{f2B} \cdot (f_{B2} - w_{B2}) \\ 0 + \varepsilon, & else \end{cases}.$$

This completes the description of optimal strategies $\{m_{A1}, m_{B2}, n_{A1}, n_{B2}\}$ and wages $\{w_{1A}, w_{2A}, w_{1B}, w_{2B}\}$. Any internally consistent set of these variables that satisfies the optimality conditions is an equilibrium. However, this intersection can be tricky.

First, note that the pure strategy combinations $\{m_{A1}, m_{B2}, n_{A1}, n_{B2}\} = \{1, 1, 1, 1\}$ and $\{0, 0, 0, 0\}$ with corresponding to wages $\{w_{1A}, w_{2A}, w_{1B}, w_{2B}\} = \{\varepsilon, \varepsilon, \varepsilon, \varepsilon\}$ approaching 0 are always equilibria. We denote them PAM and NAM representing positive and negative assortative matching. It is easy to check and verify for PAM:

$mr_{w1A} = 1, mr_{w1B} = 0, mr_{w2A} = 0, mr_{w2B} = 1, mr_{f1A} = 1, mr_{f2A} = 0, mr_{f1B} = 0, mr_{f2B} = 1, \quad w_{A1} > 0 \quad w_{B2} > 0 \quad (f_{A1} - w_{A1}) > 0 \quad (f_{B2} - w_{B2}) > 0$

Same for NAM:

$mr_{w1A} = 0, mr_{w1B} = 1, mr_{w2A} = 1, mr_{w2B} = 0, mr_{f1A} = 0, mr_{f2A} = 1\, mr_{f1B} = 1, mr_{f2B} = 0 \quad w_{B1} > 0 \quad w_{A2} > 0 \quad (f_{A2} - w_{A2}) > 0 \quad (f_{B1} - w_{B1}) > 0.$

When the matching function is symmetric CRS some rates are undetermined, but we can consider strategies simultaneously approaching 0 and 1 and take the limiting cases: $mr_{w1A} = 1, mr_{w1B} = 1, mr_{w2A} = 1, mr_{w2B} = 1, mr_{f1A} = 1, mr_{f2A} = 1, mr_{f1B} = 1, mr_{f2B} = 1\; w_{A1} = 0 + \varepsilon_1 \quad w_{A2} = 0 + \varepsilon_2 \quad w_{B1} = 0 + \varepsilon_3 \quad w_{B2} = 0 + \varepsilon_4$

PAM: $w_{A1} > w_{B1}, w_{A2} < w_{B2}, (f_{A1} - w_{A1}) > (f_{A2} - w_{A2}), (f_{B1} - w_{B1}) < (f_{B2} - w_{B2})$

Can be satisfied if: $\quad f_{A1} + f_{B2} > f_{A2} + f_{B1}$

NAM: $w_{A1} < w_{B1}, w_{B2} < w_{A2}, (f_{A1} - w_{A1}) < (f_{A2} - w_{A2}), (f_{B2} - w_{B2}) < (f_{B1} - w_{B1})$

Can be satisfied if: $\quad f_{A1} + f_{B2} < f_{A2} + f_{B1}$

Therefore, at least one of these would be an equilibrium in the limit even for the symmetric CRS case.

There is also a possibility of a mixing equilibrium where either all of the agents play mixed strategies, or only some. For instance, the equilibrium may take the form $\{m_{A1}, 0, n_{A1}, 0\}$ or $\{0, m_{B2}, 0, n_{B2}\}$, where the remaining masses are in the interval $(0, 1)$. The former is a feature of both the BSW model and the symmetric CRS model, and the latter is a feature of the symmetric CRS model.

First, consider the fully mixed equilibrium for the symmetric CRS model. After substituting interior wages and matching rates, strategies have to satisfy the following equations simultaneously:

$$\frac{\sqrt{m_{A1}n_{A1}}}{m_{A1}} * f_{A1} = \frac{\sqrt{(1-m_{A1})(1-n_{B2})}}{(1-m_{A1})} * f_{B1}$$
$$\frac{\sqrt{m_{B2}n_{B2}}}{m_{B2}} * f_{B2} = \frac{\sqrt{(1-m_{B2})(1-n_{A1})}}{(1-m_{B2})} * f_{A2}$$
$$\frac{\sqrt{m_{A1}n_{A1}}}{n_{A1}} * f_{A1} = \frac{\sqrt{(1-m_{B2})(1-n_{A1})}}{(1-n_{A1})} * f_{A2}$$
$$\frac{\sqrt{m_{B2}n_{B2}}}{n_{B2}} * f_{B2} = \frac{\sqrt{(1-m_{A1})(1-n_{B2})}}{(1-n_{B2})} * f_{B1}$$

Combining these quickly leads to the conclusion that this may only work if $f_{A1} * f_{B2} = f_{B1} * f_{A2}$. While we cannot derive specific strategies m and n, one can show that the matching rates coming out of this equilibrium would have to satisfy:

$$M = \begin{bmatrix} \sqrt{\frac{(f_{1B})^2(f_{2A})^2}{\left((f_{2B})^2+(f_{1B})^2\right)\left((f_{2B})^2+(f_{2A})^2\right)}} & \sqrt{\frac{(f_{1A})^2(f_{2B})^2}{\left((f_{1A})^2+(f_{2A})^2\right)\left((f_{2B})^2+(f_{2A})^2\right)}} \\ \sqrt{\frac{(f_{1A})^2(f_{2B})^2}{\left((f_{1A})^2+(f_{1B})^2\right)\left((f_{2B})^2+(f_{1B})^2\right)}} & \sqrt{\frac{(f_{1B})^2(f_{2A})^2}{\left((f_{1A})^2+(f_{1B})^2\right)\left((f_{1A})^2+(f_{2A})^2\right)}} \end{bmatrix}$$

For the knife-edge case $f_{xy} = \begin{bmatrix} 2 & 1 \\ 1 & 0.5 \end{bmatrix}$, the matching rate equals $M = \begin{bmatrix} 0.8 & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$.

This is exactly the mixing equilibrium we get in the zero-cost limit of the sequential targeted search model. However, as our derivation shows, the equations are contradictory and should not generate a mixed equilibrium unless $f_{A1} * f_{B2} = f_{B1} * f_{A2}$. Nevertheless, for other cases, such as the horizontal surplus $f_{xy} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and vertical surplus $f_{xy} = \begin{bmatrix} 2 & 1 \\ 1 & 0.4 \end{bmatrix}$, we get a mixing equilibrium in the zero-cost limit with matching rates quite similar to the patterns prescribed by the formula: $M = \begin{bmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{bmatrix}$ and $M = \begin{bmatrix} 0.86 & 0.33 \\ 0.33 & 0.2 \end{bmatrix}$ respectively.

Now consider the partially mixed case for the symmetric CRS model. Consider the case $\{m_{A1}, 0, n_{A1}, 0\}$, the equilibrium must satisfy:

$mr_{w1A} \cdot w_{A1} = mr_{w1B} \cdot w_{B1}$

$mr_{f1A} \cdot (f_{A1} - w_{A1}) = mr_{f2A} \cdot (f_{A2} - w_{A2})$

$w_{A1} = \alpha f_{A1} \qquad w_{B1} = \alpha f_{B1} \qquad w_{A2} = \alpha f_{A2} \qquad w_{B2} = 0 + \varepsilon$

$mr_{w2A} \cdot w_{A2} > mr_{w2B} \cdot w_{B2}$

$mr_{f1B} \cdot (f_{B1} - w_{B1}) > mr_{f2B} \cdot (f_{B2} - w_{B2})$

Substitute wages back into equations:

$mr_{w1A} \cdot \alpha f_{A1} = mr_{w1B} \cdot \alpha f_{B1}$

$mr_{f1A} \cdot (1 - \alpha) f_{A1} = mr_{f2A} \cdot (1 - \alpha) f_{A2}$

$mr_{w2A} \cdot \alpha f_{A2} > mr_{w2B} \cdot \varepsilon$

$mr_{f1B} \cdot (1 - \alpha) f_{B1} > mr_{f2B} \cdot (f_{B2} - \varepsilon)$

Substitute strategies into matching rates:

$mr_{w1A} = \frac{M(m_{1A}, n_{1A})}{m_{1,A}}, mr_{w1B} = \frac{M(1-m_{1A}, 1)}{1-m_{1,A}}$

$mr_{w2A} = \frac{M(1, 1-n_{A1})}{1}, mr_{w2B} = \frac{M(0,0)}{0}$

$mr_{f1A} = \frac{M(m_{1A}, n_{1A})}{n_{1A}}, mr_{f2A} = \frac{M(1, 1-n_{A1})}{1-n_{A1}}$

$mr_{f1B} = \frac{M(1-m_{1A}, 1)}{1}, mr_{f2B} = \frac{M(0,0)}{0}$

Substitute matching rates as well:

$\frac{M(m_{1A}, n_{1A})}{m_{1,A}} \cdot \alpha f_{A1} = \frac{M(1-m_{1A}, 1)}{1-m_{1,A}} \cdot \alpha f_{B1}$

$\frac{M(m_{1A}, n_{1A})}{n_{1A}} \cdot (1 - \alpha) f_{A1} = \frac{M(1, 1-n_{A1})}{1-n_{A1}} \cdot (1 - \alpha) f_{A2}$

$\frac{M(1, 1-n_{A1})}{1} \cdot \alpha f_{A2} > \frac{M(0,0)}{0} \cdot \varepsilon$

$\frac{M(1-m_{1A}, 1)}{1} \cdot (1 - \alpha) f_{B1} > \frac{M(0,0)}{0} \cdot (f_{B2} - \varepsilon)$

The inequalities are clearly satisfied since $\frac{M(0,0)}{0} = 0$.

We have two equations with respect to m_{1A}, n_{1A}, which would have a unique solution for most specific matching functions.

$\frac{M(m_{1A},n_{1A})}{m_{1,A}} f_{A1} = \frac{M(1-m_{1A},1)}{1-m_{1,A}} f_{B1}$

$\frac{M(m_{1A},n_{1A})}{n_{1A}} f_{A1} = \frac{M(1,1-n_{A1})}{1-n_{A1}} f_{A2}$

Consider the symmetric Cobb-Douglas case $M = \sqrt{m \cdot n}$:

$n_{1A}(1 - m_{1A}) = m_{1A}\left(\frac{f_{B1}}{f_{A1}}\right)^2$

$m_{1A}(1 - n_{1A}) = n_{1A}\left(\frac{f_{A2}}{f_{A1}}\right)^2$

Has the solution: $m_{1A} = \frac{1-\left(\frac{f_{B1}}{f_{A1}}\right)^2\left(\frac{f_{A2}}{f_{A1}}\right)^2}{1+\left(\frac{f_{B1}}{f_{A1}}\right)^2}$ $\qquad n_{1A} = \frac{1-\left(\frac{f_{B1}}{f_{A1}}\right)^2\left(\frac{f_{A2}}{f_{A1}}\right)^2}{1+\left(\frac{f_{A2}}{f_{A1}}\right)^2}$

This replicates our numerical results in Figure 3.3.

$n_{1A} = \left.\frac{1-\left(\frac{f_{B1}}{f_{A1}}\right)^2\left(\frac{f_{A2}}{f_{A1}}\right)^2}{1+\left(\frac{f_{A2}}{f_{A1}}\right)^2}\right|_{f_{A1}=2,f_{B1}=1,f_{A2}=1} = \frac{3}{4}$

$m_{1A} = \left.\frac{1-\left(\frac{f_{B1}}{f_{A1}}\right)^2\left(\frac{f_{A2}}{f_{A1}}\right)^2}{1+\left(\frac{f_{B1}}{f_{A1}}\right)^2}\right|_{f_{A1}=2,f_{B1}=1,f_{A2}=1} = \frac{3}{4}$

$\sqrt{n_{1A}m_{1A}} = \frac{3}{4}$ $\quad \sqrt{(1-n_{2B})(1-m_{1A})} = \sqrt{(1-m_{1A})(1-n_{2B})} = \frac{1}{2}$ $\quad \sqrt{n_{2B}m_{2B}} = 0.$

Thus, for the vertical surplus $f_{xy} = \begin{bmatrix} 2 & 1 \\ 1 & a < 0.5 \end{bmatrix}$, we get a semi-mixed equilibrium which produces the equilibrium matching rate $M = \begin{bmatrix} 0.75 & 0.5 \\ 0.5 & 0 \end{bmatrix}$. As we shall see below, this will be the socially optimal allocation for this surplus pattern.

BSW case: Consider also the BSW case $M(m,n) = m \cdot n$:

$mr_{w1A} = n_{1A}, mr_{w1B} = 1 - n_{B2}$

$mr_{w2A} = 1 - n_{A1}, mr_{w2B} = n_{B2}$

$mr_{f1A} = m_{1A}, mr_{f2A} = 1 - m_{B2}$

$mr_{f1B} = 1 - m_{1A}, mr_{f2B} = m_{B2}$

There are the same two pure strategies equilibria: PAM and NAM.

There is also a fully mixed-strategy equilibrium, that satisfies:

$n_{1A} \cdot f_{A1} = (1 - n_{B2}) \cdot f_{B1}$

$(1 - n_{A1}) \cdot f_{A2} = n_{B2} \cdot f_{B2}$

$m_{1A} \cdot f_{A1} = (1 - m_{B2}) \cdot f_{A2}$

$(1 - m_{1A}) \cdot f_{B1} = m_{B2} \cdot f_{B2}$

These have a closed form solution:
$n_{1A}=\frac{f_{A2}f_{B1}-f_{B2}f_{B1}}{f_{A2}f_{B1}-f_{B2}f_{A1}}, n_{B2}=\frac{f_{A2}f_{B1}-f_{A2}f_{A1}}{f_{A2}f_{B1}-f_{B2}f_{A1}}, m_{1A}=\frac{f_{A2}f_{B1}-f_{A2}f_{B2}}{f_{A2}f_{B1}-f_{B2}f_{A1}}, m_{B2}=\frac{f_{A2}f_{B1}-f_{A1}f_{B1}}{f_{A2}f_{B1}-f_{B2}f_{A1}}$

For the simple limiting case of $f_{A2}=f_{B1}=1, f_{B2}=f_{A1}=1+\varepsilon\rightarrow 1$, it is easy to check that $n_{1A}=m_{1A}=n_{2B}=m_{2B}=\frac{1}{2}$.

We have just fully reconstructed the strategies and wage rates of the BSW model and of the symmetric Cobb-Douglas numerical examples in the paper when the search/information costs equal zero. We find that in the 2x2 case there are always at least two pure strategy equilibria, representing positive and negative assortative matching. There can also be a fully-mixed equilibrium and there can be semi-mixed equilibria (when not all agents play mixed strategies). Our derivation demonstrates how cumbersome it can be to check all the combinations of equalities and inequalities of various conditions in the directed search model even in the simple 2x2 case, and that directed search always gives multiple equilibria. In the sequential targeted search model in the zero-cost limit of the 2x2 symmetric CRS model — we always find 3 equilibria — two pure-strategy (PAM and NAM) and one mixed-strategy equilibrium. This structure is fully covered by the predictions of the directed search model described above, but not all the directed search equilibria survive for small positive costs. It is much easier to consider numerically our targeted search model when approaching the zero-cost limit to find all the relevant equilibria and their properties. This would suggest that our model can serve as a great diagnostic tool to study the equilibria of directed search models, and their properties. As in reality the costs are positive, our model also allows the researcher to discard some of the directed search equilibria, which do not survive for any positive costs, indicating that they are unlikely to ever be observed in reality.

Another important question is that of social efficiency of directed search equilibria. Consider the planner's solution in general, and in each case. The planner simply sums up the expected surplus of matches and decides how to allocate masses of workers and firms:

$$\Omega=\left[\begin{array}{c} M\left(m_{1A},n_{1A}\right)f_{A1}+M\left(1-m_{2B},1-n_{1A}\right)f_{A2} \\ +M\left(1-m_{1A},1-n_{2B}\right)f_{B1}+M\left(m_{2B},n_{2B}\right)f_{B2} \end{array}\right]\rightarrow \max_{m_{A1},m_{B2},n_{A1},n_{B2}}$$

The first-order conditions are simply:

$$\frac{\partial M(m_{1A},n_{1A})}{\partial m_{1A}}f_{A1}=\frac{\partial M(1-m_{1A},1-n_{2B})}{\partial(1-m_{1A})}f_{B1}$$
$$\frac{\partial M(1-m_{2B},1-n_{1A})}{\partial(1-m_{2B})}f_{A2}=\frac{\partial M(m_{2B},n_{2B})}{\partial m_{2B}}f_{B2}$$
$$\frac{\partial M(m_{1A},n_{1A})}{\partial n_{1A}}f_{A1}=\frac{\partial M(1-m_{2B},1-n_{1A})}{\partial(1-n_{1A})}f_{A2}$$

$$\frac{\partial M(1-m_{1A},1-n_{2B})}{\partial(1-n_{2B})}f_{B1}=\frac{\partial M(m_{2B},n_{2B})}{\partial n_{2B}}f_{B2}$$

These are interior conditions for efficiency, and do not hold all at the same time. Only some may hold, or none at all. Pure-strategy corner solutions, for which none of these conditions hold — are also candidates. Different such combinations all need to be considered as candidates for efficiency, and only one would lead to the highest welfare.

We can, however, compare them to the equilibrium conditions that we have — for the cases where an agent needs to be indifferent. For instance, consider the mixed-strategy condition for worker 1.

$mr_{w1A}\cdot w_{A1}=mr_{w1B}\cdot w_{B1}$

We can substitute definitions of matching rates and the posted wages to obtain:

$$\frac{M(m_{A1},n_{A1})}{m_{A1}}\cdot\left(1-\frac{\partial M(m_{A1},n_{A1})}{\partial n_{A1}}\frac{n_{A1}}{M(m_{A1},n_{A1})}\right)f_{A1}=$$
$$=\frac{M(1-m_{1A},1-n_{B2})}{1-m_{1,A}}\cdot\left(1-\frac{\partial M(1-m_{1A},1-n_{B2})}{\partial(1-n_{B2})}\frac{(1-n_{B2})}{M(1-m_{1A},1-n_{B2})}\right)f_{B1}$$

Compare this with the corresponding planner's condition for efficiency:

$$\frac{\partial M(m_{1A},n_{1A})}{\partial m_{1A}}f_{A1}=\frac{\partial M(1-m_{1A},1-n_{2B})}{\partial(1-m_{1A})}f_{B1}$$

For efficiency of equilibria, the matching function must have constant returns to scale and a constant elasticity with respect to both inputs:

$$\frac{\partial M\left(m_{A1},n_{A1}\right)}{\partial n_{A1}}\frac{n_{A1}}{M\left(m_{A1},n_{A1}\right)}+\frac{\partial M\left(m_{1A},n_{1A}\right)}{\partial m_{1A}}\frac{m_{A1}}{M\left(m_{A1},n_{A1}\right)}=1.$$

This implies that matching functions need to be Cobb-Douglas with constant parameters for the conditions for equilibria to become equivalent to conditions for efficiency. Each equilibrium of the directed search model satisfies some conditions for an extremum, but some of them are local maxima and some are local minima. As the conditions are equivalent, but there can be only one social optimum - one of the equilibria of the directed search model must implement the planner's solution, and the others are suboptimal.

To illustrate these results, consider again the numerical cases with symmetric CRS described above. When $f_{A1}*f_{B2}>f_{B1}*f_{A2}$ - the pure-strategy PAM equilibrium is socially efficient, and this allocation is supported in equilibrium by wages arbitrarily close to 0. When $f_{A1}*f_{B2}<f_{B1}*f_{A2}$ - the semi-mixed-strategy equilibrium (e.g. with the number of matches $M=\begin{bmatrix}0.75 & 0.5\\ 0.5 & 0\end{bmatrix}$) is the socially optimal outcome. In this equilibrium, wages in three of four matching locations — those for which the workers

and firms need to be indifferent — are equal to half of the surplus, while in the fourth matching location the wage can be set arbitrarily close to zero. This structure of socially-optimal allocation is preserved for positive costs, but the corresponding semi-mixing equilibrium disappears for positive costs. In this case, for arbitrarily small positive costs, all the competitive equilibria are substantially socially inefficient.

There are two major takeaways from our derivations, compared with conventional wisdom regarding directed search models:

1. Equilibrium conditions for directed search equilibria are the same as for the planner's solution if matching technology is Cobb-Douglas. However, not all of the conditions must be satisfied as equalities. Consequently, there are always multiple equilibria in pure strategies, and there can also be equilibria in mixed and semi-mixed strategies. Only one of these directed search equilibria achieves the social optimum. The semi-mixed equilibria may not survive for positive costs, leading to cases of social inefficiency of all of the surviving competitive equilibria for arbitrarily small positive costs of search.

2. For agents that are playing mixed strategies in equilibrium, wages are set by splitting the surplus in proportions dictated by the elasticity of the matching function. For pure strategy equilibria, the wage is undetermined - it can take any value between 0 and f_{xy}. The firm with a first-mover advantage will set the wage arbitrarily close to zero (outside option of the worker).

In the symmetric CRS case that we considered in the main text, the condition under which the semi-mixing equilibrium is socially optimal in the 2x2 case is: $f_{B2}f_{A1} < f_{A2}f_{B1}$. This case leads to inefficiency of all competitive equilibria in the zero-cost limit. Otherwise, a pure strategies equilibrium (PAM or NAM) is socially optimal. In the mixing equilibrium, wages pay a nontrivial fraction of the surplus (half in the symmetric case). Pure strategy equilibria produce assortativeness and lead to posted wages essentially equal to zero.