Search, Screening and Sorting*

Xiaoming Cai† Pieter Gautier‡ Ronald Wolthoff§

September 18, 2020

Preliminary and Incomplete

Click here for the latest version.

Abstract

We investigate the effect of search frictions on labor market sorting by constructing a model which is in line with recent evidence that employers collect a pool of applicants before interviewing a subset of them. In this environment, we derive the necessary and sufficient conditions for positive and negative assortative matching, which depend on the degree of complementarity in production and the extent to which firms can interview applicants. Challenging the conventional wisdom that search frictions are necessarily a force against sorting, we find that the required degree of complementarity for positive assortative matching is *increasing* in the number of interviews: it ranges from root-supermodularity if each firm can interview a single applicant to log-supermodularity if each firm can interview all its applicants. We show that our results are robust to a large number of alternative specifications of the matching process.

JEL codes: C78, D82, D83, E24.

Keywords: sorting, complementarity, search frictions, information frictions, competing mechanisms, heterogeneity.

*We are grateful to Jim Albrecht and Axel Anderson for insightful discussions of this paper. We further thank Steve Davis, Philipp Kircher, Robert Shimer, and various seminar and conference participants for valuable comments. Wolthoff gratefully acknowledges financial support from the Social Sciences and Humanities Research Council.

†Peking University HSBC Business School. E-mail: xmingcai@gmail.com.

‡VU University Amsterdam and Tinbergen Institute. E-mail: p.a.gautier@vu.nl.

§University of Toronto. E-mail: ronald.p.wolthoff@gmail.com.

1 Introduction

One of the most important tasks for any firm is to hire the right workers. A crucial part of this process consists of screening applicants through job interviews.[1] In this paper, we are interested in the question how this screening process affects sorting patterns in the labor market. Does the extent to which firms can interview workers affect the conditions under which the labor market exhibits positive (PAM) or negative assortative matching (NAM)? If technological innovations allow firms to screen more applicants with higher precision, does that make sorting more or less likely?[2]

Unfortunately, the economic literature is silent on these questions as the work on sorting has generally abstracted from modeling the screening of applicants. The earliest work on assignment problems (Tinbergen, 1956; Shapley and Shubik, 1971; Becker, 1973; Rosen, 1974) considers frictionless environments in which there is full information about types. More recent work by Shimer and Smith (2000), Shimer (2005) and Eeckhout and Kircher (2010) allows for frictions but makes particular assumptions about the matching process and does not explore how the results depend on these assumptions.[3]

In order to answer our question, we present a new search model of the labor market. In line with recent evidence by Davis and Samaniego de la Parra (2017), we allow firms to meet and interview multiple workers before making a job offer to the most profitable candidate. We show how the extent to which firms can interview workers and the degree of complementarities in production jointly affect the allocation of workers to jobs. Perhaps surprisingly, we find that reducing search frictions by allowing firms to interview more workers is a force *against* sorting: the *easier* it is for firms to rank applicants, the *stronger* the complementarities in production that are required to obtain positive assortative matching (PAM).

Although our focus is on the labor market, our results are important for all markets where two sides of the market must form a match, where heterogeneity matters

[1] See below for some empirical evidence regarding the recruiting process. Note that 'screening' in this context has a different meaning than the homonymous game-theoretic concept. In addition to job interviews, screening workers may involve other instruments like checking references, assessments, and job tests. We use 'interview' as shorthand for the entire collection of instruments.

[2] As an example of such a technological innovation, Hoffman et al. (2018) describe how some firms have started to subject all applicants to an online job test. Based on their answers, every applicant is assigned a score by a data firm, calculated from correlations between answers and job performance among existing employees.

[3] Although Eeckhout and Kircher (2010) use buyer/seller terminology, the same idea applies.

and where one side of the market can screen a subset of agents that contacted them, i.e the housing, labor and marriage market. Also in trade, there is a growing interest in deriving patterns of international specialization (i.e. under which conditions do exporters hire the most productive workers) from fundamental properties of the production technology, see Costinot (2009).

To illustrate the importance of simultaneous interviews, we first briefly discuss the current state of the literature and then explain how allowing for simultaneous interviews alters the conventional wisdom. Becker (1973) showed that in a frictionless economy, supermodularity of the production function (or complementarities beween workers and jobs) is a sufficient condition for PAM. Then, Shimer and Smith (2000) showed that when the matching process is governed by random search frictions, we need a set of conditions which are even stronger than log-supermodularity for PAM to arise. The reason for this is that the opportunity cost of remaining unmatched is higher for the high types and this makes them more eager to match with a low type rather than running the risk to not match at all. To undo this effect, the production function must exhibit stronger complementarities. Eeckhout and Kircher (2010) (EK) show that when search is directed rather than random, we need something weaker than log-supermodularity (root-supermodularity) for PAM.[4] This is because directed search allows high types to avoid meeting low-type trading partners.

Both Shimer and Smith (2000) and Eeckhout and Kircher (2010) are very general in terms of the production technology but only allow firms to meet one worker at a time. We are equally general in terms of the production technology but we also allow firms to conduct simultaneous interviews and show that this creates an additional force that goes against PAM. This makes the analysis more difficult because firms can now invest in a large and or a high quality pool of applicants and, as we will show, both strategies can generate similar profits. Irrespective of this multiplicity, in our baseline description of the recruiting process, we prove that the necessary and sufficient degree of supermodularity is linearly increasing in the expected interview capacity, ranging from square-root-supermodularity when firms can screen only a single worker to log-supermodularity when firms can interview all their applicants.

To understand those results, start from a candidate equilibrium that has PAM and suppose a high-skilled worker considers deviating by applying to a low-type firm

[4] Shi (2001) was the first to show that under directed search supermodularity is not enough for PAM.

in order to increase his hiring probability. The smaller the screening capacity of firms, the more random the hiring process at those firms becomes and the less attractive this deviation is. Therefore, if firms do not screen much, ~~then~~ it is easier to sustain PAM and weaker complementarities in production suffice.

Things are a bit more complicated than this because one could argue that if firms screen less, low-skilled workers have more incentives to apply to high-type firms. However, high-type firms have a tool to discourage low-skilled workers from applying. They could simply offer them lower wages. This is profitable for them because reducing congestion externalities from low-skilled workers increases the probability for high-skilled workers to receive an offer so it is a cheap way for firms to offer them their market utility. To the contrary, low-type firms have no incentives to discourage high-skilled workers from applying. So for a given production technology, *decreasing* the interview capacity will make it less attractive for high-skilled workers to apply to low-type firms while at the same time, high-type firms will discourage low-type workers from applying there and this makes PAM a more likely outcome. It is also more difficult to sustain NAM when firms can also screen workers ex post because then high-type workers have an incentive to spread and avoid each other's company in the queues of applicants.

Our results are also important for the growing empirical literature that aims to identify the shape of the production function from observed matching patterns.[5] The papers in this literature have all taken the meeting technology as given. Our results imply that without information on the meeting and screening process observed matches alone cannot identify the degree of complementarity in the production technology. To facilitate the use of recruiting data in a sorting analysis, we further derive conditions for sorting in the distribution of applicants and interviews (positive assortative contacting, PAC).

Some papers have argued that increased sorting of high worker types at high wage firms has contributed to the observed increased inequality from the mid-nineties onwards, see for example Card et al. (2013) and Song et al. (2018).[6] Håkanson

[5]See the literature that started with Abowd et al. (1999). Gautier and Teulings (2006) and subsequent papers like Eeckhout and Kircher (2011), Gautier and Teulings (2015) Lise et al. (2016), Hagedorn et al. (2017), Lopes de Melo (2018), Bartolucci et al. (2018), and Bagger and Lentz (2018) made the point that wages for a given worker type are non-monotonic in firm types so the AKM-methodology of detecting sorting patterns from correlating worker and firm fixed effects fails.

[6]Card et al. (2013) use education and occupational sorting.

et al. (2018) argue that the increased sorting patterns are mainly due to increasing complementaritries in production. Our results suggest that if during the same period, new technologies like automated resume screening made it cheaper to screen workers, then this would require even stronger complementarities in the production technology.

The paper is organized as follows. Section 2 introduces the model where for ease of exposition we will assume a two-point worker distribution. Section 3 considers the market equilibrium, and shows that the optimal mechanism can be implemented by wage menus. Section 4 derives our main results regarding sorting. In Section 5, we consider various extensions, including noisy signals for every applicant and endogenous choice of screening capacity. One may expect that the most productive firms have the strongest incentives to screen workers ex post but we show that that intuition is wrong. Firms in the middle have the strongest incentives to collect a mixed pool of applicants and then screen ex post. Finally, Section 6 concludes.

2 Model

2.1 Environment

Agents. A static economy is populated by a continuum of risk-neutral firms and workers. Each firm demands and each worker supplies a single unit of indivisible labor. Both types of agents are heterogeneous. In particular, each firm is characterized by a type $y \in \mathcal{Y} = [\underline{y}, \overline{y}] \subset \mathbb{R}_+$. The measure of firms with types less or equal to y is denoted by $J(y)$, where the total measure $J(\overline{y})$ is normalized to one. Similarly, each worker is characterized by a type $x \in \mathcal{X} = [\underline{x}, \overline{x}] \subset \mathbb{R}_+$, which initially is private information, but can be learned by a firm during an interview, as we describe in more detail below. There are two types of workers: a low type x_1 and a high type x_2, with $0 < x_1 < x_2$.[7] The measure of workers with type x_i is denoted by $\ell_i > 0$. The resource set or the *endowment of agents* in the economy is thus given by $(x_1, x_2, \ell_1, \ell_2, J(y))$.

Wage Menus and Search. Each firm posts and commits to a wage menu $\boldsymbol{w} = (w_1, w_2)$, where w_i is the wage for a hire of type x_i.Workers observe all posted wage menus and apply to one firm, taking into account that at high-wage firms there will

[7]In Appendix C.2 we show for a widely used class of meeting technologies, that includes the urn-ball and geometric, how our results can be generalized to N worker types.

be more competition.[8] We initially assume that workers also observe firm types, but then show that this assumption is redundant. We capture the anonymity of the large market with the standard assumption that identical workers must use symmetric strategies (see e.g. Shimer, 2005).

A *submarket* $(\boldsymbol{w}, y)$ consists of the firms of type y that post wage menu $\boldsymbol{w}$ and all workers who apply to those firms with positive probability. For each submarket, we denote the ratio of the number of high-type applicants to the number of firms by $\mu(\boldsymbol{w}, y)$, and the ratio of the total number of applicants (regardless of their type) to the number of firms by $\lambda(\boldsymbol{w}, y)$. Naturally, these ratios—or *queue lengths*—satisfy $0 \leq \mu(\boldsymbol{w}, y) \leq \lambda(\boldsymbol{w}, y)$ for all $(\boldsymbol{w}, y)$.[9] For future reference, define $\zeta(\boldsymbol{w}, y) = \mu(\boldsymbol{w}, y)/\lambda(\boldsymbol{w}, y)$ as the fraction of high-type applicants in submarket $(\boldsymbol{w}, y)$.

Benchmark Frictions. The matching process within a submarket is frictional and exhibits constant returns to scale, in the sense that outcomes only depend on queue lengths rather the absolute measures of workers and firms. Within those boundaries, we can allow for a fairly wide class of matching processes, but we will initially focus on a specific benchmark with two stages (applying and screening) to simplify exposition.

To introduce this benchmark, consider a particular submarket with queues (μ, λ), where λ is the total queue length and μ is the queue length of x_2 workers only. Workers and firms in the submarket are randomly located on the circumference of a circle according to a uniform distribution. Workers apply clockwise to the nearest firm.[10] A firm therefore receives n applications with probability $\frac{1}{1+\lambda}(\frac{\lambda}{1+\lambda})^n$ for $n = 0, 1, 2, \ldots,$ which is a geometric distribution with mean λ. In the screening stage, firms interview their applicants in a random order. An interview reveals the type of the applicant. After every interview, and conditional on applicants remaining, there is an exogenous probability $\sigma \in [0, 1]$ that the firm can conduct another interview, while interviewing stops with complementary probability. In other words, σ is a measure of how easy it is for firms to learn applicants' types: if $\sigma = 0$, each firm can interview only a single applicant, which is a special case of the bilateral model of Eeckhout and Kircher

[8] The assumption that workers have a single chance to match (per period) is standard and captures the idea that (opportunity) costs are associated with applying to jobs. The limited work relaxing this assumption has generally focused on environments with (ex ante) homogeneous agents (see e.g. Albrecht et al., 2006; Galenianos and Kircher, 2009; Kircher, 2009; Wolthoff, 2018; Albrecht et al., 2019). An exception is Auster et al. (2020).

[9] We provide a formal derivation of these queue lengths below.

[10] As long as workers cannot keep track of the distance they have traveled, this application strategy is merely a tie-breaking rule.

(2010), while a firm can interview all its applicants if $\sigma = 1$.

It is worth reiterating that our analysis is not restricted to this particular microfoundation. In section 5, we consider various generalizations, including one in which firms can select interviewees based on noisy signals of worker types (e.g., resumes).

Matching and Production. After the interviews have been conducted, matches are formed. Firms can only hire a worker which they have interviewed.[11] If a firm has interviewed multiple applicants, it hires the most profitable one. A match between a worker of type x and a firm type of y produces output $f(x, y)$, which is strictly positive, strictly increasing, and twice continuously differentiable for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. From this output, the firm pays the worker the promised wage w_i and keeps the rest. Firms and workers which fail to match obtain a zero payoff.

Elasticity of Complementarity. For our analysis, a key characteristic of the production function is its *elasticity of complementarity* (Hicks, 1932, 1970), which is the inverse of the elasticity of substitution and can be written as

$$\rho(x, y) \equiv \frac{f_{xy}(x, y) f(x, y)}{f_x(x, y) f_y(x, y)} \in \mathbb{R}, \tag{1}$$

with extrema $\overline{\rho} = \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x, y)$ and $\underline{\rho} = \inf_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x, y)$. For future reference, note that $\rho(x, y)$ measures the elasticity of relative marginal product with respect to relative product. That is, for sufficiently small $\Delta x > 0$, we have

$$\frac{f_y(x + \Delta x, y)}{f_y(x, y)} \approx 1 + \rho(x, y) \frac{f_x(x, y)}{f(x, y)} \Delta x \approx \left(\frac{f(x + \Delta x, y)}{f(x, y)} \right)^{\rho(x,y)}.$$

In general, when x is discrete and $\rho(x, y)$ is not necessarily constant, the elasticity of relative marginal product with respect to relative product is bounded by $\underline{\rho}$ and $\overline{\rho}$, as summarized by the following lemma.

Lemma 1. *For given y, $f_y(x, y)/f(x, y)^{\underline{\rho}}$ is increasing in x, and $f_y(x, y)/f(x, y)^{\overline{\rho}}$ is decreasing in x. That is,*

$$\left(\frac{f(x_2, y)}{f(x_1, y)} \right)^{\underline{\rho}} \leq \frac{f_y(x_2, y)}{f_y(x_1, y)} \leq \left(\frac{f(x_2, y)}{f(x_1, y)} \right)^{\overline{\rho}}, \tag{2}$$

[11] This assumption can easily be rationalized by introducing a small chance that any given worker provides the firm with a sufficiently negative payoff when hired.

where the first (resp. second) inequality holds as equality if and only if $\underline{\rho}$ *(resp.* $\overline{\rho}$*) is equal to* $\rho(x, y)$ *for all* $x \in [x_1, x_2]$.

Proof. See Appendix A.1. □

Supermodularity. The elasticity of complementarity $\rho(x, y)$ is closely related to the notion of n-root-supermodularity, as defined in Eeckhout and Kircher (2010).[12]

Definition 1. *The function* $f(x, y)$ *is* n-root-supermodular *on* $\mathcal{X} \times \mathcal{Y}$ *if and only if*

$$\rho(x, y) \geq 1 - \frac{1}{n}, \tag{3}$$

for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$*; special cases include supermodularity* $(n = 1)$ *and log-supermodularity* $(n \to \infty)$*. When the inequality in* (3) *is reversed,* $f(x, y)$ *is said to be* n-root-submodular*.*

In other words, n-root-supermodularity is equivalent to $\underline{\rho} \geq 1 - 1/n$ and n-root-submodularity is equivalent to $\overline{\rho} \leq 1 - 1/n$.[13]

Special Case. We will sometimes illustrate our results with a CES production function, because it has a constant elasticity of complementarity, $\rho(x, y) = \rho$. That is, $f(x, y) = (x^{1-\rho} + y^{1-\rho})^{\frac{1}{1-\rho}}$. This production function is submodular when $\rho \leq 0$, $\frac{1}{1-\rho}$-root-supermodular when $0 < \rho < 1$, and log-supermodular when $\rho \geq 1$.

Queue Lengths and Expected Payoffs. Consider a submarket $(\boldsymbol{w}, y)$ with queues (μ, λ). For this submarket, define $\pi(\boldsymbol{w}, \mu, \lambda, y)$ and $V_i(\boldsymbol{w}, \mu, \lambda, y)$ as the expected payoff of firms and workers of type x_i, respectively. The explicit expressions for these payoffs will be given later by equations (9) and (10).

A firm of type y has to form beliefs about the queues $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ when posting a wage menu $\boldsymbol{w}$. We follow the standard approach in the literature, which imposes restrictions on these beliefs in the spirit of subgame perfection through what is known as the *market utility condition* (see e.g. McAfee, 1993; Shimer, 2005; Eeckhout and Kircher, 2010). Define the *market utility* U_i of a worker of type x_i as the maximum

[12] Eeckhout and Kircher (2010) define $f(x, y)$ to be n-root-supermodular if $\sqrt[n]{f(x, y)}$ is supermodular. Since $\frac{1}{\partial x \partial y} \sqrt[n]{f} = n^{-2} f^{1/n-2} \left(f f_{xy} - (1 - \frac{1}{n}) f_x f_y\right)$, (3) is equivalent to their definition.

[13] The equivalence between $\underline{\rho} = 1$ and log-supermodularity of $f(x, y)$ also follows from Lemma 1: when $\underline{\rho} = 1$, the first inequality in (2) can be rewritten as $\frac{\partial}{\partial y} \log f(x_2, y) \geq \frac{\partial}{\partial y} \log f(x_1, y)$.

expected payoff that this worker can obtain in equilibrium, either by visiting one of the submarkets or by remaining inactive. Then $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ must satisfy

$$\begin{cases} V_1(\boldsymbol{w}, \mu, \lambda, y) \leq U_1, \text{ with equality if } \lambda - \mu > 0, \\ V_2(\boldsymbol{w}, \mu, \lambda, y) \leq U_2, \text{ with equality if } \mu > 0. \end{cases} \tag{4}$$

For common contact technologies, including our benchmark, (4) admits a unique solution (μ, λ), which is then the firm's belief. For other technologies, there can be multiple solutions to (4). The standard assumption in the literature is then that firms are optimistic and expect the solution that maximizes their expected payoff $\pi(\boldsymbol{w}, \mu, \lambda, y)$. We follow that approach when necessary.

Let $G(\boldsymbol{w} \mid y)$ denote the (conditional) probability that a firm of type y offers a wage menu $\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}$, where $\widetilde{\boldsymbol{w}} = (\widetilde{w}_1, \widetilde{w}_2)$, $\boldsymbol{w} = (w_1, w_2)$, $\widetilde{w}_1 \leq w_1$ and $\widetilde{w}_2 \leq w_2$. Given market utility (U_1, U_2), firm optimality means $G(\boldsymbol{w} \mid y)$ must maximize $\pi(\boldsymbol{w}, \mu, \lambda, y)$ subject to the constraint (4).

Similarly, let $H_i(\boldsymbol{w}, y)$ denote the probability that workers of type x_i apply to a firm with $\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}$ and $\widetilde{y} \leq y$. Worker optimality requires that workers must obtain exactly the market utility at any firm to which they apply with positive probability, while their expected payoff must be weakly less than their market utility at firms to which they don't apply, i.e., (4) must hold. The application strategies of low-type and high-type workers are therefore given by

$$H_2(\boldsymbol{w}, y) = \frac{1}{\ell_2} \int_{\widetilde{y} \leq y} \int_{\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}} \mu(\widetilde{\boldsymbol{w}}, \widetilde{y}) \, dG(\widetilde{\boldsymbol{w}} \mid \widetilde{y}) \, dJ(\widetilde{y}) \tag{5}$$

$$H_1(\boldsymbol{w}, y) = \frac{1}{\ell_1} \int_{\widetilde{y} \leq y} \int_{\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}} [\lambda(\widetilde{\boldsymbol{w}}, \widetilde{y}) - \mu(\widetilde{\boldsymbol{w}}, \widetilde{y})] \, dG(\widetilde{\boldsymbol{w}} \mid \widetilde{y}) \, dJ(\widetilde{y}). \tag{6}$$

Note that no firm will post a wage menu $\boldsymbol{w} \geq \overline{\boldsymbol{w}} \equiv (f(x_1, \overline{y}), f(x_2, \overline{y}))$. Thus, $H_i(\overline{\boldsymbol{w}}, \overline{y})$ is the probability that workers of type x_i apply. Worker optimality requires that this probability equals 1 if $U_i > 0$, as the payoff from not sending an application is zero.

Equilibrium Definition. We can now define an equilibrium as follows.

Definition 2. *A directed search equilibrium is a triple $(G, \{H_i\}, \{U_i\})$ with the following properties:*

(i) Firm Optimality. *Given (U_1, U_2), every wage menu $\boldsymbol{w}$ in the support of $G(\cdot \mid y)$*

maximizes $\pi(\boldsymbol{w}, \mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y), y)$ *for each firm type* y*, where the queue lengths* $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ *are determined by equation* (4).

(ii) Worker Optimality. *Given* (U_1, U_2)*, the application strategy of high-type and low-type workers satisfies equation* (5) *and* (6)*, respectively. Further,* $H_i(\overline{\boldsymbol{w}}, \overline{y}) = 1$ *if* $U_i > 0$*, which determines the market utilities* (U_1, U_2).

3 Market Equilibrium

We start our analysis of the market equilibrium by establishing an equivalence: the problem of a firm in our environment is equivalent to the problem of a firm that can buy queues of workers directly in a competitive market. The difference with a "conventional" competitive market is that the firm buys a distribution of worker types rather than directly hiring a particular worker type. The immediate consequence of this result is that the market equilibrium in our environment is constrained efficient.

3.1 Efficiency

Interviewing Probability. Throughout our analysis, we will rely on the insight of Cai et al. (2018) that the meeting process within a submarket can be summarized by a simple function $\phi(\mu, \lambda)$.[14] In our context, this function represents the probability that a firm in the submarket interviews *at least* one high-type worker. The following lemma derives this probability.[15]

Lemma 2. *Consider a firm with a queue* μ *of high-type workers and a queue* $\lambda - \mu$ *of low-type workers. The probability that the firm interviews at least one high-type worker equals*

$$\phi(\mu, \lambda) = \frac{\mu}{1 + \sigma\mu + (1-\sigma)\lambda}. \tag{7}$$

Proof. See Appendix A.2. □

[14] The fact that $\phi(\mu, \lambda)$ has only two arguments is not because we have two types of workers; with more than two types, μ simply has to be redefined as the queue of workers whose type is above a certain threshold. See Cai et al. (2018) for details.

[15] Equation (7) also appears in Cai et al. (2018), who study the relation between market segmentation and meeting technologies in a world with homogeneous firms. Our focus is quite different, so we provide a derivation of $\phi(\mu, \lambda)$ for completeness.

As shown by Cai et al. (2018), the function $\phi(\mu, \lambda)$ is useful for multiple reasons. First, it provides a convenient way of calculating matching probabilities. As shown below by Lemma 3, a firm hires a high-type worker as long as it interviews at least one. Hence, $\phi(\mu, \lambda)$ describes the probability that the firm will produce $f(x_2, y)$. Similarly, evaluation of (7) in $\mu = \lambda$ gives the firm's overall matching probability (regardless of the hire's type), which we denote by $m(\lambda) \equiv \phi(\lambda, \lambda)$.

Second, the partial derivatives of $\phi(\mu, \lambda)$ have economically meaningful interpretations. The partial derivative $\phi_\lambda(\mu, \lambda) \leq 0$ captures externalities in the recruiting process as it describes how a firm's chances to hire a high-type worker change if the queue of low-type workers gets longer ($\phi_\lambda(\mu, \lambda)\Delta\lambda = \phi(\mu, \lambda + \Delta\lambda) - \phi(\mu, \lambda)$). Intuitively, the presence of low-type applicants does not affect the chance of hiring a high type if and only if a firm can interview all its applicants (i.e. $\sigma = 1$).[16] In contrast, $\phi_\mu(\mu, \lambda)$ describes how a firm's probability of hiring a high-type worker changes if the queue of such workers increases, while the total queue remains constant (i.e. changing the composition of the applicant pool). From the perspective of a high-type applicant, this partial derivative represents the probability that the worker is hired and increases surplus because there was no other high-type worker who was interviewed.[17]

The expression in equation (7) has the following properties:

- A0. $\phi(\mu, \lambda)$ is strictly increasing and concave in μ, i.e. replacing low-type workers with high-type workers in a submarket increases a firm's probability of interviewing at least one high-type worker, but at a decreasing rate;
- A1. for any given $\zeta \in (0, 1]$, $\phi(\lambda\zeta, \lambda)$ is strictly increasing and strictly concave in λ, i.e. holding the fraction of high-type workers constant, adding more workers to the submarket increases a firm's probability of interviewing at least one high type, but at a decreasing rate;
- A2. for any given $\zeta \in (0, 1]$, $\phi_\mu(\lambda\zeta, \lambda)$ is strictly decreasing in λ, i.e. holding the fraction of high-type workers constant, adding more workers to the submarket reduces the probability that a high-type worker creates surplus.

Surplus. To derive expected surplus, consider a firm of type y which faces a queue (μ, λ).s With probability $m(\lambda) \equiv \phi(\lambda, \lambda)$, the firm receives at least one applica-

[16] This property is known as *invariance*, see Lester et al. (2015), Cai et al. (2018) and section 5.3.

[17] To see this, note that $\phi_\mu(\mu, \lambda)\Delta\mu = \phi(\mu + \Delta\mu, \lambda) - \phi(\mu, \lambda)$, where the right-hand side is the probability that additional surplus is generated when we replace $\Delta\mu$ low-type workers with high-types. Naturally, additional surplus is generated if and only if these $\Delta\mu$ workers are the only high types that are interviewed.

tion, hence generating at least a surplus $f(x_1, y)$; with probability $\phi(\mu, \lambda)$, the firm interviews at least one high-type worker, hence generating an additional surplus $f(x_2, y) - f(x_1, y)$. The expected surplus is thus

$$S(\mu, \lambda, y) = m(\lambda) f(x_1, y) + \phi(\mu, \lambda)[f(x_2, y) - f(x_1, y)]. \tag{8}$$

Payoffs. To simplify exposition, suppose that a firm of type y posts a wage menu $\boldsymbol{w}$ satisfying $f(x_2, y) - w_2 > f(x_1, y) - w_1$, meaning that more productive workers are more profitable. Later, in Lemma 3, we will show that this restriction is without loss of generality. Given queues (μ, λ), the firm then has an expected payoff equal to

$$\pi(\boldsymbol{w}, \mu, \lambda, y) = \phi(\mu, \lambda)[f(x_2, y) - w_2] + [m(\lambda) - \phi(\mu, \lambda)][f(x_1, y) - w_1]. \tag{9}$$

Intuitively, the firm hires a high-type worker if it interviews at least one such worker, which happens with probability $\phi(\mu, \lambda)$. Similarly, the firm hires a low-type worker if it interviews no high-type workers but at least one low-type worker, which happens with probability $m(\lambda) - \phi(\mu, \lambda)$.

A worker of type x_i matches with the firm with probability $\psi_i(\mu, \lambda)$, which equals

$$\psi_1(\mu, \lambda) = \frac{m(\lambda) - \phi(\mu, \lambda)}{\lambda - \mu} \qquad \text{or} \qquad \psi_2(\mu, \lambda) = \frac{\phi(\mu, \lambda)}{\mu}$$

by a simple accounting identity.[18] Given these matching probabilities, a worker's expected payoff from applying to the firm equals

$$V_i(\boldsymbol{w}, \mu, \lambda, y) = \psi_i(\mu, \lambda) w_i. \tag{10}$$

This expression reveals that worker values do not depend on firm types directly in submarkets where high-type workers are preferred $(f(x_2, y) - w_2 > f(x_1, y) - w_1)$. We will show below that in equilibrium all submarkets have this property, such that workers do not actually need to observe firms types to choose where to apply.

Competitive Market for Queues. As is standard in the literature, we can use the market utility condition (4) to substitute the wages w_1 and w_2 out of (9) and

[18]Note that these expressions are also valid for $\mu = 0$ or $\mu = \lambda$, in which case we take the corresponding limit.

rewrite the firm's problem with the queue lengths as choice variables. This yields

$$\max_{0\leq\mu\leq\lambda} \Pi\left(\frac{\mu}{\lambda},\lambda,y\right) \equiv S(\mu,\lambda,y) - \lambda U_1 - \mu\left(U_2 - U_1\right), \tag{11}$$

where, for use in Section 3.2, the arguments of the firm profit function Π are the fraction of high-type applicants μ/λ and the queue length λ.[19] Equation (11) has a straightforward interpretation: it is the payoff of a firm buying queues of low-type and high-type workers in a competitive market at prices equal to their respective market utilities. This formulation will be the starting point for our sorting analysis below.

Suppose that (μ^*,λ^*) solves (11). By equation (10), firms can then attract these queues by posting the wage menu

$$(w_1^*, w_2^*) = \left(U_1/\psi_1(\mu^*,\lambda^*),\ U_2/\psi_2(\mu^*,\lambda^*)\right). \tag{12}$$

Hence, the problem of a firm posting a wage menu $\boldsymbol{w}$ and the competitive problem (11) are equivalent.

Productivity versus Profitability. We have only considered wage menus where firms strictly prefer hiring high-type workers, i.e. $f(x_2,y) - w_2 > f(x_1,y) - w_1$. To see that this restriction is justified, suppose that a firm posts a wage menu where low-type workers yield a higher profit ex post, i.e. $f(x_2,y) - w_2 \leq f(x_1,y) - w_1$, and attracts a queue (μ,λ). Workers must again obtain their market utility, so that the expected transfer from the firm to the workers must be $\mu U_2 + (\lambda-\mu)U_1$; however, the expected surplus must be strictly smaller than $S(\mu,\lambda,y)$ in equation (8). Thus, the firm's expected profit must be strictly smaller than the maximum profit in (11).

The following lemma formalizes this idea by showing that any interior solution to (11) can be implemented with a wage menu in which high-type workers are more profitable ex post. We omit the case in which the firm attracts either only low-type or only high-type workers; these corner solutions are trivial, since the firm can always discourage a certain type of workers from applying by offering them a zero wage.[20]

Lemma 3. *Suppose that (μ^*,λ^*) is an interior solution to firms' problem (11) with*

[19] We have implicitly assumed that $0<\mu<\lambda$ such that both market utility conditions hold with equality. However, it is easy to see that (11) also holds if $\mu = 0$ or $\mu = \lambda$.

[20] A similar result appears in Shimer (2005) for the case of urn-ball meetings. Our proof of Lemma 3 generalizes his result to arbitrary meeting technologies.

$0 < \mu^* < \lambda^*$, and $\boldsymbol{w}^*$ is the associated wage menu given in equation (12), then $f(x_2, y) - w_2^* > f(x_1, y) - w_1^*$.

Proof. See Appendix A.3. □

As mentioned above, this lemma and eq. (10) jointly imply that all our results continue to hold if workers do not observe y.

Uniqueness of Queues. For a given wage menu $\boldsymbol{w}$, the associated queues (μ, λ) are determined by the market utility condition (4). The following lemma shows that there exists a unique solution to this system of non-linear equations.

Lemma 4. *For any wage menu $\boldsymbol{w}$, there exists exactly one solution (μ, λ) to the market utility condition.*

Proof. See Appendix A.4. □

Efficiency. In sum, we have demonstrated that the directed search equilibrium with wage menus $\boldsymbol{w}$ coincides with the equilibrium in a competitive market where firms can buy queues directly at prices equal to workers' market utility. Hence, by the first welfare theorem, we obtain the following efficiency result.

Proposition 1. *The directed search equilibrium is constrained efficient.*

Marginal Contributions to Surplus. Since the decentralized equilibrium is constrained efficient, the expected payoffs of firms and workers equal their marginal contribution to surplus. Note that adding more low-type workers to a submarket only increases λ while adding more high-type workers increases both μ and λ. Thus, the marginal contribution of low-type and high-type workers at a firm of type y with queues (μ, λ) are $S_\lambda(\mu, \lambda, y)$ and $S_\mu(\mu, \lambda, y) + S_\lambda(\mu, \lambda, y)$, respectively. If we increase the number of firms by a factor $1 + \Delta s$, the additional surplus is $(1+\Delta s)S(\mu/(1+\Delta s), \lambda/(1+\Delta s), y) - S(\mu, \lambda, y)$, which, by taking $\Delta s \to 0$, implies that the marginal contribution to surplus of firms is $S(\mu, \lambda, y) - \mu S_\mu(\mu, \lambda, y) - \lambda S_\lambda(\mu, \lambda, y)$.

Using $S(\mu, \lambda, y)$ from equation (8), $f^1 \equiv f(x_1, y)$ and $\Delta f = f(x_2, y) - f(x_1, y)$, we get

$$R(\mu,\lambda,y) = (m(\lambda) - \lambda m'(\lambda))\, f^1 + (\phi(\mu,\lambda) - \mu\phi_\mu(\mu,\lambda) - \lambda\phi_\lambda(\mu,\lambda))\, \Delta f \tag{13}$$

$$T_1(\mu,\lambda,y) = m'(\lambda) f^1 + \phi_\lambda(\mu,\lambda)\Delta f \tag{14}$$

$$T_2(\mu,\lambda,y) = m'(\lambda) f^1 + (\phi_\mu(\mu,\lambda) + \phi_\lambda(\mu,\lambda))\, \Delta f, \tag{15}$$

where R, T_1, and T_2 are the marginal contribution to surplus of firms, low-type workers, and high-type workers, respectively.

Concavity of Surplus. The surplus function $S(\mu,\lambda,y)$ is not necessarily strictly concave at a point (μ,λ). To see this, consider its Hessian $\mathcal{H}(\mu,\lambda,y)$, which equals

$$\mathcal{H}(\mu,\lambda,y) = \begin{pmatrix} \phi_{\mu\mu}\Delta f & \phi_{\mu\lambda}\Delta f \\ \phi_{\mu\lambda}\Delta f & m'' f^1 + \phi_{\lambda\lambda}\Delta f \end{pmatrix},$$

where we omit the arguments of the derivatives of $\phi(\mu,\lambda)$ and $m(\lambda)$ for simplicity.

In the bilateral case $\sigma = 0$, we have $\phi(\mu,\lambda) = m(\lambda)\mu/\lambda$ such that $\phi_{\mu\mu} = 0$, which means that the Hessian is never negative definite and surplus is never strictly concave. Below, we will therefore focus on cases in which $\phi_{\mu\mu} < 0$, i.e. $\sigma > 0$; the results will extend to the bilateral case by continuity. Given $\phi_{\mu\mu} < 0$, the Hessian is negative definite if and only if its determinant is positive. This yields the following result.

Lemma 5. *Surplus $S(\mu,\lambda,y)$ is concave at a point (μ,λ) with $0 < \mu < \lambda$ if (i) $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2 \geq 0$ or (ii) $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2 < 0$ and*

$$\kappa(y) \equiv \frac{f(x_2,y) - f(x_1,y)}{f(x_1,y)} < \frac{-m''}{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}}. \tag{16}$$

Proof. See Appendix A.5. □

Note that $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2$ is the determinant of the Hessian matrix of $\phi(\mu,\lambda)$. If it is (weakly) positive for any μ and λ, then $\phi(\mu,\lambda)$ is jointly concave in (μ,λ), which is the case analyzed in Cai et al. (2017). Such meeting processes feature non-negative meeting externalities: an additional low-type worker does not make it harder for firms to meet high-type workers. A special case form the invariant meeting processes where $\phi_\lambda = 0$ and hence $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2 = 0$ so that the concavity condition in (16) is always satisfied. For our benchmark meeting process, this is the case when $\sigma = 1$;

for all other σ, we have $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2 < 0$. Since $m(\lambda)$ is always strictly concave in λ, the concavity condition in (16) is always satisfied when $\kappa(y) \to 0$ (i.e. worker heterogeneity disappears).

3.2 Optimal Queue Length and Composition

In our sorting analysis below, the fraction of high-type applicants across firms will play an important role. It is therefore convenient to reformulate the firm's problem in the following way: first, firms choose the fraction of high-type workers in their pool of applicants, which we denote by $\zeta \equiv \mu/\lambda \in [0,1]$; second, they choose the total queue length λ. In other words, the firms' problem is given by

$$\max_{\zeta\in[0,1]} \Pi^*(\zeta, y) \equiv \max_{\zeta\in[0,1]} \max_{\lambda>0} \Pi(\zeta, \lambda, y), \tag{17}$$

where $\Pi^*(\zeta, y) \equiv \max_{\lambda>0} \Pi(\zeta, \lambda, y)$ is the firm's value in the second step above, and $\Pi(\zeta, \lambda, y) = S(\zeta\lambda, \lambda, y) - \lambda U_1 - \zeta\lambda\,(U_2 - U_1)$, as defined in equation (11).

Optimal Queue Length. Working backwards, we first consider the choice of the queue length λ. Since $\phi(\zeta\lambda, \lambda)$ is strictly concave in λ for all $\zeta > 0$ and $m(\lambda) \equiv \phi(\lambda, \lambda)$, the payoff $\Pi(\zeta, \lambda, y)$ is *strictly* concave in λ for a given $\zeta \in [0, 1]$. Thus, assuming that firms of type y are active in hiring, the following first-order condition (FOC) with respect to λ determines a unique optimal queue length as a function of ζ and y, which we denote by $\lambda^o(\zeta, y)$:

$$U_1 + \zeta(U_2 - U_1) = m'(\lambda)\, f^1 + \frac{\partial \phi\,(\zeta\lambda, \lambda)}{\partial \lambda}\Delta f, \tag{18}$$

where $\partial \phi\,(\zeta\lambda, \lambda)\,/\partial\lambda \equiv \zeta\phi_\mu\,(\zeta\lambda, \lambda) + \phi_\lambda\,(\zeta\lambda, \lambda)$. To understand (18), note that the first term denotes the marginal contribution to surplus of a low-type applicant when all applicants are low-type. The second term corrects for the fact that a fraction ζ of workers are high-productivity workers.

To understand how the optimal queue length varies with firm type, we can differentiate equation (18) with respect to y. This yields

$$\lambda_y^o(\zeta, y) = \frac{\partial \lambda^o(\zeta, y)}{\partial y} = -\frac{m' f_y^1 + \frac{\partial \phi}{\partial \lambda}\Delta f_y}{m'' f^1 + \frac{\partial^2 \phi}{\partial \lambda^2}\Delta f}, \tag{19}$$

where we have suppressed arguments from $m(\lambda^o(\zeta, y))$ and $\phi(\zeta\lambda^o(\zeta, y), \lambda^o(\zeta, y))$. Since $\phi(\zeta\lambda, \lambda)$ is strictly increasing and concave in λ for $\zeta > 0$, the numerator in (19) is positive if $\Delta f_y \geq 0$ and the denominator is negative. In other words, when the opportunity costs of remaining unmatched are larger for more productive firms (i.e. supermodularity of the production function), those firms are more willing to invest in longer queues (holding ζ constant).

Optimal Queue Composition. Assuming that firms have solved for the optimal queue length $\lambda^o(\zeta, y)$, we next consider their choice of ζ. That is, firms now solve the maximization problem $\max_{\zeta\in[0,1]} \Pi^*(\zeta, y)$, where $\Pi^*(\zeta, y) = \Pi(\zeta, \lambda^o(\zeta, y), y)$. In general, $\Pi^*(\zeta, y)$ is not necessarily quasi-concave in ζ, so firms' maximization problem may admit multiple solutions. Denote by $Z(y)$ the set of all optimal ζ for firms of type y and let $\zeta(y)$ be an arbitrary element from $Z(y)$.

Suppose first that $\zeta(y)$ is interior, i.e. $\zeta(y) \in (0, 1)$. With a slight abuse of notation, $\zeta(y)$ and the corresponding optimal queue length $\lambda(y) = \lambda^o(\zeta(y), y)$ must then satisfy the FOC

$$\frac{\partial \Pi^*(\zeta, y)}{\partial \zeta}\Big|_{\zeta=\zeta(y)} = 0 \Leftrightarrow \phi_\mu(\zeta(y)\lambda(y), \lambda(y))\Delta f = U_2 - U_1, \tag{20}$$

where we used the envelope theorem and treated the total queue λ as constant in this exercise. Note that the left-hand side of the above equation is exactly the difference between the marginal contribution to surplus of high-type and low-type workers, given by equations (15) and (14), respectively. In words, a larger ζ increases the firm's probability of matching with a high-type worker, but comes at a cost as these workers are more expensive.[21] Finally, recall that by Lemma 5, an interior solution ζ which satisfies the FOC (20) also satisfies the second-order condition (SOC) if equation (16) holds.

The optimal ζ does not need to be interior. In case of corner solutions, i.e., $\zeta(y) = 0$ or $\zeta(y) = 1$, the following FOCs must be satisfied:

$$\zeta(y) = 0 \Rightarrow \phi_\mu(0, \lambda(y))\Delta f - (U_2 - U_1) \leq 0 \tag{21}$$

$$\zeta(y) = 1 \Rightarrow \phi_\mu(\lambda(y), \lambda(y))\Delta f - (U_2 - U_1) \geq 0, \tag{22}$$

[21] The firm can increase ζ by $\Delta\zeta$ while keeping λ the same by increasing the queue length of high-type workers by $\lambda\Delta\zeta$ and decreasing the queue length of low-type workers by $\lambda\Delta\zeta$.

where $\lambda(y) = \lambda^o(0, y)$ in equation (21), implicitly defined by $m'(\lambda)f^1 = U_1$, while $\lambda(y) = \lambda^o(1, y)$ in equation (22), implicitly defined by $m'(\lambda)f^2 = U_2$.

For future use, note that differentiating equation (18) with respect to ζ and evaluating the result at $\zeta = \zeta(y)$ gives

$$\lambda^o_\zeta(\zeta(y), y) = \frac{\partial \lambda^o(\zeta, y)}{\partial \zeta}\Big|_{\zeta=\zeta(y)} = -\frac{\lambda(y)\frac{\partial \phi_\mu}{\partial \lambda}\Delta f}{m'' f^1 + \frac{\partial^2 \phi}{\partial \lambda^2}\Delta f}, \quad \text{if } \zeta(y) \in (0, 1) \tag{23}$$

where we have used equation (20) to substitute out $U_2 - U_1$ and suppressed arguments from $m(\lambda(y))$ and $\phi(\zeta(y)\lambda(y), \lambda(y))$. For a given firm type y, this equation denotes the effect of a higher ζ on the optimal queue length, evaluating ζ at its equilibrium value $\zeta(y)$. It shows that along the equilibrium path, $\lambda^o_\zeta(\zeta(y), y)$ is negative since with a higher fraction of high-type workers, firms will reduce the total queue length to reduce negative hiring spillovers from low-productivity workers ($\partial\phi_\mu/\partial\lambda < 0$).

Multiplicity. As mentioned above, a firm's optimal choice of the fraction of high-type workers ζ may not be unique. That is, $Z(y)$ may contain multiple elements, which generally renders the analysis intractable. However, Cai et al. (2018) show that under a single-crossing condition which is satisfied by our benchmark meeting technology, the optimal queue compositions take the following simple form.

Lemma 6 (Cai et al., 2018)**.** *For any given firm type y, $Z(y)$ contains at most two elements, and when it does contain two elements, one of the two must be zero.*

That is, firms of a particular type y may have two optimal strategies. Some firms may go for *quality* by encouraging high-type workers to apply and limiting the number of low-type applicants in order to reduce congestion. Other firms of the same type may go for *quantity* and aim for a large hiring probability by attracting many low-type workers; however, this stops high-skilled workers from applying altogether. The above lemma shows that these two scenarios can be optimal simultaneously, but there are no other possibilities.

Limit Case. In general, the FOC—i.e. equation (20), (21) or (22)—is necessary but not sufficient for the optimum. Intuitively, however, when worker heterogeneity is sufficiently small, the tradeoff between quality and quantity disappears: the firms' problem becomes concave, the FOC sufficient, and the solution both unique and continuous. The following proposition formally establishes this result.

Proposition 2. *Fix $(x_1, \ell_1, \ell_2, J(y))$ and let $x_2 \to x_1$. For sufficiently small $x_2 - x_1$, each firm then has a unique optimal queue $(\mu(y), \lambda(y))$. Both $\mu(y)$ and $\lambda(y)$ are continuous in y, and if $0 < \mu(y_0) < \lambda(y_0)$ for some point y_0, then both $\mu(y)$ and $\lambda(y)$ and hence $\zeta(y) \equiv \mu(y)/\lambda(y)$ are differentiable at point y_0.*

Proof. See Appendix A.6. □

Inspection of the proof reveals that Proposition 2 only requires Property A0 and a weaker version of A1, namely $m(\lambda)$ is strictly increasing and concave in λ. This proposition, although simple and intuitive, will prove useful in the following section for constructing examples of equilibrium allocations that exhibit sorting.

4 Sorting

In this section, we analyze under what conditions the market equilibrium exhibits sorting. We derive a condition that states that positive assortative matching occurs when the relative benefits for a high-type firm of hiring a high-type worker exceed the costs. The benefits depend positively on the elasticity of complementarity of the production function. The costs depend on the likelihood that high-type workers end up in an applicant pool with other high-type workers, in which case a high-type worker does not contribute to surplus and it is a force against sorting. Below, we first provide our definitions of positive and negative sorting and then derive necessary and sufficient conditions.

4.1 Definition of Sorting

Much of the literature (see e.g. Becker, 1973; Shi, 2001; Eeckhout and Kircher, 2010) defines sorting in terms of a monotonic matching function which maps each worker type x to their employer type y.[22] This definition is not suitable in our environment, because firms do not necessarily hire a unique worker type. Instead, we require a set-based notion of sorting. Following Shimer and Smith (2000) and Shimer (2005), we therefore define sorting as first-order stochastic dominance (FOSD) in firms' distributions of hires.[23] In our environment, this definition can be expressed in terms of

[22]See Lindenlaub (2017) for a generalization to multidimensional types.

[23]Strictly speaking, Shimer and Smith (2000) use a *weaker* notion of sorting which is based on the bounds of the support of the distribution of hires; however, their definition is equivalent to FOSD of

the probability that a firm hires a high-type worker, conditional on hiring someone,

$$h(\zeta(y), \lambda(y)) \equiv \frac{\phi(\zeta(y)\lambda(y), \lambda(y))}{m(\lambda(y))}. \tag{24}$$

Recall that $\zeta(y)$ is an arbitrary element from $Z(y)$, the set of all optimal ζ for firms of type y, and $\lambda(y)$ is the corresponding optimal queue length, i.e., $\lambda(y) \equiv \lambda^o(\zeta(y), y)$.

Definition 3. *The market equilibrium exhibits positive (resp. negative) assortative matching if and only if* $h(\zeta(y), \lambda(y))$ *is weakly increasing (resp. decreasing) in* y.

While the literature has traditionally restricted attention to sorting patterns in matches, our environment yields additional predictions. After all, given that firms may interview multiple applicants and subsequently select the most desirable one, there is a meaningful distinction between an application on the one hand and a match on the other hand. Hence, in addition to PAM/NAM, we can also analyze the assortativeness of applications, for which we use the following definition.

Definition 4. *The market equilibrium exhibits positive (resp. negative) assortative contacting if and only if* $\zeta(y)$ *is weakly increasing (decreasing) in* y.

In words, PAC occurs if the fraction of high-type applicants increases in y. If the reverse occurs, we have NAC.

4.2 Quantity versus Quality

Before starting our sorting analysis, we first define a few new variables which are crucial for understanding firms' tradeoff between quantity and quality. They are determined by the search technology only and enable us to write the sorting conditions below in a uniform way.

A Simple Elasticity. If all workers were homogeneous, then more productive firms would attract more workers. With heterogeneous workers, firms now have two margins: queue length λ (quantity) and the fraction of high-type applicants ζ (quality).

this distribution in the random-search environment that they consider. In contrast, Shimer (2005) proves a *stronger* sorting result (high-type workers are more likely to be employed in high-type jobs than in low-type jobs) for a special case (multiplicatively separable production function and urn-ball meetings); however, he acknowledges that the data demands to test this result "may be unrealistic" and suggests FOSD of the distribution of hires as a "more easily testable" alternative.

They can adjust the quality by replacing low-type workers with high-type workers in their queues, which increase the match surplus if and only if the marginal high-type worker turns out to be the only high-type worker in the applicant pool, which occurs with probability $\phi_\mu(\zeta\lambda,\lambda)$. To investigate how firms' choices between quantity and quality change with firm types, it is natural to first fix ζ and consider how the probability $\phi_\mu(\zeta\lambda,\lambda)$ changes with λ. We are thus led to the following elasticity.

$$\varepsilon_w(\zeta,\lambda) \equiv \frac{\partial \log \phi_\mu(\zeta\lambda,\lambda)}{\partial \log \lambda} = \frac{\zeta\lambda\phi_{\mu\mu}(\zeta\lambda,\lambda) + \lambda\phi_{\mu\lambda}(\zeta\lambda,\lambda)}{\phi_\mu(\zeta\lambda,\lambda)}. \tag{25}$$

Because $\phi_\mu(\lambda\zeta,\lambda)$ is always decreasing in λ, $\varepsilon_w(\zeta,\lambda)$ is always strictly negative.

The Key Elasticity. As we mentioned above, firms have two margins: quantity λ and quality ζ. The quantity part determines the number of matches $m(\lambda)$, and the quality part determines the number of high-type matches $\phi(\lambda\zeta,\lambda)$. The marginal effect of an additional worker (varying λ) on the number of matches is given by $m'(\lambda)$, and the marginal effect of replacing a low-type applicant with a high-type applicant (varying ζ) on the number of high-type matches is $\phi_\mu(\lambda\zeta,\lambda)$. To capture the relative weight between quantity and quality faced by the firms, we consider the elasticity of $\phi_\mu(\zeta\lambda,\lambda)$ with respect to $m'(\lambda)$, which is given by the following.[24]

$$a^c(\zeta,\lambda) \equiv \frac{\partial \log \phi_\mu(\zeta\lambda,\lambda)}{\partial \log m'(\lambda)} = \varepsilon_w(\zeta,\lambda)\frac{m'(\lambda)}{\lambda m''(\lambda)}, \tag{26}$$

with extrema $\overline{a}^c \equiv \sup_{\zeta,\lambda} a^c(\zeta,\lambda)$ and $\underline{a}^c \equiv \inf_{\zeta,\lambda} a^c(\zeta,\lambda)$. Put differently, $a^c(\zeta,\lambda)$ measures the relative percentage changes of match quality and matching probability due to a longer queue while holding queue composition fixed.

The Matching Probability Adjusted Elasticity. In considering the relative percentage changes of match quality and matching probability, we kept ζ fixed. If, instead, we keep $h(\zeta,\lambda)$ fixed while varying both ζ and λ, then the firm must choose its queue composition and queue length according to $d\zeta = -\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}d\lambda$, which induces a percentage change of ϕ_μ equal to

$$\frac{1}{\phi_\mu}\left(\frac{\partial\phi_\mu}{\partial\zeta}d\zeta + \frac{\partial\phi_\mu}{\partial\lambda}d\lambda\right) = \left(1 - \frac{\partial\phi_\mu/\partial\zeta}{\partial\phi_\mu/\partial\lambda}\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}\right)\frac{\partial\phi_\mu}{\partial\lambda}\frac{d\lambda}{\phi_\mu}.$$

[24]To understand Equation (26), note that $\frac{\partial \log \phi_\mu(\zeta\lambda,\lambda)}{\partial \log m'(\lambda)} = \frac{\partial \log \phi_\mu(\zeta\lambda,\lambda)/\partial \log\lambda}{d \log m'(\lambda)/d\log\lambda}$.

The corresponding elasticity, which we will use in our analysis of assortative matching, therefore equals

$$a^m(\zeta,\lambda) = a^c(\zeta,\lambda)\left(1 - \frac{\partial\phi_\mu/\partial\zeta}{\partial\phi_\mu/\partial\lambda}\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}\right), \tag{27}$$

with extrema $\overline{a}^m \equiv \sup_{\zeta,\lambda} a^m(\zeta,\lambda)$ and $\underline{a}^m \equiv \inf_{\zeta,\lambda} a^m(\zeta,\lambda)$.

It is worth highlighting that $a^m(\zeta,\lambda)$ reduces to $a^c(\zeta,\lambda)$ when the queue is homogeneous, i.e. $a^m(\zeta,\lambda) = a^c(\zeta,\lambda)$ when $\zeta = 0$ or $\zeta = 1$.[25] As we will see later for our benchmark model and various other meeting technologies, the infimum (resp. supremum) of a^c and a^m can be reached or approached with $\zeta = 0$ (resp. $\zeta = 1$). It is therefore not surprising that *in general* $\overline{a}^c = \overline{a}^m$ and $\underline{a}^c = \underline{a}^m$.

4.3 Conditions for Sorting

We can now analyze the conditions for PAC/PAM to hold locally around y. To do so, we will distinguish between $Z(y)$ having a single element and it having two elements.

Unique Optimal Queue Composition. We first consider the simple case where $\zeta(y)$ is both unique and interior around some y. This allows us to differentiate the FOCs (18) and (20) with respect to y, giving us two equations that jointly determine $\lambda'(y)$ and $\zeta'(y)$, i.e. how the optimal queue length and composition vary with y.

It is instructive to first differentiate (20) to obtain $\zeta'(y)$ as a function of $\lambda'(y)$, i.e.

$$\zeta'(y) = -\frac{\phi_\mu}{\partial\phi_\mu/\partial\zeta}\frac{\Delta f_y}{\Delta f} - \frac{\partial\phi_\mu/\partial\lambda}{\partial\phi_\mu/\partial\zeta}\lambda'(y). \tag{28}$$

Hence, $\zeta'(y) \geq 0$, i.e. PAC holds locally, if and only if

$$\frac{\Delta f_y}{\Delta f} \geq -\frac{\partial\phi_\mu/\partial\lambda}{\phi_\mu}\lambda'(y) = -\varepsilon_w(\zeta(y),\lambda(y))\frac{\lambda'(y)}{\lambda(y)}. \tag{29}$$

As (29) reveals, how the optimal queue composition varies with firm productivity depends on a race between two forces. The left-hand side measures the complementarities in production, which are a force for PAC. However, if higher-type firms attract

[25] To see this, note that $\phi(0,\lambda) = 0$ for any λ, which implies that $\partial h/\partial\lambda = 0$. Similarly, $\phi(\lambda,\lambda) = m(\lambda)$ by definition, which again implies that $\partial h/\partial\lambda = 0$.

longer queue lengths, then that reduces the probability that high-type applicants create surplus, which is a force against PAC, measured by the right-hand side.

The analysis for PAM follows a similar path. Differentiating $h(\zeta(y), \lambda(y))$ with respect to y (along the equilibrium path) shows that

$$\frac{dh(\zeta(y), \lambda(y))}{dy} \geq 0 \quad \Longleftrightarrow \quad \zeta'(y) \geq -\frac{\partial h/\partial \lambda}{\partial h/\partial \zeta}\lambda'(y).$$

Using equation (28) to eliminate $\zeta'(y)$ from the above equation then yields

$$\frac{\Delta f_y}{\Delta f} \geq -\varepsilon_w(\zeta(y), \lambda(y))\frac{\lambda'(y)}{\lambda(y)}\left(1 - \frac{\partial \phi_\mu/\partial \zeta}{\partial \phi_\mu/\partial \lambda}\frac{\partial h/\partial \lambda}{\partial h/\partial \zeta}\right). \tag{30}$$

The interpretation of (30) follows the same logic as in the PAC case. Firms can invest in the expected match quality by choosing an appropriate expected pool of applicants. Again by equation (20), if (30) holds then a firm with a higher y will take advantage of its type by choosing a higher conditional probability to hire a high type, $h(\zeta(y), \lambda(y))$.

Equations (29) and (30) take $\lambda'(y)$ as given, but it is of course endogenous. We can differentiate FOC (18) to obtain a second relation between $\zeta'(y)$ and $\lambda'(y)$, which can be used to eliminate $\lambda'(y)$. These steps give us the following result.

Lemma 7. *Assume $\zeta(y)$ is both interior and continuous at a point y. PAC (resp. PAM) then holds locally at y if and only if for $i = c$ (resp. $i = m$), we have*

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq a^i \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}}, \tag{31}$$

where we suppress the arguments of $\phi(\zeta(y)\lambda(y), \lambda(y))$, $m(\lambda(y))$ and $a^i(\zeta(y), \lambda(y))$.

Proof. See Appendix A.7. □

Multiplicity in Optimal Queue Composition. Consider now the case in which $Z(y)$ is not a singleton. By Lemma 6, it then has two elements, with one of them equal to zero. Denote these elements by ζ_0 and ζ_1, satisfying $0 = \zeta_0 < \zeta_1$, and their corresponding queue lengths by λ_0 and λ_1, respectively. Since firm y must be indifferent, their expected payoff—or, equivalently, their marginal contribution to surplus—must be the same for a queue $(0, \lambda_0)$ as for (ζ_1, λ_1). By (13), we have

$$m(\lambda_0) - \lambda_0 m'(\lambda_0) = m(\lambda_1) - \lambda_1 m'(\lambda_1) + \left(\phi(\zeta_1\lambda_1, \lambda_1) - \lambda_1 \frac{d\phi(\zeta_1\lambda_1, \lambda_1)}{d\lambda}\right)\frac{\Delta f}{f^1}, \quad (32)$$

where the left-hand side is the firm's marginal contribution to surplus with a queue $(0, \lambda_0)$, divided by f^1, and the right-hand side is the corresponding value with a queue (ζ_1, λ_1).

Similarly, if $\zeta_1 \in (0, 1)$, then low-type workers are present in both queues and their marginal contribution to surplus must be the same. Equation (14) then yields

$$m'(\lambda_0) = m'(\lambda_1) + \phi_\lambda(\zeta_1\lambda_1, \lambda_1)\frac{\Delta f}{f^1}, \quad \text{if } \zeta_1 \in (0, 1). \quad (33)$$

If $\zeta_1 = 1$, then the low-type workers are not present in the shorter queue. In this special case, firm y's optimality requires that the left-hand side above is larger than the right-hand side.

For firm types y' close to y, by continuity, $Z(y')$ is either the corner solution 0, or some ζ' close to ζ_1, or both. We say that PAC/PAM holds locally around a multiplicity point y if for y' sufficiently close to y, the corner solution 0 is the only solution for $y' < y$, but that it does not belong to $Z(y')$ if $y' > y$.[26] The above requirement implies that $Z(y')$ is unique for y' slightly above y, else 0 must belong to $Z(y')$. Note that the above condition for local PAC/PAM does *not* require that for y' slightly above y, $\zeta(y') \geq \zeta_1$ (PAC) or $h(\zeta(y'), \lambda(y')) \geq h(\zeta_1, \lambda_1)$ (PAM). Those conditions are only necessary when $Z(y)$ is unique (below in the next section we elaborate on this further).

By the envelope theorem, a *sufficient* condition for PAC/PAM to hold around point y is $\Pi_y(0, \lambda_0, y) < \Pi_y(\zeta_1, \lambda_1, y)$, where $\Pi(\zeta, \lambda, y)$ is the expected profit of a firm y with queue (ζ, λ) which is defined in equation (11).[27] The condition $\Pi_y(0, \lambda_0, y) <$

[26] Recall that PAC requires that any element of $Z(y')$ is smaller (resp. larger) than any element of $Z(y)$ when $y' < y$ (resp. $y' > y$). Similarly, if $Z(y')$ contains a nonzero element for $y' < y$, or contains zero for $y' > y$, then PAM fails around y because $h(\zeta, \lambda) = 0$ if and only if $\zeta = 0$.

[27] The envelope theorem states that for a firm with type y' close to y, if it is constrained to choose only ζ' close to ζ_1, then the maximum expected profit is approximately (first-order): $\Pi(\zeta_1, \lambda_1, y) + \Pi_y(\zeta_1, \lambda_1, y)\Delta y$ where $\Delta y = y' - y$. Similarly, if the firm is constrained to choose $\zeta' = 0$, then the maximum expected profit is approximately $\Pi(0, \lambda_0, y) + \Pi_y(0, \lambda_0, y)\Delta y$. When $\Pi_y(\zeta_1, \lambda_1, y) > \Pi_y(\zeta_1, \lambda_1, y)$, then a firm type $y' > y$ strictly prefers to choose ζ' around ζ_1 instead of zero, and a firm type $y' < y$ strictly prefers zero. As mentioned before, by continuity it is without loss of generality to constrain the firm to choose between zero and all ζ' close to ζ_1. See Milgrom and Segal

$\Pi_y(\zeta_1, \lambda_1, y)$ can be written explicitly as follows.

$$m(\lambda_0) < m(\lambda_1) + \phi(\zeta_1\lambda_1, \lambda_1)\frac{\Delta f_y}{f_y^1}. \tag{34}$$

If the reverse inequality $\Pi_y(0, \lambda_0, y) > \Pi_y(\zeta_1, \lambda_1, y)$ holds, then the opposite is true, i.e., for y' slightly above y, $Z(y')$ is unique and equals zero, and for y' slightly below y, $Z(y')$ is unique and equals some ζ' around ζ_1. The above result implies that a *necessary* condition for PAC/PAM to hold around y is that the $<$ sign in the above inequality (34) is replaced by $\leq$, because otherwise (with $>$) the optimal ζ for firm types slightly above y must be zero, and PAC/PAM fails. Finally, the case of NAC/NAM is similar except that we need to reverse the relevant inequalities.

From Local to Global. Suppose that PAC/PAM holds locally at all points, then we show below that PAC/PAM holds globally. To see this, note first that when PAC/PAM holds locally at some multiplicity point y, then all points sufficiently close to y but not equal to y have a unique optimal ζ, which implies that multiplicity points are isolated from each other. Furthermore, there exists at most one multiplicity point. Suppose otherwise that there are two consecutive points y' and y'' with $Z(y') = \{0, \zeta_1'\}$ and $Z(y'') = \{0, \zeta_1''\}$ (because multiplicity points are isolated, we can talk about *consecutive* multiplicity points). Then by construction, $Z(y)$ is unique for all y between y' and y''. Since PAC/PAM holds locally for all points between y' and y'', $\zeta(y)$ must be continuously increasing, which can not be true because its limit value at the left end is ζ_1' and its limit value at the right end is zero.[28] Hence, we get a contradiction. There exists at most one multiplicity point below which $Z(y)$ is unique and equals zero, and above which $Z(y)$ is unique and continuously increasing. A similar logic applies for NAC/NAM.

A Simple Necessary Condition for Sorting. In general, it is difficult to completely solve the model and show whether or not $Z(y)$ is unique. However, the limit case considered by Proposition 2 where worker heterogeneity disappears is particularly simple since all firms have a unique optimal ζ. Furthermore, equation (31), the

(2002) for a discussion of envelope theorems.

[28] To be precise, here we used the fact that $Z(y)$ is an upper-hemi continuous correspondence, which follows from the theorem of the maximum.

condition for PAC/PAM in this case, reduces to something particularly simple when $x_2, x_1 \to x$. The left-hand side reduces to $\rho(x, y)$, and the right-hand side simply becomes a^i. Our next result provides a simple necessary condition for assortative contacting/matching to hold based on this limit case.

Proposition 3. *A necessary condition for PAC (resp. PAM) to hold for any endowment of agents is that for $i = c$ (resp. $i = m$), we have*

$$\underline{\rho} \equiv \inf_{x,y} \rho(x, y) \geq \sup_{\zeta,\lambda} a^i(\zeta, \lambda) \equiv \overline{a}^i. \tag{35}$$

Similarly, a necessary condition for NAC (resp. NAM) to hold for any endowment of agents is that for $i = c$ (resp. $i = m$), we have

$$\overline{\rho} \equiv \sup_{x,y} \rho(x, y) \leq \inf_{0 \leq \mu \leq \lambda} a^i(\zeta, \lambda) \equiv \underline{a}^i. \tag{36}$$

Proof. See Appendix A.8. □

The intuition for the above result will be discussed later after Proposition 4.

Sufficiency. Of course, the question remains whether the above necessary condition is also sufficient for any endowment of agents. In the following section, we show that the answer is 'yes' for our benchmark meeting technology. In Section 5, we show that this conclusion extends to other examples of meeting technologies.

4.4 How Screening Affects Sorting

Our analysis so far has been quite general and has not really made use of the functional form of $\phi(\mu, \lambda)$ given by (7). The following lemma establishes that this functional form yields very simple expressions for the right-hand side of equations (35) and (36).

Lemma 8. *When $\phi(\mu, \lambda)$ is given by (7), then $a^m(\zeta, \lambda) \leq a^c(\zeta, \lambda)$, and*

$$\overline{a}^c = \overline{a}^m = \frac{1 + \sigma}{2} \quad \text{and} \quad \underline{a}^c = \underline{a}^m = \frac{1 - \sigma}{2}. \tag{37}$$

Proof. See Appendix A.10. □

We have shown in Proposition 3 and Proposition 10 in Appendix A.9 that $\underline{\rho} \geq (1 + \sigma)/2$ is both sufficient and necessary for PAC/PAM to hold at points where

$Z(y)$ is interior and unique, similarly, for NAC/NAM to hold at these points we need $\overline{\rho} \leq (1-\sigma)/2$. Proposition 4 below shows that the above two conditions are sufficient for PAC/PAM and NAC/NAM to hold at jump points, respectively. Thus we have derived the necessary and sufficient condition for sorting to hold for any endowment of agents.

Proposition 4. *Assume that ϕ is given by equation (7) with $\sigma > 0$. Then the market equilibrium exhibits PAC/PAM for any endowment of agents if and only if $\underline{\rho} \geq (1+\sigma)/2$. In contrast, the market equilibrium exhibits NAC/NAM for any endowment of agents if and only if $\overline{\rho} \leq (1-\sigma)/2$.*

Proof. See Appendix A.11. □

When $\sigma \to 0$ and meetings are bilateral, PAC/PAM requires square-root supermodularity, in line with the results in Eeckhout and Kircher (2010). At the other extreme, log-supermodularity is required for PAC/PAM when $\sigma = 1$ and firms can interview all their applicants. In contrast, for NAC/NAM, a stronger degree of substitutability is required as the expected number of interviews goes up: the production function should be square-root-submodular if $\sigma = 0$ and submodular when $\sigma = 1$. Figure 1 plots these results for the case of CES production where $\rho(x, y)$ is constant.

The planner's problem is to assign workers and firms to different segments such that aggregate output is maximized. The planner takes into account that within each segment, meeting frictions frustrate the assignment problem in two ways: (i) some workers do not meet firms and simultaneously, some firms do not meet any workers, (ii) firms select the highest worker type that they screen but some resources are wasted since workers with a type below or equal to the selected worker contribute nothing to surplus. The planner wants to reduce the likelihood of both events. If the planner forces all firms to choose the same fraction of high-type workers in their queues, then the only way that firms with a relatively high y can reduce the likelihood of remaining unmatched is by choosing a longer queue (see equation (19)), which leads to a higher matching probability. The increase in queue length changes the relative weight of match quality and match probability, which is captured by $a^c(\zeta, \lambda)$, the elasticity of $\phi_\mu(\lambda\zeta, \lambda)$ with respect to $m'(\lambda)$ (see equation (26)). When this elasticity is large, then there is a relatively large drop in match quality associated with a longer queue. In order to accommodate this drop, firms reduce the fraction of high-type workers, which constitutes a force against PAC. In contrast, a higher elasticity of complementarity is a force towards PAC because then high-type firms are more keen to match with high-type workers. Since the force against PAC is bounded above by $\overline{a}^c$ and the force towards PAC (elasticity of complementarity) is bounded below by $\underline{\rho}$, it is not surprising that $\underline{\rho} \geq \overline{a}^c$ is the condition for PAC. The case for PAM is similar, since $a^m(\zeta, \lambda)$ measures the same relative weight while keeping $h(\zeta, \lambda)$ fixed.

The force against positive sorting is in general strongest when the relative supply of high-type workers is large, because then it is more likely that there are other high type workers in the pool of applicants. When σ increases, the force against PAC/PAM becomes stronger because then firms are more likely to also encourage low-type workers to apply.

The force against negative sorting is generally strongest when there are relatively few high-type workers because then it is less likely that other high type workers contacted the firm. In that case there is no need to encourage high type workers to apply to low y firms where the queues are shorter because the queues at the high type firms will also be short so it will be unlikely that they crowd out each other there. When σ is larger, this effect becomes even stronger because high type workers won't be crowded out by low type workers so they can safely visit high y firms and consequently NAC/NAM becomes harder to sustain.

Given definition 1, we can alternatively state Proposition 4 as follows.

Corollary 1. *The market equilibrium exhibits PAC/PAM for any endowment of agents if and only if the production function is* $2/(1-\sigma)$*-root-supermodular. In contrast, the market equilibrium exhibits NAC/NAM for any endowment of agents if and only if the production function is* $2/(1+\sigma)$*-root-submodular.*

We have showed that the simple necessary condition in Proposition 3 is also sufficient for our benchmark contact technology. In the following section, we will show that this result extends to other classes of contact technologies.

5 Extensions

In this section, we explore various extensions of our environment and discuss their impact on the sorting conditions.

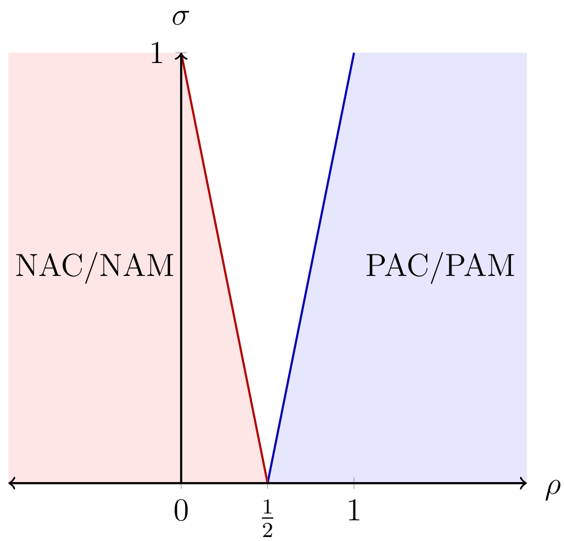

Figure 1: Combinations of ρ and σ that give rise to PAC/PAM (blue) or NAC/NAM (red) for any endowment of agents, assuming a CES production function.

5.1 Alternative Contact Technologies

We have derived our main result, Proposition 4, for a specific micro-foundation for the contact technology, such that $\phi(\mu, \lambda)$ was given by (7). This technology was chosen because of its analytic tractability, but many reasonable alternatives exist.[29] Since most of our analysis is presented in a general way, such alternatives can be analyzed by simply updating the form of $\phi(\mu, \lambda)$. The necessary conditions for sorting continue to be given by Proposition 3. Proving that these conditions are also sufficient is more challenging because one needs to show that the contact technology satisfies all the required regularity conditions (see Proposition 10 and the proof of Proposition 4 for details).

To illustrate this, we show below that how our main result carries over to a contact technology where workers send their applications in the first stage according to an urn-ball process (as in Shimer, 2005), and the screening process in the second stage remains geometric (see also Wolthoff, 2018). We prove in Appendix A.2 that $\phi(\mu, \lambda)$ is given by

$$\phi(\mu, \lambda) = \frac{\mu}{\sigma\mu + (1-\sigma)\lambda}\left(1 - e^{-\sigma\mu-(1-\sigma)\lambda}\right). \tag{38}$$

[29] See Lester et al. (2015) and Cai et al. (2017) for various examples.

Lemma 9. *When $\phi(\mu, \lambda)$ is given by (38), then $a^m(\zeta, \lambda) \leq a^c(\zeta, \lambda)$, and*

$$\overline{a}^c = \overline{a}^m = \frac{1+\sigma}{2} \qquad \text{and} \qquad \underline{a}^c = \underline{a}^m = 0. \tag{39}$$

Proof. See Appendix A.13. □

Proposition 5. *Assume that ϕ is given by equation (38). Then the market equilibrium exhibits PAC/PAM for any endowment of agents if and only if $\underline{\rho} \geq (1+\sigma)/2$. In contrast, the market equilibrium exhibits NAC/NAM for any endowment of agents if and only if $\overline{\rho} \leq 0$ (i.e. f is submodular).*

Proof. See Appendix A.14. □

Proposition 6. *Suppose that ϕ satisfies Property A0, A1, A2, and A3. Furthermore, assume that $0 \leq \underline{a}^i \leq \overline{a}^i \leq 1$ for $i \in \{c, m\}$, then the market equilibrium exhibits PAC/PAM for any endowment of agents iff $\underline{\rho} \geq 1$ (log-supermodularity). In contrast, the market equilibrium exhibits NAC/NAM for any endowment of agents if $\overline{\rho} \leq 0$ (i.e. f is submodular).*

5.2 Bilateral Contact Technologies

Much of the search literature uses bilateral contact technologies, which could be interpreted as firms being able to interview only a single applicant.[30] In that case, $\phi(\mu, \lambda) = m(\lambda)\mu/\lambda$, i.e. a firm interviews a high-type applicant if it has at least one applicant and the applicant selected for the interview is of high type. The following lemma establishes that $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$ then become independent of ζ and both reduce to the elasticity of substitution of the total number of matches in a submarket, which is precisely the object that Eeckhout and Kircher (2010) show to be important in their study of sorting patterns for bilateral technologies.

Lemma 10. *When the contact technology is bilateral, we have*

$$a^c(\zeta, \lambda) = a^m(\zeta, \lambda) = \frac{m'(\lambda)(\lambda m'(\lambda) - m(\lambda))}{\lambda m(\lambda) m''(\lambda)}. \tag{40}$$

[30]Examples include Moen (1997) and Acemoglu and Shimer (1999).

Proof. See Appendix A.15. □

With bilateral technologies, firms always find it optimal to attract either only low-type or only high-type workers. For the analysis of sorting, we therefore only need to consider multiplicity points. Our analysis in Section 4.3 can be easily applied to this case. We do so in Appendix C.1 for our setup with two types of workers and demonstrate that we recover the sorting results of Eeckhout and Kircher (2010), who assume a continuum of worker types.

5.3 Invariant Contact Technologies

Another common choice in the search literature are invariant technologies, with as most prominent example urn-ball.[31] Screening in these technologies is perfect in the sense that the presence of low-type applicants does not make it harder (or easier) for a firm to identify a high-type applicant. That is, $\phi_\lambda(\mu,\lambda) = 0$ for all μ and λ, or equivalently, $\phi(\mu,\lambda) = \phi(\mu,\mu) \equiv m(\mu)$, where $m(\mu)$ is always assumed to be strictly concave (see Cai et al., 2017).

The elasticities $\varepsilon_f(\zeta,\lambda)$ and $\varepsilon_w(\zeta,\lambda)$, defined by (**??**) and (25), also only depend on $\mu = \lambda\zeta$ in this case; with a slight abuse of notation, they can be written as

$$\epsilon_f(\mu) = \frac{\mu m'(\mu)}{m(\mu)} \quad \text{and} \quad \epsilon_w(\mu) = \frac{\mu m''(\mu)}{m'(\mu)}. \tag{41}$$

The firm elasticity now simply measures the relative increase in the probability that a firm hires a high type worker due to an increase in the number of high type workers per vacancy. The worker elasticity gives the relative decrease in the probability that a high type worker is hired due to an increase in the number of high type workers per vacancy. The following lemma then presents $a^c(\zeta,\lambda)$ and $a^m(\zeta,\lambda)$ for invariant technologies.

Lemma 11. *When the contact technology is invariant, we have*

$$a^c(\zeta,\lambda) = \frac{\epsilon_w(\lambda\zeta)}{\epsilon_w(\lambda)} \quad \textit{and} \quad a^m(\zeta,\lambda) = \frac{\epsilon_w(\lambda\zeta)}{\epsilon_w(\lambda)}\frac{\epsilon_f(\lambda)}{\epsilon_f(\lambda\zeta)}, \tag{42}$$

with extrema $\underline{a}^c = \underline{a}^m = 0$ *and* $\overline{a}^c, \overline{a}^m \geq 1$.

[31] See Lester et al. (2015).

Proof. See Appendix A.16. □

When contacts are invariant, the surplus function $S(\mu, \lambda, y)$ defined in equation (8) simplifies to $m(\lambda) f(x_1, y) + m(\mu)[f(x_2, y) - f(x_1, y)]$. The firms' problem becomes concave in (μ, λ), which implies that for each y, there exists exactly one optimal queue $(\mu(y), \lambda(y))$, which, by the theorem of the maximum, is continuous in y. Therefore, the scenario where $\zeta(y)$ contains jumps, i.e., a firm's optimal choice of the fraction of high-type workers is not unique, never occurs, so we only need to consider the requirement that $\zeta(y)$ is unique and interior.

Our next result shows that the necessary conditions from Proposition 3 are also sufficient when the contact technology is invariant.

Proposition 7. *Suppose the contact technology is invariant. The market equilibrium then exhibits PAC (resp. PAM) for any endowment of agents if and only if* $\underline{\rho} \geq \overline{a}^c$ *(resp.* $\underline{\rho} \geq \overline{a}^m$*). In contrast, the market equilibrium exhibits NAC/NAM for any endowment of agents if and only if* $f(x, y)$ *is submodular.*

Proof. See Appendix A.17. □

Under the following two mild assumptions, which are satisfied by urn-ball or a geometric technology, the sorting results become particularly simple.[32]

Assumption INV-1. $\varepsilon_w(\mu)$ *is decreasing in* μ.

Assumption INV-2. $\varepsilon_w(\mu)/\varepsilon_f(\mu)$ *is decreasing in* μ.

Assumption INV-1 (resp. INV-2) implies that $a^c(\zeta, \lambda)$ (resp. $a^m(\zeta, \lambda)$) is increasing in ζ so that $\overline{a}^c = 1$ (resp. $\overline{a}^m = 1$). The condition for PAC/PAM then becomes $\underline{\rho} \geq 1$, which is equivalent to $f(x, y)$ being log-supermodular.

Corollary 2. *When the contact technology is invariant and satisfies Assumption INV-1 (resp. INV-2), the market equilibrium exhibits PAC (resp. PAM) for any endowment of agents if and only if* $f(x, y)$ *is log-supermodular.*

In Appendix C.2, we show that the results for invariant technologies can be generalized to N worker types in a straightforward manner. When the meeting technology is neither bilateral nor invariant, the multiplicity problem becomes very tedious and therefore for those cases, we stick to two worker types.

[32]However, one can construct invariant technologies that do not satisfy these assumptions, e.g. a mixture between urn-ball and geometric: $m(\mu) = t(1 - e^{-\mu}) + (1 - t)\mu/(1 + \mu)$ with $t \in [0, 1]$. Numerically, one can see that $\overline{a}^c > 1$ and $\overline{a}^m = 1$ when $t = 0.2$, while $\overline{a}^c, \overline{a}^m > 1$ when $t = 0.98$.

5.4 Signals

In our benchmark model, firms have no information about applicants' types when selecting interviewees. In practice, there often exist relatively easy ways to obtain a signal, e.g. from a quick look at applicants' resumes. As we show in this section, our baseline environment can be extended quite easily to capture this idea.

Environment with Signals. Consider an environment which is like our benchmark model, except that firms can costlessly observe a signal for every applicant. For high-type applicants, the signal is positive with certainty. In contrast, a low-type applicant generates a correct negative signal with probability $\tau \in [0,1]$ and an incorrect positive signal with complementary probability. In other words, τ is a measure of the amount of information contained by signals: they are worthless if $\tau = 0$, but perfectly reveal applicants' types when $\tau = 1$. Using this information, firms will prioritize applicants with positive signals when selecting interviewees and only select applicants with negative signals if interview capacity remains. As before, an interview reveals the true type of an applicant, allowing firms to hire the most profitable candidate and pay a type-contingent wage.

Isomorphism. The following proposition establishes that this modified environment is isomorphic to our baseline model, as long as we transform the parameter σ to account for the fact that firms also obtain information from signals.

Proposition 8. *In our environment with signals, consider a firm with a queue μ of high-type workers and a queue $\lambda - \mu$ of low-type workers. The probability that the firm interviews at least one high-type worker equals*

$$\phi(\mu, \lambda) = \frac{\mu}{1 + \widehat{\sigma}\mu + (1 - \widehat{\sigma})\lambda},$$

where $\widehat{\sigma} = 1 - (1 - \tau)(1 - \sigma) \in [0, 1]$.

Proof. See Appendix A.18. □

As a direct consequence of this proposition, all our earlier results carry over to the environment with signals, except that they apply to $\widehat{\sigma}$ instead of σ.

5.5 Endogenous Screening

The screening intensity σ was exogenous in our baseline model. In real life, however, firms can generally influence the number of applicants that they interview. In this section, we therefore endogenize σ and discuss how it affects our sorting results.

Environment with Endogenous Screening. Consider an environment which is like our benchmark model, except that firms additionally choose (and post) their recruiting intensity $\sigma \in [0,1]$ at a linear cost $c\sigma$, where $c \geq 0$.[33] That is, they solve

$$\max_{\sigma,\mu,\lambda} \quad \frac{\lambda}{1+\lambda} f^1 + \frac{\mu}{1+\sigma\mu+(1-\sigma)\lambda} \Delta f - \lambda U_1 - \mu \Delta U - c\sigma. \tag{43}$$

Since the second term above is convex in σ and $c\sigma$ is linear, the above profit function is *convex* in σ. The maximum is therefore reached at a corner, i.e. when $\sigma = 0$ or 1. In other words, firms only need to choose between two extremes: no screening or perfect screening, which correspond to a bilateral and an invariant meeting technology, respectively. To determine firms' choice, we compare the profits from the two options.

Profits with No Screening. Consider a firm of type y choosing $\sigma = 0$. This firm's optimal queue then exists of either low-type workers or high-type workers, but not both.[34] Suppose the firm attracts workers of type x_i. Equation (43) then reduces to $\max_{\lambda_i} m(\lambda_i) f(x_i, y) - \lambda_i U_i$. Because $m(\lambda) = \lambda/(1+\lambda)$ is strictly concave, the FOC of this problem is both necessary and sufficient. Assuming that $f(x_i, y) > U_i$, the optimal queue length is $\lambda_i = \sqrt{f(x_i, y)/U_i} - 1$, which yields an expected payoff of

$$\pi_i(y) = \left(\sqrt{f(x_i, y)} - \sqrt{U_i}\right)^2. \tag{44}$$

Naturally, the firm chooses the type of workers it wishes to attract based on whether $\pi_1(y)$ or $\pi_2(y)$ is higher. The profits from $\sigma = 0$ therefore equal $\max\{\pi_1(y), \pi_2(y)\}$.

[33] Posting contracts that include σ in addition to wages is necessary for constrained efficiency in this environment. More restrictive contract spaces and more general cost functions are left for future research. Wolthoff (2018) endogenizes σ in a similar way as us, but with a cost function that is sufficiently convex (in an otherwise quite different model).

[34] See Section 5.2 or Eeckhout and Kircher (2010).

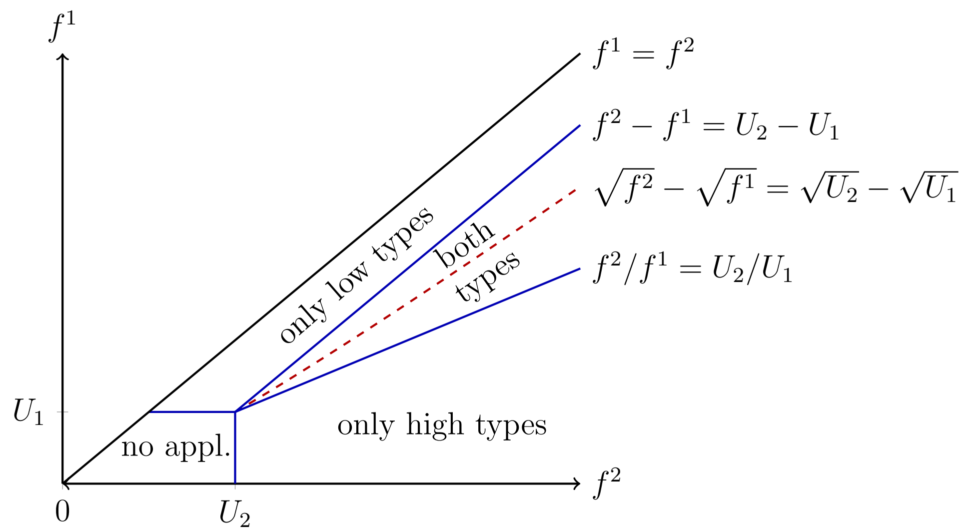

Figure 2: Optimal applicant pool for a firm, conditional on $\sigma = 1$.

Profits with Perfect Screening. When the firm chooses $\sigma = 1$, (43) reduces to

$$\overline{\pi}(y) \equiv \max_{0 \le \mu \le \lambda} \frac{\lambda}{1+\lambda} f^1 + \frac{\mu}{1+\mu} \Delta f - \lambda U_1 - \mu \Delta U - c. \tag{45}$$

This problem is concave, so the only complexity lies in the constraint $0 \le \mu \le \lambda$. As we show in Appendix B and Figure 2, there are four possibilities with respect to the optimal applicant pool of a firm, conditional on choosing $\sigma = 1$: the firm may attract (i) no applicants, such that $\overline{\pi}(y) = -c$; (ii) only low-type applicants, such that $\overline{\pi}(y) = \pi_1(y) - c$; (iii) only high-type applicants, such that $\overline{\pi}(y) = \pi_2(y) - c$; or (iv) both types of applicants, in which case the FOCs imply $\mu = \sqrt{\Delta f / \Delta U} - 1$ and $\lambda = \sqrt{f^1/U_1} - 1$, such that

$$\overline{\pi}(y) = \left(\sqrt{f^1} - \sqrt{U_1}\right)^2 + \left(\sqrt{\Delta f} - \sqrt{\Delta U}\right)^2 - c. \tag{46}$$

Choice of Screening Intensity. Clearly, a necessary condition for $\sigma = 1$ to yield higher profits than $\sigma = 0$ is that the firm attracts both types of applicants. In what follows, we will therefore focus on this case, which occurs when

$$f(x_2, y) - f(x_1, y) > U_2 - U_1 \quad \text{and} \quad f(x_2, y)/f(x_1, y) < U_2/U_1. \tag{47}$$

As the red dashed line in Figure 2 shows, the region described by (47) is divided into two parts by the curve $\pi_1(y) = \pi_2(y)$, or equivalently

$$\sqrt{f^2} - \sqrt{f^1} = \sqrt{U_2} - \sqrt{U_1}. \tag{48}$$

We therefore have to distinguish between two cases when calculating the difference in profits between $\sigma = 0$ and $\sigma = 1$ in this region, i.e. $\Delta\pi(y) \equiv \overline{\pi}(y) - \max\{\pi_1(y), \pi_2(y)\}$. The following lemma formalizes this.

Lemma 12. *Suppose that $\pi_1(y) = \pi_2(y) > 0$, then (47) holds. In the region characterized by (47), the difference in profits between $\sigma = 1$ and $\sigma = 0$ equals*

$$\Delta\pi(y) = \begin{cases} \left(\sqrt{\Delta f} - \sqrt{\Delta U}\right)^2 - c & \text{if } \pi_1(y) \geq \pi_2(y), \quad (49a) \\ 2\left(\sqrt{f^2 U_2} - \sqrt{f^1 U_1} - \sqrt{\Delta f \Delta U}\right) - c & \text{if } \pi_1(y) \leq \pi_2(y). \quad (49b) \end{cases}$$

Proof. See Appendix A.19. □

This characterization of $\Delta\pi(y)$ completes the analysis of the firm's choice problem given in (43): the firm chooses $\sigma = 1$ if $\Delta\pi(y)$ is positive and $\sigma = 0$ otherwise.

Necessary Conditions for Sorting. Next, we analyze how the solution varies with firm type. Specifically, under what conditions do PAC/PAM or NAC/NAM continue to hold? When $c = 0$, all firms will choose $\sigma = 1$. In this limit case, we already know that for PAC/PAM we need log-supermodularity and for NAC/NAM we need submodularity. Below we analyze whether they are also sufficient. The results are given in Proposition 9.

Sufficient Condition for NAC/NAM. Suppose that $f(x, y)$ is strictly submodular and that there exists some y^{EK} which solves (48).[35] is strictly decreasing in y. Hence, $\pi_2(y) < \pi_1(y)$ for firms with $y > y^{EK}$ and vice versa. Using this fact, the following Lemma then establishes how $\Delta\pi(y)$ varies with y.

Lemma 13. *If $f(x, y)$ is strictly submodular, then $\Delta\pi(y)$ is strictly increasing in y for $y \leq y^{EK}$ and strictly decreasing in y for $y \geq y^{EK}$.*

[35] That is, y^{EK} is the firm type that is indifferent between low-type and high-type workers in a bilateral world like Eeckhout and Kircher (2010).] Since submodularity implies square-root submodularity, the left-hand side of (48)

Proof. See Appendix A.20. □

This result implies that firms with type y^{EK} have the strongest incentive to screen. If in equilibrium $\Delta\pi(y^{EK}) \leq 0$, then all firms will choose $\sigma = 0$, in which case submodularity implies NAC/NAM by Proposition 4. In contrast, if $\Delta\pi(y^{EK}) > 0$, then Lemma 13 above implies that firms with low y will choose $\sigma = 0$ and attract only high-type workers; firms with y around y^{EK} will choose $\sigma = 1$ and attract both types of workers; finally, firms with high y will choose $\sigma = 0$ and attract only low-type workers.[36] As we have showed before, submodularity of $f(x, y)$ ensures that NAC/NAM holds for the firms that choose $\sigma = 1$. Therefore, strict submodularity is sufficient for NAC/NAM to hold globally.

Sufficient Condition for PAC/PAM. Suppose now that $f(x, y)$ is log-supermodular, which then implies that productivity dispersion $\kappa(y)$ is increasing in y, with a minimum $\kappa(\underline{y})$ and a maximum $\kappa(\overline{y})$. Below we fix the endowment of agents and analyze how the equilibrium allocations and hence sorting change with c. Suppose initially that c is very large (for example $c \geq f(x_2, \overline{y})$) so that all firms choose $\sigma = 0$. In this case, PAC/PAM holds since f is strictly square-root supermodular: firms with $y > y^{EK}$ will choose x_2 workers and firms with $y < y^{EK}$ will choose x_1 workers. Furthermore, if $y \leq y^{EK}$, then $\Delta\pi(y)$ is given by equation (49a), so it is strictly increasing in y. If $y \geq y^{EK}$, then $\Delta\pi(y)$ is given by equation (49b), from which its monotonicity is not immediately clear. Suppose that in this case (all firms optimally choose $\sigma = 0$), the right derivative of $\Delta\pi(y)$ at point y^{EK} is strictly positive: $\Delta\pi'_+(y^{EK}) > 0$, then the maximum value of $\Delta\pi(y)$ must be reached at some point $y^* > y^{EK}$. Now define $c^* = \Delta\pi(y^*)$ and gradually change c around this point. How does this change the sorting patterns? As long as $c \geq c^*$, no firm is willing to invest in screening so the equilibrium allocation remains the same. When c is slightly below c^*, then firms with types sufficiently close to y^* will choose $\sigma = 1$. Note that the equilibrium market utilities U_1 and U_2 will change slightly, so that y^{EK} also changes only slightly. Thus, as before firms with y slightly above (the new) y^{EK} will choose $\sigma = 0$ and x_2 workers only, and firms with y sufficiently close to y^* will choose both x_1 and x_2 workers. Hence, PAM/PAC fails to hold for c slightly below c^*.

Thus given an endowment of agents, a necessary condition for PAM/PAC to hold for all c is that when all firms choose $\sigma = 0$, then at y^{EK}, the right derivative is

[36] Note that the first (resp. last) statement becomes void if $\Delta\pi(\underline{y}) \geq 0$ (resp. $\Delta\pi(\overline{y}) \geq 0$).

negative: $\Delta\pi'_+(y^{EK}) \leq 0$. Proposition 9 below shows that for any log-supermodular function $f(x, y)$, we can find an endowment of agents such that this condition fails and hence PAM/PAC fails for certain c. However, we also show that if the endowment satisfies a certain constraint, then PAC/PAM continues to hold for all c.

Given x_1 and x_2, with a slight abuse of notation we define $\rho(x_1, x_2, y)$ implicitly by

$$\frac{f_y(x_2, y)}{f_y(x_1, y)} = \left(\frac{f(x_2, y)}{f(x_1, y)}\right)^{\rho(x_1, x_2, y)} \tag{50}$$

By Lemma 1, $\rho(x_1, x_2, y) \in [\underline{\rho}, \overline{\rho}]$. Note that $\rho(x_1, x_2, y)$ is the discrete version of $\rho(x, y)$ defined in (1): When $x_1, x_2 \to x$, $\rho(x_1, x_2, y) \to \rho(x, y)$.

Below we will need the following function: Define

$$\Omega(\kappa) \equiv \frac{1}{2} + \frac{\ln(\sqrt{\kappa} + \sqrt{1+\kappa})}{\ln(1+\kappa)}. \tag{51}$$

Lemma 14. *$\Omega(\kappa)$ is strictly decreasing with $\lim_{\kappa \to 0} \Omega(\kappa) = \infty$ and $\lim_{\kappa \to \infty} \Omega(\kappa) = 1$.*

Proof. See Appendix A.21. □

Summary. The following proposition summarizes our results.

Proposition 9. *In our environment with endogenous screening, the following holds:*

(i) The market equilibrium exhibits NAC/NAM for any endowment of agents and any cost c if (resp. only if) $f(x, y)$ is strictly (resp. weakly) submodular.

(ii) Given any log-supermodular function f and $(x_1, x_2, J(y))$, if for some $y \in (\underline{y}, \overline{y})$,

$$\rho(x_1, x_2, y) < \Omega(\kappa(y)), \tag{52}$$

then we can find (ℓ_1, ℓ_2) and c such that PAC/PAM fails in equilibrium. On the other hand, if for all $y \in (\underline{y}, \overline{y})$, we have

$$\rho(x_1, x_2, y) \geq \Omega(\kappa(y)) \tag{53}$$

then the equilibrium must exhibit PAC/PAM for any (ℓ_1, ℓ_2) and c.

Proof. See Appendix A.22. □

When $x_1, x_2 \to x$, the left-hand side of (52) approaches $\rho(x, y)$, whereas the right-hand side approaches infinity since $\kappa(y) \to 0$. Thus we can certainly construct an endowment of agents such that (52) holds. On the other hand, for any CES production function with $\rho > 1$, (53) holds for sufficiently large $\kappa(y)$. That is, both (52) and (53) are not void.

One may have expected that whenever the production technology is log-supermodular that the most productive firms will have the strongest incentives to invest in screening and PAM/PAC always arises. This intuition turns out to be wrong because, to avoid congestion created by low-type workers, more productive firms can discourage low-type workers from applying by posting a sufficiently low wage for them (ex-ante screening). Thus in general firms in the middle have the strongest incentive to screen ex post. The issue is that some firms with types below those screening firms may be productive enough to hire high-type workers only (but not productive enough to pay the screening cost), and when that happens, PAM/PAC fail even when $f(x, y)$ is log-supermodular.

6 Conclusion

A firm with a vacancy can typically screen workers both ex ante and ex post. By announcing the terms of trade ex ante, it can discourage certain types from applying while ex post, after it observes the realized pool of applicants, it can select the most profitable worker. In this paper we show how the meeting and screening technology affects the interaction between ex post and ex ante screening, the chosen wage posting mechanisms and sorting patterns. Perhaps surprisingly, we find that if ex post screening is easier (firms can screen more applicants), this makes sorting harder. That is, stronger complementarities in the production technology are necessary to get positive assortative matching. The more workers a firm can screen, the stronger the incentives for high type workers to avoid ending up in the same pool of applicants and this is a force against sorting which is by itself efficient (a social planner also wants to reduce the probability that valuable resources are wasted because they end up in the same pool). An important implication of our mode is that identifying complementarities in the production function from observed sorting patterns in the labor market requires not only data on observed hirings but also on the entire pool of applicants.

Appendix A Proofs

A.1 Proof of Lemma 1

Consider the $\underline{\rho}$ case first. Taking the derivative of $\log f_y(x,y) - \underline{\rho} \log f(x,y)$ with respect to x gives

$$\frac{\partial}{\partial x}\left(\log f_y - \underline{\rho}\log f\right) = \frac{f_{xy}}{f_y} - \underline{\rho}\frac{f_x}{f} = \frac{f_{xy}f - \underline{\rho} f_x f_y}{f f_y} \geq 0,$$

where we have suppressed the arguments of $f(x,y)$ and its partial derivatives for simplicity.

Similarly,

$$\frac{\partial}{\partial x}\left(\log f_y - \overline{\rho}\log f\right) = \frac{f_{xy}}{f_y} - \overline{\rho}\frac{f_x}{f} = \frac{f_{xy}f - \overline{\rho} f_x f_y}{f f_y} \leq 0.$$

Thus $\log f_y(x_2,y) - \underline{\rho}\log f(x_2,y) \geq \log f_y(x_1,y) - \underline{\rho}\log f(x_1,y)$, and $\log f_y(x_2,y) - \overline{\rho}\log f(x_2,y) \geq \log f_y(x_1,y) - \overline{\rho}\log f(x_1,y)$, which together imply (2). □

A.2 Proof of Lemma 2

In our benchmark model, a firm's number of applicants n_A follows a geometric distribution with support $\mathbb{N}_0$ and mean λ, i.e., $\mathbb{P}\left[n_A \geq n \,|\, \lambda\right] = \left(\frac{\lambda}{1+\lambda}\right)^n$ for $n = 0, 1, 2, \ldots$. For use in Section 5.1, we will make this proof more general by allowing applications to be invariant, which Lester et al. (2015) and Cai et al. (2017) define as follows: applications are invariant if and only if the probability that a firm with queues (μ, λ) receives at least one high-type applicant depends only on μ (and not on λ). Hence, it equals the probability $\mathbb{P}\left[n_A \geq 1 \,|\, \mu\right]$ that a firm receives at least one applicant when the queue has length μ and consists only of high-type workers. That is,

$$\begin{aligned}
\mathbb{P}\left[n_A \geq 1 \,|\, \mu\right] &= 1 - \sum_{n=0}^{\infty} \mathbb{P}\left[n_A = n \,|\, \lambda\right]\left(1 - \frac{\mu}{\lambda}\right)^n \\
&= 1 - \sum_{n=0}^{\infty} \mathbb{P}\left[n_A \geq n \,|\, \lambda\right]\left(1 - \frac{\mu}{\lambda}\right)^n + \sum_{n=0}^{\infty} \mathbb{P}\left[n_A \geq n+1 \,|\, \lambda\right]\left(1 - \frac{\mu}{\lambda}\right)^n \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left[n_A \geq n \,|\, \lambda\right]\frac{\mu}{\lambda}\left(1 - \frac{\mu}{\lambda}\right)^{n-1}. \qquad (54)
\end{aligned}$$

where the first equality uses the fact that the probability that an applicant is high-type is μ/λ and is independent across applicants, while the second and the third equality follow from summation by parts. Note that Equation (54) holds both for the geometric and for the urn-ball technology.

Next, we consider a firm's interview capacity. Suppose that their *potential* number of interviews, n_C, follows a geometric distribution with support $\mathbb{N}_1$ and mean $(1-\sigma)^{-1}$. That is, $\mathbb{P}[n_C \geq n \,|\, \sigma] = \sigma^{n-1}$ for $n = 1, 2, \ldots$. Since interviewing might be constrained by the number of applications, the firm's *actual* number of interviews is $n_I = \min\{n_A, n_C\} \in \mathbb{N}_0$, distributed according to $\mathbb{P}\left[n_I \geq n \,|\, \lambda, \sigma\right] = \mathbb{P}\left[n_A \geq n \,|\, \lambda\right] \sigma^{n-1}$.

The probability that an interview reveals a high-type worker is μ/λ, and is independent across applicants. The probability that the firm interviews at least one high-type worker therefore equals

$$\phi(\mu,\lambda) = 1 - \sum_{n=0}^{\infty} \mathbb{P}\left[n_I = n|\lambda,\sigma\right] \left(1 - \frac{\mu}{\lambda}\right)^n = \sum_{n=1}^{\infty} \mathbb{P}\left[n_I \geq n|\lambda,\sigma\right] \frac{\mu}{\lambda} \left(1 - \frac{\mu}{\lambda}\right)^{n-1} \tag{55}$$

Substituting $\mathbb{P}\left[n_I \geq n \,|\, \lambda, \sigma\right] = \mathbb{P}\left[n_A \geq n \,|\, \lambda\right] \sigma^{n-1}$ yields

$$\begin{aligned}
\phi(\mu,\lambda) &= \frac{\mu}{\sigma\mu + (1-\sigma)\lambda} \sum_{n=1}^{\infty} \mathbb{P}\left[n_A \geq n \,|\, \lambda\right] \frac{\sigma\mu + (1-\sigma)\lambda}{\lambda} \left(1 - \frac{\sigma\mu + (1-\sigma)\lambda}{\lambda}\right)^{n-1} \\
&= \frac{\mu}{\sigma\mu + (1-\sigma)\lambda} \mathbb{P}\left[n_A \geq 1 \,|\, \sigma\mu + (1-\sigma)\lambda\right]. \qquad (56)
\end{aligned}$$

If a firm has a geometric number of applicants, as in our baseline model, then $\mathbb{P}\left[n_A \geq 1 \,|\, \lambda\right] = \lambda/(1+\lambda)$ and we obtain equation (7). In contrast, if the firm's number of applicants is Poisson, as in Section 5.1, then $\mathbb{P}\left[n_A \geq 1 \,|\, \lambda\right] = 1 - e^{-\lambda}$ and we obtain equation (38). $\square$

A.3 Proof of Lemma 3

This proof is based on Shimer (2005), but extends his result to arbitrary $\phi(\mu,\lambda)$. Because $\phi(\mu,\lambda)$ is concave in μ, we have

$$\psi_1\left(\mu^*,\lambda^*\right) \leq \phi_\mu(\mu^*,\lambda^*) \leq \psi_2\left(\mu^*,\lambda^*\right).$$

The above two inequalities can not hold as equalities simultaneously because that would imply that $\phi(\mu, \lambda^*)$ is linear for $\mu \in [0, \lambda^*]$, in which case the firm's problem can never be solved by an interior solution (see Section 5.2 for an extensive discussion of this case). By the above two inequalities, the wages must satisfy

$$w_1^* = \frac{U_1}{\psi_1(\mu^*, \lambda^*)} \geq \frac{U_1}{\phi_\mu(\mu^*, \lambda^*)} \quad \text{and} \quad w_2^* = \frac{U_2}{\psi_2(\mu^*, \lambda^*)} \leq \frac{U_2}{\phi_\mu(\mu^*, \lambda^*)}.$$

Moreover, the FOC of equation (11) with respect to μ implies

$$\phi_\mu(\mu^*, \lambda^*)(f(x_2, y) - f(x_1, y)) = U_2 - U_1.$$

Therefore,

$$w_2^* - w_1^* < \frac{U_2 - U_1}{\phi_\mu(\mu^*, \lambda^*)} = \frac{\phi_\mu(\mu^*, \lambda^*)(f(x_2, y) - f(x_1, y))}{\phi_\mu(\mu^*, \lambda^*)} = f(x_2, y) - f(x_1, y).$$

where the first inequality is strict because the two inequalities above can not hold as equalities simultaneously. □

A.4 Proof of Lemma 4

If $w_1 \leq U_1$ and $w_2 \leq U_2$, then apparently the only queue that is attracted is $(0, 0)$. If $w_1 \leq U_1$ and $w_2 > U_2$, then no x_1 workers will apply and x_2 workers will apply till the queue length is such that $w_2 m(\lambda)/\lambda = U_2$. Note that $m(\lambda)/\lambda = 1/(1+\lambda)$, which implies that the queue length is determined uniquely. The case $w_1 > U_1$ and $w_2 \leq U_2$ follows the same logic.

We are then led to consider the only case left: $w_1 > U_1$ and $w_2 > U_2$. We first consider necessary conditions if the wage menu attracts i) both x_1 and x_2 workers; ii) x_1 workers only; iii) x_2 workers only.

Consider case i) first. By equation (7) and (10), the market utility condition can be written as

$$\frac{1 + (1-\sigma)\lambda}{(1+\lambda)(1 + (1-\sigma+\sigma\zeta)\lambda)} w_1 = U_1 \tag{57}$$

$$\frac{1}{1 + (1-\sigma+\sigma\zeta)\lambda} w_2 = U_2 \tag{58}$$

From the second equation, we can solve λ in terms of ζ and then plug the value of λ into the first equation, which gives

$$\frac{w_1}{U_1} = \frac{w_2}{U_2}\left(1 + \frac{\left(\frac{w_2}{U_2} - 1\right)\sigma}{(1-\sigma)\frac{w_2}{U_2} + \zeta\sigma}\right) \tag{59}$$

Thus it is easy to see that the right-hand side of the above equation is strictly increasing in ζ, which then implies a unique solution for ζ and λ. Furthermore, it also implies that

$$\frac{(\frac{w_2}{U_2})^2}{(1-\sigma)\frac{w_2}{U_2} + \sigma} < \frac{w_1}{U_1} < \frac{\frac{w_2}{U_2} - \sigma}{1-\sigma} \tag{60}$$

where the term on the left (right) is obtained by setting $\zeta = 1$ ($\zeta = 0$) on the right-hand side of equation (59).

For case ii), we have $\zeta = 0$. Setting $\zeta = 0$ in equations (57) and (58) gives

$$\frac{1}{1+\lambda}w_1 = U_1$$
$$\frac{1}{1+(1-\sigma)\lambda}w_2 \leq U_2$$

where we have replaced "=" in equation (58) with "$\leq$" since x_2 workers choose not to visit the firm. We can solve the first equation for λ and then plug it into the second equation, which implies

$$\frac{\frac{w_2}{U_2} - \sigma}{1-\sigma} \leq \frac{w_1}{U_1} \tag{61}$$

For case iii), $\zeta = 1$ and we have

$$\frac{1+(1-\sigma)\lambda}{(1+\lambda)^2}w_1 \leq U_1$$
$$\frac{1}{1+\lambda}w_2 = U_2$$

Follow the same logic as above and we have

$$\frac{(\frac{w_2}{U_2})^2}{(1-\sigma)\frac{w_2}{U_2}+\sigma} \geq \frac{w_1}{U_1} \tag{62}$$

Therefore, cases i), ii), iii) imply a mutually exclusive relation between w_1/U_1 and w_2/U_2. For a given (w_1, w_2) with $w_1 > U_1$ and $w_2 > U_2$, there will be exactly one that holds. And within each case, the solution $(\zeta\lambda, \lambda)$ is also unique, as we have showed above. □

A.5 Proof of Lemma 5

Given $\phi_{\mu\mu} < 0$, the Hessian is negative definite if and only if its determinant is positive, i.e. $\Delta f \left[m''\phi_{\mu\mu}f^1 + \left(\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2\right)\Delta f\right] > 0$. Using $\Delta f > 0$ and the definition of $\kappa(y)$, this gives condition (16). □

A.6 Proof of Proposition 2

As $x_2 \to x_1 = x$, U_2 and U_1 will approach to the same value and both will be strictly larger than $\underline{U}$, which is the market utility of workers if all workers are the same type x and all firms are the same and have type $\underline{y}$, the infimum of firm types. Next, as $x_2 \to x_1$, the total queue length of each firm's choice will be bounded from above. The marginal contribution to surplus of a worker who applies to a firm of type y is $m'(\lambda)f(x, y)$. Define $\overline{\lambda}$ implicitly by $m'(\overline{\lambda})f(x, \overline{y}) = \underline{U}$, where $\overline{y}$ is the supremum of firm types. Then, we can restrict each firm's choice of queues to be in the convex set $\Delta \equiv \{(\mu, \lambda) \,|\, 0 \leq \mu \leq \lambda \leq \overline{\lambda}\}$.

Recall that a firm's SOC is $1 + \kappa(y)\left(\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}\right)/m'' > 0$. Since the above set Δ is compact, the factor $\left(\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}\right)/m''$ is also bounded, which implies that for sufficiently small $\kappa(y)$ or equivalently $x_2 - x_1$, (16) will hold for all (μ, λ) in the above set Δ, which implies that for each firm type y, the surplus function will be concave in this set. Therefore, by the theorem of the maximum, the firm's' solution $(\mu(y), \lambda(y))$ will be unique and continuous. Furthermore, when $0 < \mu(y) < \lambda(y)$, $\mu(y)$ and $\lambda(y)$, or equivalently $\zeta(y)$ and $\lambda(y)$, are jointly determined by the FOCs: (18) and (20). Hence by the implicit function theorem, they are both differentiable at that point. □

A.7 Proof of Lemma 7

Differentiating eq. (18) along the equilibrium path yields

$$\zeta'(y)(U_2 - U_1) = m' f_y^1 + m'' \lambda'(y) f^1 + (\zeta(y)\phi_\mu + \phi_\lambda)\Delta f_y \\ + \left[\zeta'(y)\phi_\mu + \zeta \frac{\partial \phi_\mu}{\partial \zeta}\zeta'(y) + \zeta \frac{\partial \phi_\mu}{\partial \lambda}\lambda'(y) + \frac{\partial \phi_\lambda}{\partial \zeta}\zeta'(y) + \frac{\partial \phi_\lambda}{\partial \lambda}\lambda'(y)\right]\Delta f,$$

where we have suppressed the arguments $\mu(y)$ and $\lambda(y)$ from the functions m and ϕ. By equation (20), we can substitute $\phi_\mu \Delta f$ for $U_2 - U_1$ on the left-hand side. The resulting equation and equation (28) are two linear equations in $\zeta'(y)$ and $\lambda'(y)$. A simple but tedious calculation then yields

$$\lambda'(y) = \frac{m' f_y^1 + \left(\phi_\lambda - \phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}}\right)\Delta f_y}{-m'' f^1 + \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\Delta f}. \tag{63}$$

Plugging equation (63) into equations (29) and (30) gives the desired result, where we use the definitions of $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$ from equations (26) and (27), respectively. □

A.8 Proof of Proposition 3

We first consider the PAC case. The other cases (PAM, NAC, and NAM) follow the same logic.

Suppose that (35) does not hold for $i = c$ and there exists x', y', μ', and λ' such that

$$\rho(x', y') < a^c(\mu', \lambda')$$

Then by continuity, we can assume that $0 < \mu' < \lambda'$ (note the strict inequality). Furthermore, by continuity there exists small $\epsilon_0 \in (0, (\lambda' - \mu')/2)$ such that the above inequality will hold for all $x \in [x' - \epsilon_0, x' + \epsilon_0]$, $y \in [y' - \epsilon_0, y' + \epsilon_0]$, $\mu \in [\mu' - \epsilon_0, \mu' + \epsilon_0]$, and $\lambda \in [\lambda' - \epsilon_0, \lambda' + \epsilon_0]$. Fix ϵ_0 from now on.

We set $x_1 = x'$, and $\ell_2 = \mu'$ and $\ell_1 = \lambda' - \mu'$, where ℓ_i is the total measure of x_i workers, $i = 1, 2$. Next, we pick x_2, $\underline{y}$, and $\overline{y}$ such that $x_2 - x' = y' - \underline{y} = \overline{y} - y'$. We denote this difference by ϵ_1 and let $\epsilon_1 \to 0$. The firm distribution $J(y)$ is uniform on $[\underline{y}, \overline{y}]$. For sufficiently small ϵ_1, by Proposition 2 the equilibrium $\lambda(y)$ is unique, continuous, and for all y, $\lambda(y) \in [\lambda' - \epsilon_0, \lambda' + \epsilon_0]$. Furthermore, $\mu(y)$ is continuous and the average $\mu(y)$ is μ' ($\int_{\underline{y}}^{\overline{y}} \mu(y) dJ(y) = \mu'$). Therefore, by continuity there exists some y_0 such that $\mu(y_0) = \mu'$. To sum up, at point y_0 we have $\mu' = \mu(y_0) < \lambda(y_0) \in (\lambda' - \epsilon_0, \lambda' + \epsilon_0)$. Also by Proposition 2, $\zeta(y)$ is differentiable at point y_0. Hence the assumptions of Lemma 7 are satisfied at point y_0.

When $\epsilon_1 \to 0$ ($x_2 \to x_1 = x'$), the left-hand side of eq. (31) approaches $\rho(x', y')$. For the right-hand side, $a^c(\mu(y_0), \lambda(y_0)) \to a^c(\mu', \lambda')$ (note the choice of y_0 depends on the value of ϵ_1). Furthermore, at point y_0 we have

$$1 - \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) \frac{\Delta f_y}{f_y^1} \approx 1 - \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) \frac{f_{xy}(x', y')}{f_y(x', y')} \epsilon_1 \to 1$$

where on the right-hand side we have suppressed arguments of $m(\lambda')$ and $\phi(\mu', \lambda')$. Similarly, the denominator on the right-hand side of eq. (31) also goes to 1. Therefore, for equation (31) to hold at point y_0, we need $\rho(x', y') \geq a^c(\mu', \lambda')$. We have thus reached a contradiction. □

A.9 On the sufficiency of (35) and (36) for points where $\zeta(y)$ is unique and interior.

Before we state and prove Proposition 10, we first present the following useful technical lemma. The first two parts are trivial, whereas the third part is non-trivial and critical for our results.

Lemma 15. *(i) If $\rho > 1$, then $((1+\kappa)^\rho - 1)/\kappa$ is strictly increasing for $\kappa > 0$; (ii) if $\rho \in (0,1)$, then $((1+\kappa)^\rho - 1)/\kappa$ is strictly decreasing for $\kappa > 0$; and (iii) if $\rho \in (0,1)$, then $((1+\kappa)^\rho - 1)/\kappa + \frac{1-\rho}{2}((1+\kappa)^\rho - 1)$ is strictly increasing for $\kappa > 0$.*

Proof. For (i) and (ii), define $g(\kappa) = (1+\kappa)^\rho$, then if $\rho \in (0,1)$, $g(\kappa)$ is strictly concave, and if $\rho > 1$, $g(\kappa)$ is strictly convex. Observe that $((1+\kappa)^\rho - 1)/\kappa = (g(\kappa) - g(0))/(\kappa - 0)$, which is strictly increasing in κ if $g(\kappa)$ is strictly convex, and it is strictly decreasing in κ if $g(\kappa)$ is strictly concave.

For (iii), by direct computation we have

$$\frac{d}{d\kappa} \left[\frac{(1+\kappa)^\rho - 1}{\kappa} + \frac{1-\rho}{2}((1+\kappa)^\rho - 1) \right] = \frac{2(1+\kappa)^{1-\rho} - 2 - \kappa(1-\rho)(2-\kappa\rho)}{2\kappa^2(1+\kappa)^{1-\rho}}$$

The derivative of the numerator on the right-hand side is $\frac{d}{d\kappa}[2(1+\kappa)^{1-\rho} - 2 - \kappa(1-\rho)(2-\kappa\rho)] = 2(1-\rho)[(1+\kappa)^{-\rho} - (1-\kappa\rho)]$. Since $(1+\kappa)^{-\rho}$ is convex in κ, $(1+\kappa)^{-\rho} - (1-\kappa\rho) > 0$, which implies that the numerator on the right-hand side of the above equation is strictly increasing in κ and hence strictly positive since at $\kappa = 0$, it is zero. Thus we have proved (iii). □

Proposition 10. *If for both $i = c$ and m and any μ and λ, we have*

$$\overline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \geq \max(1 - \overline{a}^i, 0) \tag{64}$$

then (35) implies (31).

Else, assume that (i) ϕ_λ/ϕ_μ is weakly decreasing in μ, (ii) $0 \leq \underline{a}^i < 1$, and (iii) for both $i = c$ and m and any μ and λ, we have

$$\underline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \leq 0 \tag{65}$$

then (36) implies that (31) holds with $\leq$ instead of $\geq$.

Proof. First we consider the case PAC/PAM where $\underline{\rho} \geq \overline{a}^i$. Since f is supermodular, $\Delta f_y \geq 0$ and the left-hand side of (31) is positive. We then prove a slightly stronger version of (31):

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq \overline{a}^i \frac{1 - \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) \frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \frac{\Delta f}{f^1}}$$

where we have replaced $a^i(\mu, \lambda)$ with its supreme $\overline{a}^i$. This is justified because if the second term on the right-hand side is negative, then we have nothing to prove; if it is positive, then we have a stronger version of the original inequality. Since the denominator on the right-hand side is positive because of firms' SOC, multiplying it on both sides and rearranging terms gives

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} + \frac{\Delta f_y}{f_y^1} \left(\overline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \right) \geq \overline{a}^i \tag{66}$$

If $\underline{\rho} \geq 1$, then

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq \frac{(1+\kappa(y))^{\underline{\rho}} - 1}{\kappa(y)} \geq \underline{\rho} \geq \overline{a}^i$$

where the first inequality is because of (2), and the second inequality is because of $\underline{\rho} \geq 1$. Since the second term on the left-hand side of (66) is positive because of (64), the above inequality then implies (66).

Now, suppose $\underline{\rho} \in (0,1]$. Recall we assume that $\underline{\rho} \geq \overline{a}^i$, which implies $\overline{a}^i \leq 1$. Note that the second term on the left-hand side of (66) is greater than $1 - \overline{a}^i$, which on its turn greater than $1 - \underline{\rho}$ because we assumed that $\underline{\rho} \geq \overline{a}^i$. Thus the left-hand side of (66) is greater than the following

$$\frac{(1+\kappa(y))^{\underline{\rho}} - 1}{\kappa(y)} + \left((1+\kappa(y))^{\underline{\rho}} - 1\right)(1 - \underline{\rho})$$

which, by part iii) of Lemma 15, reaches its minimum value $\underline{\rho}$ at $\kappa(y) = 0$. Hence we proved (66).

Next, we consider the NAC/NAM case. First note that by taking the derivative with respect to μ, ϕ_λ/ϕ_μ being weakly decreasing in μ is equivalent to $\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \geq 0$.

If $\Delta f_y \leq 0$, then the left-hand side of (31) is negative. The numerator on the right-hand side is positive because $\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \geq 0$, and the denominator is also positive because of the firm's SOC. Thus (31) holds with $\leq$.

Next, consider the case $\Delta f_y \geq 0$, which implies that $0 \leq \overline{\rho}$. Next, we prove the following inequality.

$$1 \leq \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right) \frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right) \frac{\Delta f}{f^1}} \tag{67}$$

which is equivalent to

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right) - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right) \leq 0. \tag{68}$$

Note that

$$\frac{f^1 \Delta f_y}{f^1_y \Delta f} \leq \frac{(1+\kappa(y))^{\overline{\rho}} - 1}{\kappa(y)} \leq \overline{\rho} \leq \underline{a}^i$$

where the first inequality follows from (2), and the second part ii) of Lemma 15, and the last from our assumption $\overline{\rho} \leq \underline{a}^i$. Thus by (65), we have (68) and hence (67). Thus

$$\frac{f^1 \Delta f_y}{f^1_y \Delta f} \leq \overline{\rho} \leq \underline{a}^i \leq a^i \leq a^i \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f^1_y}}{1 - \frac{1}{m''}\left(\frac{\phi^2_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}}$$

We have proved the second part of Proposition 10. □

A.10 Proof of Lemma 8

We first consider $a^c(\zeta, \lambda)$. Since $\phi(\mu, \lambda)$ is given by equation (7) and $a^c(\zeta, \lambda)$ is defined by equation (26), by direct calculation we get

$$a^c(\zeta, \lambda) = \frac{1+\lambda}{2\lambda}\left(1 + \frac{1}{1+(1-\sigma)\lambda} - \frac{2}{1+\sigma\zeta\lambda+(1-\sigma)\lambda}\right) \tag{69}$$

Note that $a^c(\zeta, \lambda)$ is strictly increasing in ζ. Thus for a given λ, we have

$$\max_\zeta a^c(\zeta, \lambda) = a^c(1, \lambda) = \frac{1}{2}\frac{1+\sigma+(1-\sigma)\lambda}{1+(1-\sigma)\lambda} \tag{70}$$

$$\min_\zeta a^c(\zeta, \lambda) = a^c(0, \lambda) = \frac{1}{2}\frac{(1-\sigma)(1+\lambda)}{1+(1-\sigma)\lambda} \tag{71}$$

Note that $a^c(0, \lambda) + a^c(1, \lambda) = 1$ and

$$\frac{da^c(1,\lambda)}{\lambda} = -\frac{\sigma(1-\sigma)}{2(1+(1-\sigma)\lambda)^2} < 0 \quad \text{and} \quad \frac{da^c(0,\lambda)}{\lambda} = \frac{\sigma(1-\sigma)}{2(1+(1-\sigma)\lambda)^2} > 0$$

Therefore, $a^c(1, \lambda)$ approaches its supreme when $\lambda \to 0$ and $a^c(0, \lambda)$ approaches its infimum when $\lambda \to 0$. Therefore, $\sup_{\zeta,\lambda} a^c(\zeta, \lambda) = \lim_{\lambda\to 0} a^c(1, \lambda) = (1+\sigma)/2$ and $\inf_{\zeta,\lambda} a^c(\zeta, \lambda) = \lim_{\lambda\to 0} a^c(0, \lambda) = (1-\sigma)/2$. Note that both the infimum and the supremem can not be reached because we require $\lambda > 0$.

Next we consider $a^m(\mu, \lambda)$. Similar as above, by direct computation we have

$$a^m(\zeta, \lambda) = \frac{\lambda(1+\lambda)(1-\sigma) + 2\sigma\zeta\lambda}{2\lambda\,(1+(1-\sigma)\lambda)} \tag{72}$$

Note that $a^m(\zeta, \lambda)$ is strictly increasing in ζ. For a given λ, $a^m(\zeta, \lambda)$ therefore reaches its minimum at $\zeta = 0$ and its maximum at $\zeta = 1$. Because $a^m(0, \lambda) = a^c(0, \lambda)$ and $a^m(1, \lambda) = a^m(1, \lambda)$, the rest of the proof is the same as for $a^c(\zeta, \lambda)$.

Finally, note that

$$a^c(\zeta, \lambda) - a^m(\zeta, \lambda) = \frac{\zeta(1-\zeta)\sigma^2\lambda}{(1+(1-\sigma)\lambda)(1+\sigma\zeta\lambda+(1-\sigma)\lambda)} \geq 0$$

Thus when $\sigma > 0$, $a^c(\zeta, \lambda) = a^m(\zeta, \lambda)$ if and only if $\zeta = 0$ or 1. □

A.11 Proof of Proposition 4

Our proof consists of two parts. First we show that assortative sorting holds at points where $\zeta(y)$ is unique and interior. By Proposition 10, that amounts to show that the regularity conditions in Proposition 10 are satisfied when $\phi(\mu, \lambda)$ is given by equation (7). Second, we show that the necessary and sufficient condition for assortative sorting to hold at points where $\zeta(y)$ is interior and unique is *sufficient* for it to hold at points where $\zeta(y)$ is a jump point.

Part I. We now verify that the regularity conditions in Proposition 10 are satisfied. By Lemma 8, $\overline{a}^i = (1+\sigma)/2$ and $\underline{a}^i = (1-\sigma)/2$ for both $i = c$ and m. Since $\phi(\mu, \lambda)$ is given by equation (7), by direct computation the left-hand side of (64) is

$$\frac{(1-\sigma)(1+\lambda)^2(2+(1-\sigma)\lambda)}{4(1+(1-\sigma)\lambda)(1+\sigma\mu+(1-\sigma)\lambda)} \geq \frac{(1-\sigma)(1+\lambda)(2+(1-\sigma)\lambda)}{4(1+(1-\sigma)\lambda)} \geq \frac{(1-\sigma)}{2}$$

where the first inequality is because the denominator reaches its maximum at $\mu = \lambda$, and the second inequality is because $1+\lambda \geq 1+(1-\sigma)\lambda$ and $2+(1-\sigma)\lambda \geq 2$. So we have proved (64).

Next, we verify the regularity conditions for NAC/NAM in Proposition 10. (i) Note that $\phi_\lambda/\phi_\mu = -(1-\sigma)\mu/(1+(1-\sigma)\lambda)$, which is decreasing in μ, (ii) $\underline{a}^i =$

$(1-\sigma)/2 \in [0,1)$, and (iii) by direct computation the left-hand side of (65) is

$$-\frac{\lambda(1+\lambda)^2(1-\sigma)^2}{4(1+(1-\sigma)\lambda)(1+\sigma\mu+(1-\sigma)\lambda)} \leq 0.$$

Part II). We now show that $\underline{\rho} \geq 1/2$ and $\overline{\rho} \leq (1-\sigma)/2$ is sufficient for PAC/PAM and NAC/NAM to hold respectively at firm types where $\zeta(y)$ is a jump point. Note that the former ($\underline{\rho} \geq 1/2$) is weaker than what we actually need ($\underline{\rho} \geq (1+\sigma)/2$).

Consider equations (32) and (33). If $\zeta_1 = 1$ (a corner solution), equation (33) does not need to hold, and we are back to the case of Eeckhout and Kircher (2010), who showed that when $\underline{\rho} \geq 1/2$, equation (32) implies (34), and when $\underline{\rho} \leq 1/2$, equation (32) implies (34) ($\leq$ replaced by $\geq$ in this case). The result of Eeckhout and Kircher (2010) is also restated and proved in Proposition 11.

If $\zeta_1 < 1$, then we also solve for $\kappa(y) \equiv \Delta f/f^1$ and λ_0 in terms of ζ_1 and λ_1 from equations (32) and (33). The solutions are

$$\kappa(y) = \frac{4\sigma(\lambda+\lambda\sigma(z-1)+1)^2}{(\lambda+1)(\lambda-\sigma+\lambda\sigma(z-1)+1)^2}, \tag{73}$$

$$\lambda_0 = -\frac{\lambda(\lambda+\sigma(-\lambda+(\lambda+2)z-1)+1)}{\sigma+\lambda(\sigma+\sigma z-1)-1} \tag{74}$$

Consider the case of PAC/PAM first. If $\underline{\rho} \leq 1/2$, then by (2), $\Delta f_y/f_y^1 \geq (1+\kappa(y))^{1/2} - 1$. Thus (34) holds if the following is positive.

$$(1+\kappa(y)) - \left(1+\frac{m(\lambda_0)-m(\lambda_1)}{\phi(\zeta_1\lambda_1,\lambda_1)}\right)^2 = \frac{4\lambda_1\sigma^3(1-\zeta_1)(1+\lambda_1(1-\sigma(1-\zeta_1)))}{(1+\lambda_1)^2(1-\sigma+\lambda_1(1-\sigma(1-\zeta_1)))^2} > 0$$

For the NAC/NAM case, (34) holds with $>$ if the following is negative.

$$\begin{aligned}(1+\kappa(y))^{\overline{\rho}} - \left(1+\frac{m(\lambda_0)-m(\lambda_1)}{\phi(\zeta_1\lambda_1,\lambda_1)}\right) &< \frac{1-\sigma}{2}\kappa(y) - \left(\frac{m(\lambda_0)-m(\lambda_1)}{\phi(\zeta_1\lambda_1,\lambda_1)}\right) \\ &= -\frac{2\sigma^2\lambda_1(1-\sigma(1-\zeta_1))(1+\lambda_1(1-\sigma(1-\zeta_1)))}{(1+\lambda_1)(1-\sigma+\lambda_1(1-\sigma(1-\zeta_1)))^2} \leq 0\end{aligned}$$

where the first inequality follows from $(1+\kappa)^{\overline{\rho}} < 1+\overline{\rho}\kappa \leq 1+\frac{1-\sigma}{2}\kappa$, and for the equality we used equations (73) and (74). $\square$

A.12 Proof of Lemma 9

Lemma 16. *When $\phi(\mu, \lambda)$ is given by (7) or (38), it satisfies Property A0, A1, A2 and A3.*

Proof. We first show that whenever $\phi(\mu, \lambda) = \frac{\mu}{\sigma\mu+(1-\sigma)\lambda} m(\sigma\mu + (1-\sigma)\lambda)$, where $m(\cdot)$ is strictly concave, then it satisfies properties A0, A1 and A2. Note that both (7) and (38) are special cases.

For A0, note $\phi_\mu(\lambda\zeta, \lambda) = \frac{1}{\zeta\sigma\chi-\sigma\chi+\chi}(\zeta\sigma\chi m'(\chi) - \sigma m(\chi) + m(\chi))$, where we set $\chi = \sigma\lambda\zeta + (1-\sigma)\lambda$. Thus ϕ_μ has the same sign as the term in the parenthesis, which is linear in σ. At $\sigma = 1$, $\phi_\mu = m'(\chi) > 0$, and at $\sigma = 0$, $\phi_\mu = m(\chi)/\chi > 0$. Thus ϕ_μ is always strictly positive. Similarly, $\frac{1}{\sigma}\phi_{\mu\mu}(\lambda\zeta, \lambda) = \frac{1}{\chi^2((\zeta-1)\sigma+1)}(\zeta\sigma\chi^2 m''(\chi) + 2(\sigma-1)(m(\chi) - \chi m'(\chi)))$. Again $\phi_{\mu\mu}$ has the same sign as the term in the parenthesis, which is linear in σ. At $\sigma = 1$, $\phi_{\mu\mu} = m''(\chi) < 0$, and at at $\sigma = 0$, $\phi_{\mu\mu} = -\frac{2}{\chi^2}(m(\chi) - \chi m'(\chi)) < 0$. Thus $\phi_{\mu\mu} < 0$.

For A1, note $\phi(\lambda\zeta, \lambda) = \frac{\zeta}{\sigma\zeta+(1-\sigma)} m(\sigma\lambda\zeta + (1-\sigma)\lambda)$, which is strictly concave in λ for $\zeta > 0$ because $m(\cdot)$ is strictly concave.

For A2, direct computation yields that

$$\frac{\partial\phi_\mu(\lambda\zeta, \lambda)}{\partial\lambda} = \zeta\phi_{\mu\mu}(\lambda\zeta, \lambda) + \phi_{\mu\lambda}(\lambda\zeta, \lambda) = -\frac{(1-\sigma)(m(\chi) - \chi m'(\chi))}{\chi^2} + \sigma\zeta m''(\chi)$$

where again we set $\chi = \lambda\zeta\sigma + (1-\sigma)\lambda$. At $\sigma = 1$, the above expression equals $\zeta m''(\chi) < 0$, and at $\sigma = 0$, it equals $-(m(\chi) - \chi m'(\chi))/\chi^2 < 0$. Since it is linear in σ, the above expression is always strictly negative.

For *A*3, we first verify the case of (7). Since ϕ is given by (7), we have

$$\mathcal{H}(\zeta, \lambda) = \frac{(\lambda+1)^3(\sigma-1)^2}{4\sigma(1+(1-\sigma)\lambda)((\zeta-1)\lambda\sigma+\lambda+1)} > 0 \tag{75}$$

By direct computation, we have

$$\frac{\partial\mathcal{H}(\zeta, \lambda)}{\partial\zeta} = \frac{\lambda(\lambda+1)^3(\sigma-1)^2}{4(\lambda(\sigma-1)-1)((\zeta-1)\lambda\sigma+\lambda+1)^2}$$

$$\frac{\partial\mathcal{H}(\zeta, \lambda)}{\partial\lambda} = -\frac{(\lambda+1)^2(\sigma-1)^2\left(-(\zeta-2)\left(\lambda^2-1\right)\sigma + (\zeta-1)(\lambda-2)\lambda\sigma^2 - (\lambda+1)^2\right)}{4\sigma(\lambda(-\sigma)+\lambda+1)^2((\zeta-1)\lambda\sigma+\lambda+1)^2}$$

Thus $-\frac{\partial\phi_\mu(\lambda z, \lambda)/\partial z}{\partial\phi_\mu(\lambda z, \lambda)/\partial\lambda} < -\frac{\partial H(\lambda z, \lambda)/\partial z}{\partial H(\lambda z, \lambda)/\partial\lambda}$: Property *A*3 holds trivially for (7).

Next, consider contact technology (38). The specific expression of $\mathcal{H}(\zeta,\lambda)$ is of limited value to us, but note that by Mathematica, $\lim_{\lambda\to 0}\mathcal{H}(\zeta,\lambda) = (1-\sigma)^2/4\sigma \geq 0$. Below we first prove that $\partial\mathcal{H}(\zeta,\lambda)/\partial\lambda > 0$ when $\sigma < 1$. By direct computation (Mathematica)

$$\frac{\partial\mathcal{H}(\zeta,\lambda)}{\partial\lambda} = -\frac{(s-1)^2\left(-\chi+e^\chi-1\right)e^{\lambda-\chi}}{\lambda(\lambda-\chi)\left(\lambda\left(2(s-1)e^\chi - s(\chi(\chi+2)+2)+2(\chi+1)\right)+s\chi^3\right)^2}\mathcal{T}(s)$$

where

$$\mathcal{T}(s) = \lambda^2\left(-\chi+e^\chi-1\right)\left(2(s-1)e^\chi - s(\chi(\chi+2)+2)+2(\chi+1)\right) - 2\lambda\left(-\chi(\chi+2)+2e^\chi-2\right)\left(-s\left(\chi^2\right.\right.$$

$$\mathcal{T}(1) = \chi^2(\lambda-\chi)\left(\lambda(\chi+1)-e^\chi(\lambda+\chi-4)-\chi(\chi+3)-4\right)$$

$$\mathcal{T}(0) = -2\lambda\left(-\chi+e^\chi-1\right)\left((\lambda-2)e^\chi-\lambda(\chi+1)+\chi(\chi+2)+2\right)$$

Next,

$$\partial H(\lambda z,\lambda)/\partial z - \partial H(\lambda z,\lambda)/\partial\lambda\frac{\partial\phi_\mu(\lambda z,\lambda)/\partial z}{\partial\phi_\mu(\lambda z,\lambda)/\partial\lambda} = -\frac{\lambda^2(s-1)^2\,(}{\chi(s\chi-\lambda)\left(\lambda\left(-s\left(\chi^2+\chi+1\right)+(s-1)e^\chi+\chi+1\right)+\right.}$$

where

$$\mathcal{T}(\lambda) = -\lambda\left(s^2(\chi(\chi(2\chi(\chi(\chi+3)+6)+15)+12)+6)-3(s-1)e^\chi(\chi+2)\left(s\left(\chi^2+\chi+2\right)-\chi-2\right)+\right.$$

with

$$\frac{1}{1-s}\mathcal{T}'(\lambda=\chi) = \mathcal{T}(s) \equiv -\left(2(s-1)e^{2\chi}(4\chi-3)+s(\chi(\chi(2\chi(\chi+2)+1)-4)-6)+e^\chi(-s(\chi(\chi(5\chi+\right.$$

$$\mathcal{T}(1) = -\chi^2\left((\chi-2)(2\chi+3)+e^\chi(6-5\chi)\right)$$

$$\mathcal{T}(0) = (-\chi + e^{\chi} - 1)(-\chi(5\chi + 2) + e^{\chi}(8\chi - 6) + 6)$$

At any point (z, λ) where $H(\lambda z, \lambda) > 0$, we have $\partial H(\lambda z, \lambda)/\partial \lambda > 0$ and

$$-\frac{\partial \phi_\mu(\lambda z, \lambda)/\partial z}{\partial \phi_\mu(\lambda z, \lambda)/\partial \lambda} < -\frac{\partial H(\lambda z, \lambda)/\partial z}{\partial H(\lambda z, \lambda)/\partial \lambda}. \tag{76}$$

□

A.13 Proof of Lemma 9

Proof. We first consider $a^c(\zeta, \lambda)$.

□

A.14 Proof of Proposition 5

Proof. TO BE ADDED □

A.15 Proof of Lemma 10

Since $\phi(\mu, \lambda) = \mu m(\lambda)/\lambda$, we have $\phi_\mu(\zeta\lambda, \lambda) = m(\lambda)/\lambda$, which then in turn implies that $\partial \phi_\mu(\zeta\lambda, \lambda)/\partial \zeta = 0$. Therefore, by equations (26) and (27), $a^c(\mu, \lambda) = a^m(\mu, \lambda) = \varepsilon_w(\mu, \lambda)\frac{m'(\lambda)}{\lambda m''(\lambda)}$ for any μ and λ. Next, we consider $\varepsilon_w(\mu, \lambda)$. Since $\phi_{\mu\mu}(\mu, \lambda) = 0$ and $\phi_{\mu\lambda}(\mu, \lambda) = (m(\lambda)/\lambda)' = (\lambda m'(\lambda) - m(\lambda))/\lambda^2$, by equation (25), we have

$$\varepsilon_w(\mu, \lambda) = \frac{\mu \phi_{\mu\mu}(\mu, \lambda) + \lambda \phi_{\mu\lambda}(\mu, \lambda)}{\phi_\mu(\mu, \lambda)} = \frac{\lambda m'(\lambda) - m(\lambda)}{m(\lambda)}.$$

This gives equation (40) for $a^c(\mu, \lambda)$ and $a^m(\mu, \lambda)$. A discussion of why the right-hand side in equation (40) equals the elasticity of substitution of the total number of matches in a submarket can be found in Eeckhout and Kircher (2010). □

A.16 Proof of Lemma 11

The desired expression for a^c follows readily from equations (26) and (41). To derive the expression for a^m, note that $\phi(\zeta\lambda, \lambda) = m(\zeta\lambda)$ implies that $\phi_\mu(\zeta\lambda, \lambda) = m'(\zeta\lambda)$

and $h(\zeta, \lambda) = m(\zeta\lambda)/m(\lambda)$. Therefore, the last factor in (27) can be rewritten as

$$1 - \frac{\partial\phi_\mu/\partial\zeta}{\partial\phi_\mu/\partial\lambda}\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta} = 1 - \frac{\lambda m''(\zeta\lambda)}{\zeta m''(\zeta\lambda)}\frac{\frac{\zeta m'(\zeta\lambda)m(\lambda) - m(\zeta\lambda)m'(\lambda)}{m(\lambda)^2}}{\lambda m'(\zeta\lambda)/m(\lambda)} = \frac{m(\zeta\lambda)m'(\lambda)}{\zeta m'(\zeta\lambda)m(\lambda)} = \frac{\epsilon_f(\lambda)}{\epsilon_f(\zeta\lambda)}.$$

By equation (41) and the concavity of $m(\mu)$, this expression is between 0 and 1, so $a^c(\zeta, \lambda) \geq a^m(\zeta, \lambda)$ for all ζ and λ. Therefore, $\overline{a}^c \geq \overline{a}^m \geq a^m(1, \lambda) = 1$. Next, consider $\underline{a}^c$ and $\underline{a}^m$. Since $m(\mu)$ is strictly concave and strictly increasing, $m'(\mu) > 0$ for any $\mu \geq 0$. Since $m(0) = 0$, by L'Hopital's rule we have $\lim_{\mu\to 0}\epsilon_f(\mu) = \lim_{\mu\to 0}\mu m'(\mu)/m(\mu) = 1$. Moreover, $\lim_{\mu\to 0}\epsilon_w(\mu) = \lim_{\mu\to 0}\mu m''(\mu)/m'(\mu) = 0 \cdot m''(0)/m'(0) = 0$ (recall $m'(0) > 0$). Thus $\lim_{\zeta\to 0} a^c(\zeta, \lambda) = \lim_{\zeta\to 0}\epsilon_w(\lambda\zeta)/\epsilon_w(\lambda) = 0$, and $\lim_{\zeta\to 0} a^m(\zeta, \lambda) = \lim_{\zeta\to 0}\epsilon_w(\lambda\zeta)/\epsilon_w(\lambda)\cdot\epsilon_f(\lambda)/\epsilon_f(\lambda\zeta) = \epsilon_w(0)/\epsilon_w(\lambda)\cdot\epsilon_f(\lambda)/\epsilon_f(0) = 0$. Since ϵ_f is always positive and ϵ_w is always negative, a^c and a^m are always positive. Thus we have $\underline{a}^c = \underline{a}^m = 0$. □

A.17 Proof of Proposition 7

The part of necessity follows from Proposition 3, where the proof is valid for any non-bilateral contact technology (ϕ is strictly concave in μ). Here we only need to prove the sufficiency part.

We first consider the PAC/PAM case. By assumption $\underline{\rho} \geq \overline{a}^i$, which implies that $\underline{\rho} \geq 1$, since for invariant contact technologies, $\overline{a}^i \geq 1$ (see Lemma 11). Consider the left-hand side of (**??**). We have

$$\frac{f^1\Delta f_y}{f_y^1\Delta f} \geq \frac{(1+\kappa(y))^{\underline{\rho}} - 1}{\kappa(y)} \geq \underline{\rho} \geq \overline{a}^i \geq a^i$$

where the first inequality is because of (2), and the second inequality is because that when $\underline{\rho} \geq 1$, $((1+\kappa)^{\underline{\rho}} - 1)/\kappa$ is increasing in κ so that $((1+\kappa)^{\underline{\rho}} - 1)/\kappa \geq \underline{\rho}$.

Next, consider the NAC/NAM case, where we have assumed that $\overline{\rho} \leq 0 = \underline{a}^c = \underline{a}^m$ (see Lemma 11). Again by (2), we have

$$\frac{f^1\Delta f_y}{f_y^1\Delta f} \leq \frac{(1+\kappa(y))^{\overline{\rho}} - 1}{\kappa(y)} \leq 0 = \underline{a}^i \leq a^i.$$

Thus we have proved our claim. □

A.18 Proof of Proposition 8

First we consider the unconditional probability that an applicant has a good signal. It is easy to see that

$$\mathbb{P}(\widetilde{x}_2) = \frac{\mu}{\lambda} + \frac{\lambda - \mu}{\lambda}(1-\tau)$$

If the applicant has a good signal $((\widetilde{x}_2))$, then the probability that he is high-type (x_2) is

$$\mathbb{P}(x_2 \,|\, \widetilde{x}_2) = \frac{\mathbb{P}(x_2)\mathbb{P}(\widetilde{x}_2 \,|\, x_2)}{\mathbb{P}(\widetilde{x}_2)} = \frac{\mu}{\mu + (1-\tau)(\lambda - \mu)}$$

where the first equality is simply the Bayes rule. From the above analysis, the queue length of applicants with a good signal is $\widetilde{\lambda} = \mu + (\lambda - \mu)(1-\tau)$.

Next, we consider the probability that the manager hires a high-type worker, $\phi(\mu, \lambda)$. For this we can ignore the existence of applicants with bad signals because they are low-type workers for sure and they will not affect the meeting process between firms and workers with good signals because geometric meeting technology is invariant. Now firms face a queue of length $\widetilde{\lambda}$, of which high-type workers have queue length μ. Therefore, by equation (7) the probability that firms hire a type-type worker is $\phi(\mu, \lambda) = \mu/(1 + \sigma\mu + (1-\sigma)\widetilde{\lambda})$, which is exactly the equation given in Proposition 8.

Alternatively, we can prove the same claim directly. $\phi(\mu, \lambda) = 1 - \sum_{n=0}^{\infty} P_n(\widetilde{\lambda}) \left(1 - \frac{\mu}{\widetilde{\lambda}}\right)^n$, where $P_n(\widetilde{\lambda})$ is given by equation (**??**). □

A.19 Proof of Lemma 12

Note that $\pi_1(y) = \pi_2(y) > 0$ means that $f^1 > U_1$, $f^2 > U_2$, and $\sqrt{f^2} - \sqrt{f^1} = \sqrt{U_2} - \sqrt{U_1}$. The last expression can be rewritten as $\sqrt{f^2/f^1} - 1 = \sqrt{U_1/f^1}(\sqrt{U_2/U_1} - 1)$. Since $\sqrt{U_1/f^1} < 1$, it follows that $\sqrt{U_2/U_1} - 1 > \sqrt{f^2/f^1} - 1$, and thus $U_2/U_1 > f^2/f^1$. Similarly, we can rewrite $\sqrt{f^2} - \sqrt{f^1} = \sqrt{U_2} - \sqrt{U_1}$ as $(f^2 - f^1)/(\sqrt{f^2} + \sqrt{f^1}) = (U_2 - U_1)/(\sqrt{U_2} + \sqrt{U_1})$. Because $f^1 > U_1$ and $f^2 > U_2$, we have $\Delta f > \Delta U$.

Equation (49) then follows from substituting the relevant version of (44) into $\Delta\pi(y) = \overline{\pi}(y) - \max\{\pi_1(y), \pi_2(y)\}$. □

A.20 Proof of Lemma 13

For submodular f, we have $\pi_2(y) > \pi_1(y)$ if $y < y^{EK}$, and vice versa. Hence,

$$\Delta\pi'(y) = \begin{cases} -f_y^2\left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_2}{f^2}}\right) + f_y^1\left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_1}{f^1}}\right) & \text{if } y < y^{EK}, \quad (77a) \\ 2\left(\sqrt{\Delta f} - \sqrt{\Delta U}\right)\left(f_y^2 - f_y^1\right) & \text{if } y > y^{EK}. \quad (77b) \end{cases}$$

To establish the sign of (77a), note that (47) implies $f^2/f^1 < U_2/U_1$, which is equivalent to $\Delta U/\Delta f > U_1/f^1$ or $\Delta U/\Delta f > U_2/f^2$. The coefficient of f_y^2 in (77a) is therefore negative. Since f is submodular, $f_y^2 \leq f_y^1$, and we have

$$\Delta\pi'(y) \geq -f_y^1\left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_2}{f^2}}\right) + f_y^1\left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_1}{f^1}}\right) = f_y^1(\sqrt{U_2/f^2} - \sqrt{U_1/f^1}),$$

where the right-hand side is strictly positive because $U_2/U_1 > f^2/f^1$. Hence, $\Delta\pi'(y) > 0$ for $y < y^{EK}$, i.e., $\Delta\pi(y)$ is strictly increasing in y for $y \leq y^{EK}$.

To establish the sign of (77b), note that (47) implies $\Delta f > \Delta U$ and that $\Delta f_y \leq 0$ when f is weakly submodular. Hence, $\Delta\pi'(y) \leq 0$ for $y > y^{EK}$. When f is strictly submodular, then the above inequalities are strict so that $\Delta\pi(y)$ is strictly decreasing in y for $y \geq y^{EK}$. □

A.21 Proof of Lemma 14

By L'Hospital's Rule, $\lim_{\kappa\to 0}\Omega(\kappa) = \lim_{\kappa\to 0}\frac{1}{2} + \frac{1}{\sqrt{\kappa}+\sqrt{1+\kappa}}\left(\frac{1}{2\sqrt{\kappa}} + \frac{1}{2\sqrt{1+\kappa}}\right)(1+\kappa) = \infty$. In contrast, when $\kappa \to \infty$, we have $\kappa \approx 1+\kappa$ and $\lim_{\kappa\to\infty}\Omega(\kappa) = \lim_{\kappa\to\infty}\frac{1}{2} + \frac{\ln(\sqrt{\kappa}+\sqrt{\kappa})}{\ln(\kappa)} = 1$.

Next, we prove that $\Omega(\kappa)$ is strictly decreasing. By direct computation,

$$\Omega'(\kappa) = \frac{\ln(1+\kappa) - 2\sqrt{\frac{\kappa}{1+\kappa}}\ln(\sqrt{\kappa}+\sqrt{1+\kappa})}{4\sqrt{\kappa(1+\kappa)}\ln(1+\kappa)}.$$

The derivative of the numerator above is $-\ln(\sqrt{\kappa}+\sqrt{1+\kappa})\sqrt{\frac{1+\kappa}{\kappa}}(1+\kappa)^{-2} < 0$. At $\kappa = 0$, the numerator is zero, which implies that it is strictly negative and hence $\Omega'(\kappa) < 0$ when $\kappa > 0$. □

A.22 Proof of Proposition 9

NAC/NAM. As we mentioned before, for the NAC/NAM case, the necessity of submodularity of $f(x, y)$ follows from the limit case $c = 0$ (see Proposition 4). Next, we consider the sufficiency of strict submodularity of $f(x, y)$. We have already showed that when $y \leq y^{EK}$ and $f^2/f^1 < U_2/U_1$, $\Delta\pi(y)$ is strictly increasing. Furthermore, Lemma 12, $f^2/f^1 < U_2/U_1$ at $y = y^{EK}$. Since f is submodular and hence strictly log-submodular, f^2/f^1 is strictly decreasing. Hence, there exists at most one y' such that $f(x_2, y')/f(x_1, y') = U_2/U_1$. Firms of $y \leq y'$ will choose $\sigma = 0$ and x_2 workers; firms of $y \in [y', y^{EK}]$ will choose $\sigma = 1$ and both x_1 and x_2 workers.

Next, we consider $y \geq y^{EK}$ and $f^2 - f^1 > U_2 - U_1$. Since f is strictly submodular, $f^2 - f^1$ is strictly decreasing. Furthermore, at $y = y^{EK}$, $f^2 - f^1 > U_2 - U_1$ (see Lemma 12). Thus there exists at most one y'' such that $f(x_2, y'') - f(x_1, y'') = U_2 - U_1$. Firms of $y \in [y^{EK}, y'']$ will choose $\sigma = 1$ and both x_1 and x_2 workers; firms of $y \geq y''$ will choose $\sigma = 0$ and x_1 workers. Thus firms of $y \in [y', y'']$ all choose $\sigma = 1$, and we know from Proposition 4 that submodularity ensures that NAC/NAM holds for these firms. Therefore, NAC/NAM holds globally.

Note that we can not weaken the requirement of strict submodularity to mere submodularity for the sufficient condition. To see this, set $f(x, y) = x + y$. We initially set c large enough so that all firms choose $\sigma = 0$. Then for $y \geq y^{EK}$, $\pi_d(y) = \pi_{d,1}(y)$ is a constant by equation (**??**). We then decrease c till $c = \pi_d(y^{EK})$. Then all firms with $y \geq y^{EK}$ are indifferent between choosing $\sigma = 0$ with x_1 workers and $\sigma = 1$ with both x_1 and x_2 workers. This indeterminacy contradicts with NAC/NAM.

PAC/PAM. Next, we consider PAC/PAM. Assume that $f(x, y)$ is log-supermodular. Then $f^2 - f^1$ is strictly increasing, which implies that there exists at most one y' such that $f^2 - f^1 = U_2 - U_1$. When $y \leq y'$, $\pi_d(y) = 0$; when $y' \leq y \leq y^{EK}$, $\pi_d(y) = \pi_{d,1}(y)$, which is given by equation (**??**) and is strictly increasing in y. Next, we consider $y \geq y^{EK}$. Since f^2/f^1 is increasing, let $y'' = \sup\{y \mid f^2/f^1 = U_2/U_1\}$ (if this set is empty, then set $y'' = \infty$). Then for $y \geq y''$, $\pi_d(y) = 0$. If $y \in [y^{EK}, y'']$, $\pi_d(y) = \pi_{d,2}(y)$, which is given by equation (**??**).

Below we shorten $\rho(x_1, x_2, y)$ as $\rho(y)$ and define

$$\delta(y) \equiv \sqrt{\frac{\kappa(y)}{\Delta U/U_1}}. \tag{78}$$

For $y \in [y^{EK}, y'']$, $\pi'_{d,2}(y)$ given by equation (**??**) can be rewritten as

$$\begin{aligned}
\pi'_{d,2}(y) &= f_y^1 \sqrt{\frac{\Delta U}{\kappa(y) f^1}} \left[-(1+\kappa(y))^{\rho(y)} \left(1 - \sqrt{\frac{\kappa(y)}{1+\kappa(y)}} \sqrt{\frac{U_2}{\Delta U}} \right) + 1 - \sqrt{\frac{\kappa(y)}{\Delta U / U_1}} \right] \\
&= f_y^1 \sqrt{\frac{\Delta U}{\kappa(y) f^1}} (1+\kappa(y))^{\rho(y)-\frac{1}{2}} \sqrt{\kappa(y) + \delta(y)^2} - \left((1+\kappa(y))^{\rho(y)} - 1 + \delta(y) \right) \\
&= f_y^1 \sqrt{\frac{\Delta U}{\kappa(y) f^1}} \frac{(1+\kappa(y))^{2\rho(y)-1} \left(\kappa(y) + \delta(y)^2 \right) - \left((1+\kappa(y))^{\rho(y)} - 1 + \delta(y) \right)^2}{(1+\kappa(y))^{\rho(y)-\frac{1}{2}} \sqrt{\kappa(y) + \delta(y)^2} + \left((1+\kappa(y))^{\rho(y)} - 1 + \delta(y) \right)}.
\end{aligned}$$

Thus $\pi'_{d,2}(y)$ has the same sign as the numerator of the last term in the last line. Single out the numerator and define

$$\mathcal{S}(\delta, \kappa, \rho) = (1+\kappa)^{2\rho-1} \left(\kappa + \delta^2 \right) - \left((1+\kappa)^{\rho} - 1 + \delta \right)^2, \tag{79}$$

which is a quadratic function of δ with strictly positive second-order coefficient. Note that $\mathcal{S}(1, \kappa, \rho) = 0$ and $\frac{\partial \mathcal{S}(\delta,\kappa,\rho)}{\partial \delta}\big|_{\delta=1} = 2(1+\kappa)^{\rho}((1+\kappa)^{\rho-1} - 1) \geq 0$, which implies that the equation $\mathcal{S}(\delta, \kappa, \rho) = 0$ must have another root $\delta \leq 1$. Furthermore

$$\mathcal{S}(0, \kappa, \rho) = \kappa(1+\kappa)^{2\rho-1} - \left((1+\kappa)^{\rho} - 1 \right)^2,$$

Thus $\mathcal{S}(0, \kappa, \rho) \leq 0$ if and only if $\sqrt{\frac{\kappa}{1+\kappa}}(1+\kappa)^{\rho} \leq (1+\kappa)^{\rho} - 1$, or equivalently $\rho \geq \Omega(\kappa)$.

If for each $y \in (\underline{y}, \overline{y})$, $\rho(y) \geq \Omega(\kappa(y))$, then $\mathcal{S}(0, \kappa(y), \rho(y)) \leq 0$, which implies that $\mathcal{S}(\delta(y), \kappa(y), \rho(y)) < 0$ and hence $\pi'_{d,2}(y) < 0$. Therefore, $\pi_d(y) = 0$ if $y \leq y'$; it is strictly increasing if $y \in [y', y^{EK}]$; it is strictly decreasing if $y \in [y^{EK}, y'']$; it equals zero again when $y \geq y''$. In equilibrium, firms with $\pi_d(y) \geq c$ choose $\sigma = 1$. Similar to the NAC/NAM case, PAC/PAM holds in equilibrium.

Step 1: If for some $y^* \in (\underline{y}, \overline{y})$, $\rho(y^*) < \Omega(\kappa(y^*))$. Then $\mathcal{S}(0, \kappa(y^*), \rho(y^*)) > 0$ where $\mathcal{S}$ is defined above in equation (79). Thus by continuity, we can find δ^* small enough such that $\mathcal{S}(\delta^*, \kappa(y^*), \rho(y^*)) > 0$. Next, consider the following two equations with two unknowns (U_1, U_2),

$$\sqrt{f(x_2, y^*)} - \sqrt{f(x_1, y^*)} = \sqrt{U_2} - \sqrt{U_1}$$
$$\delta^* = \sqrt{\frac{(f(x_2, y^*) - f(x_1, y^*))/f(x_1, y^*)}{(U_2 - U_1)/U_1}}.$$

The first equation above is a reminiscent of equation (??) which defines y^{EK} and the second is a reminiscent of equation (78); in the latter two equations, we treat the market utilities as known and solve for y^{EK} and $\delta(y)$; here we treat y^* and δ^* as known and solve for market utilities. Denote the unique solution by (U_1^*, U_2^*).

Step 2: We assume that (U_1^*, U_2^*) are market utilities and solve for firms' problem. For $y \leq y^*$, then firms' problem is $\max_\lambda m(\lambda)f(x_1, y) - \lambda U_1^*$; For $y \geq y^*$, then firms' problem is $\max_\lambda m(\lambda)f(x_2, y) - \lambda U_2^*$. Denote the solution by $\lambda^*(y)$ for all y.

Step 3: Set $\ell_1 = \int_{\underline{y}}^{y^*} \lambda^*(y)dJ(y)$ and $\ell_2 = \int_{y^*}^{\overline{y}} \lambda^*(y)dJ(y)$. Then by construction, (U_1^*, U_2^*) are indeed market utilities and $y^* = y^{EK}$ for the particular equilibrium where all firms choose $\sigma = 0$. Furthermore, by construction $\pi'_{d,2}(y^{EK}) > 0$.

Step 4: We then decrease c till it is slightly lower than $\max \pi_d(y)$. Then firms with $y \leq y^{EK}$ will choose $\sigma = 0$ and $x_!$ workers; firms with y slightly above y^{EK} will choose $\sigma = 1$ and x_2 workers; there exist firms that are larger than y^* will choose $\sigma = 1$ and both x_1 ad x_2 workers. Therefore, PAC/PAM fails in equilibrium. □

Appendix B Derivation of Figure 2

Consider a firm of type y which chooses $\sigma = 1$. There are four possibilities regarding the firm's optimal applicant pool:

(i) *No applicants.* If $f(x_1, y) \leq U_1$ and $f(x_2, y) \leq U_2$, then the firm will not attract any applicants, such that $\overline{\pi}(y) = -c$.

(ii) *Only low-type applicants.* If $f(x_1, y) > U_1$ and $f(x_2, y) - f(x_1, y) \leq U_2 - U_1$, the firm will attract low-type workers, but not high-type workers as their marginal product is less than their marginal cost; in this case, $\overline{\pi}(y) = \pi_1(y) - c$.

(iii) *Only high-type applicants.* If $f(x_2, y) > U_2$ and $f(x_2, y)/f(x_1, y) \geq U_2/U_1$, the firm will attract only high-type workers since their relative productivity is higher than their relative cost; in this case, $\overline{\pi}(y) = \pi_2(y) - c$.

(iv) *Both types of applicants.* If $f(x_2, y) - f(x_1, y) > U_2 - U_1$ and $f(x_2, y)/f(x_1, y) <$

U_2/U_1, then the firm strictly prefers a mix of both types of workers in their application pool. By the FOCs, the optimal queue is given by $\mu = \sqrt{\Delta f/\Delta U} - 1$ and $\lambda = \sqrt{f^1/U_1} - 1$. In this case, $\overline{\pi}(y)$ is given by (46).

Appendix C Online Appendix

C.1 General Bilateral Contact Technologies

Under bilateral contact technologies, the firm's problem is solved by attracting either x_1 or x_2 workers but not both.[37] Therefore, we do not need need to consider the scenario where $Z(y)$ is unique and interior; we only need to consider jump points of $\zeta(y)$, where the fraction of high type workers can both be 0 and 1 but not any number in between. Denote such a jump point by y^{EK}. Then the indifference condition of a firm with type y^{EK}, equation (32), now becomes, (note that $\zeta_1 = 1$)

$$(m(\lambda_0) - \lambda_0 m'(\lambda_0))\, f(x_1, y) = (m(\lambda_1) - \lambda_1 m'(\lambda_1))\, f(x_2, y). \tag{80}$$

The condition for PAC/PAM, (34), now reduces to

$$m(\lambda_1)\frac{f_y(x_2, y)}{f_y(x_1, y)} > m(\lambda_0). \tag{81}$$

In other words, PAC/PAM requires that firm types above y^{EK} strictly prefer x_2 workers and firm types below y^{EK} strictly prefer x_1 workers. The following results provide the conditions for assortative sorting and are proved by Eeckhout and Kircher (2010) in a framework with a continuum of worker types. Here we derive the same results in a simple framework with two worker types.

Proposition 11. *[Eeckhout and Kircher, 2010] Suppose contacts are bilateral, i.e. $\phi(\mu, \lambda) = m(\lambda)\mu/\lambda$.*

The market equilibrium exhibits PAC/PAM for any endowment of workers and firms if (only if resp.) $\underline{\rho} > \overline{a}^c = \overline{a}^m = \overline{a}^{EK}$ ($\underline{\rho} \geq \overline{a}^c = \overline{a}^m = \overline{a}^{EK}$ resp.).

In contrast, the market equilibrium exhibits NAC/NAM for any endowment of workers and firms if (only if resp.) $\overline{\rho} < \underline{a}^c = \underline{a}^m = \underline{a}^{EK}$ ($\overline{\rho} \leq \underline{a}^c = \underline{a}^m = \underline{a}^{EK}$ resp.).

[37] Note that this result can be derived from firms' SOC: With bilateral contact technologies $\phi_{\mu\mu} = 0$, the right-hand side of (16) is zero and is never satisfied for an interior ζ.

Proof. Because Proposition 3 is derived under the assumption that $\phi(\mu,\lambda)$ is strictly concave in μ (the contact technology is not bilateral), in addition to the sufficiency part we also offer a proof for the necessity part.

Consider first the sufficiency part of PAC/PAM. In this case we need to show that when $\underline{\rho} > \overline{a}^{EK}$, equation (80) implies (81). To see this, by taking logs, we can rewrite equation (80) as follows.

$$\log\left(f^2\right) - \log\left(f^1\right) = \log\left(m(\lambda_0) - \lambda_0 m'(\lambda_0)\right) - \log\left(m(\lambda_1) - \lambda_1 m'(\lambda_1)\right)$$

which can then be rewritten as

$$\int_{x_1}^{x_2} \frac{f_x(x,y)}{f(x,y)}dx = \int_{\lambda_1}^{\lambda_0} \frac{-\lambda m''(\lambda)}{m(\lambda) - \lambda m'(\lambda)}d\lambda \tag{82}$$

Note that

$$\int_{x_1}^{x_2} \frac{f_{xy}(x,y)}{f_y(x,y)} - \underline{\rho}\frac{f_x(x,y)}{f(x,y)}dx > \int_{\lambda_1}^{\lambda_0} \frac{m'(\lambda)}{m(\lambda)} - \underline{\rho}\frac{-\lambda m''(\lambda)}{m(\lambda) - \lambda m'(\lambda)}d\lambda$$

because by the definition of $\underline{\rho}$, the integrand on the left-hand side is weakly positive, and by the assumption that $\underline{\rho} > \overline{a}^{EK}(\lambda)$, the integrand on the right-hand side is strictly negative. Combing the above inequalities with equation (82) yields

$$\int_{x_1}^{x_2} \frac{f_{xy}(x,y)}{f_y(x,y)}dx > \int_{\lambda_1}^{\lambda_0} \frac{m'(\lambda)}{m(\lambda)}d\lambda, \tag{83}$$

which is simply another way to write (81) (by the same logic that (82) is a rewrite of (80)).

Next consider the necessity of $\underline{\rho} \geq \overline{a}^{EK}$ for the PAC/PAM. Suppose otherwise that $\underline{\rho} < \overline{a}^{EK}$. Then there exists x^*, y^*, λ^* such that $\rho(x^*,y^*) < a^{EK}(\lambda^*)$, and by continuity there exists some $\varepsilon > 0$ such that this strict inequality holds for all (x,y,λ) with $\max(|x - x^*|, |y - y^*|, |\lambda - \lambda^*|) < \varepsilon$. Below we will choose an endowment of workers and firms such that NAC/NAM holds.

Step 1: Set $y = y^*$, by continuity we can find Δx and $\Delta\lambda$ such that $x_1 = x^* - \Delta x$, $x_2 = x^* + \Delta x$, $\lambda_0 = \lambda^* + \Delta\lambda$, and $\lambda_1 = \lambda^* - \Delta\lambda$ such that $\Delta x, \Delta\lambda < \varepsilon/2$ and equation (80) holds.

Step 2: Set the market utility of x_1 workers as $m'(\lambda_0)f(x_1,y)$ and the market utility of x_2 workers as $m'(\lambda_1)f(x_2,y)$.

Step 3: Define $\pi_i(y)$, $i=1,2$, as the maximum expected profit by attracting x_i workers only. That is, $\pi_i(y)=\max_{\lambda\geq 0} m(\lambda)f(x_i,y)-\lambda U_i$ (the same definition will also be used in Section 5.5). Then by construction, we have $\pi_1(y^*)=\pi_2(y^*)$ and $\pi_1'(y^*)=m(\lambda_0)f(x_1,y^*)>m(\lambda_1)f(x_2,y^*)=\pi_2'(y*)$. Thus by continuity we can find Δy small enough such that $\pi_1(y)<\pi_2(y)$ for $y\in[y^*-\Delta y,y^*)$ and $\pi_1(y)>\pi_2(y)$ for $y\in(y^*,y^*+\Delta y]$. Finally, for any firm type distribution $J(y)$ on $[y^*-\Delta y,y^*+\Delta y]$, we can pick the number of workers (ℓ_1,ℓ_2) such that firms' demand of labor equals the supply so that U_1 and U_2 are indeed the equilibrium market utilities of workers.

The case of NAC/NAM follows the same logic. □ □

For our benchmark technology with $\sigma=0$, $a^{EK}(\lambda)=1/2$ for any λ. Thus Proposition 11 implies that Proposition 3 gives the right answer for this special case (by continuity), even though it is derived under the assumption that the contact technology is *not* bilateral ($\sigma>0$) so that we can avoid the technical issue of division by zero ($\phi_{\mu\mu}(\mu,\lambda)$).[38]

C.2 Invariant technologies with N Worker Types

In the special case of invariant contact technologies, the firm's problem is concave so the solution (the optimal queue) is always unique. Below, we briefly extend our analysis of sorting to an arbitrary number of worker types for this case.

Suppose that there are N worker types: $x_1<x_2<\cdots<x_N$. Consider a firm of type y who faces a queue $(\mu_1,\mu_2,\ldots,\mu_N)$ where μ_i is the queue length of all workers with types higher or equal to x_i for $i=1,2,\ldots,N$.. Thus the queue length of x_i workers is $\mu_i-\mu_{i+1}$ with the convention that $\mu_{N+1}=0$. The probability that the firm meets at least one worker with a type higher than or equal to x_i is $m(\mu_i)$. The

[38]There is a small twist: In Proposition 3, when $\sigma>0$, the necessary and sufficient conditions for positive or negative sorting coincide exactly, because the supremeum and the infimum of $a^c(\zeta,\lambda)$ and $a^m(\zeta,\lambda)$ are never reached as maximum and minimum, respectively. However, when $\sigma=0$, $a^c(\zeta,\lambda)$, which coincides with $a^c(\mu,\lambda)$ and $a^{EK}(\lambda)$, is always a constant: 1/2. With CES production function $\rho=1/2$, all firms will be indifferent between hiring only x_1 and only x_2 workers. See equation (44) below. This indifference makes necessary the distinction between necessary and sufficient conditions for bilateral contact technologies.

expected surplus is then given by

$$S(\mu_1,\ldots,\mu_N,y) = m(\mu_1)f(x_1,y) + \sum_{i=2}^{N} m(\mu_i)[f(x_i,y) - f(x_{i-1},y)] \tag{84}$$

The above equation is a direct generalization of equation (8): with probability $m(\mu_1)$, the firm meets at least one worker, which generates at least surplus $f(x_1, y)$; with probability $m(\mu_2)$ the firm meets at least one worker with a type higher than or equal to x_2, which generates at least an additional surplus $f(x_2, y) - f(x_1, y)$, and so on. Note that $S(\mu_1,\ldots,\mu_N, y)$ is strictly concave in $(\mu_1,\ldots,\mu_N)$.

From equation (84) we can proceed as before. Adding one more x_i workers increase $\mu_1,\ldots,\mu_i$ simultaneously, which implies that the marginal contribution to surplus of x_i workers in the queue is $\sum_{k=1}^{i} m'(\mu_k)[f(x_k, y) - f(x_{k-1}, y)]$ for $i \geq 1$ with the convention $f(x_0, y) = 0$. Suppose that the optimal queue is $(\mu_1(y),\ldots,\mu_N(y))$, and to simplify the exposition, suppose that all types of workers are present in the queue, i.e., $\mu_1(y) > \mu_2(y) > \cdots > \mu_N(y)$. Then the firm's FOC with respect to x_i workers is

$$\sum_{k=1}^{i} m'(\mu_k(y))[f(x_k,y) - f(x_{k-1},y)] = U_i$$

where U_i is the market utility of x_i workers. With the convention $U_0 = 0$, the above equation then implies that for $i \geq 1$

$$m'(\mu_i(y))[f(x_i,y) - f(x_{i-1},y)] = U_i - U_{i-1} \tag{85}$$

Differentiaing the above equation along the equilibrium path as before yields

$$\frac{\mu_i'(y)}{\mu_i(y)} = \frac{m'(\mu_i(y))}{-\mu_i(y)m''(\mu_i(y))}\frac{f(x_i,y) - f(x_{i-1},y)}{f_y(x_i,y) - f_y(x_{i-1},y)}$$

where $i \geq 1$ (recall that $f(x_0, y) = 0$). From the above equation one can show that the condition for PAC/PAM is

$$\frac{f^1}{f^1_y}\frac{f(x_i,y) - f(x_{i-1},y)}{f_y(x_i,y) - f_y(x_{i-1},y)} \geq a^i\left(\frac{\mu_i(y)}{\mu_1(y)}, \mu_1(y)\right). \tag{86}$$

which is a generation of equation (**??**). Similarly, Proposition 7 and Corollary 2 continue to hold with N worker types.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Shimer, R. (1999). Holdups and efficiency with search frictions. *International Economic Review*, 40(4):827–851.

Albrecht, J. W., Cai, X., Gautier, P. A., and Vroman, S. B. (2019). Multiple applications, competing mechanisms, and market power. mimeo.

Albrecht, J. W., Gautier, P. A., and Vroman, S. B. (2006). Equilibrium directed search with multiple applications. *Review of Economic Studies*, 73(4):869–891.

Auster, S., Gottardi, P., and Wolthoff, R. (2020). Simultaneous search and adverse selection. mimeo.

Bagger, J. and Lentz, R. (2018). An empirical model of wage dispersion with sorting. *Review of Economic Studies*, forthcoming.

Bartolucci, C., Devicienti, F., and Monzón, I. (2018). Identifying sorting in practice. *American Economic Journal: Applied Economics*, 10(4):408–438.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 81(4):813–846.

Cai, X., Gautier, P., and Wolthoff, R. (2017). Search frictions, competing mechanisms and optimal market segmentation. *Journal of Economic Theory*, 169:453–473.

Cai, X., Gautier, P., and Wolthoff, R. (2018). Meetings and mechanisms. mimeo.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Costinot, A. (2009). An elementary theory of comparative advantage. *Econometrica*, 77(4):1165–1192.

Davis, S. J. and Samaniego de la Parra, B. (2017). Application flows. mimeo.

Eeckhout, J. and Kircher, P. (2010). Sorting and decentralized price competition. *Econometrica*, 78:539–574.

Eeckhout, J. and Kircher, P. (2011). Identifying sorting - in theory. *Review of Economic Studies*, 78(3):872–906.

Galenianos, M. and Kircher, P. (2009). Directed search with multiple job applications. *Journal of Economic Theory*, 114:445–471.

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

Gautier, P. A. and Teulings, C. N. (2015). Sorting and the output loss due to search frictions. *Journal of the European Economic Association*, 13(6):1136–1166.

Hagedorn, M., Law, T. H., and Manovskii, I. (2017). Identifying equilibrium models of labor market sorting. *Econometrica*, 85(1):29–65.

Hicks, J. (1932). *The Theory of Wages*. Macmillan, London.

Hicks, J. (1970). Elasticity of substitution again: Substitutes and complements. *Oxford Economic Papers*, 22(3):289–296.

Hoffman, M., Kahn, L. B., and Li, D. (2018). Discretion in hiring. *Quarterly Journal of Economics*, 133(2):765–800.

Håkanson, C., Lindqvist, E., and Vlachos, J. (2018). Firms and skills: The evolution of worker sorting. mimeo.

Kircher, P. (2009). Efficiency of simultaneous search. *Journal of Political Economy*, 117:861–913.

Lester, B., Visschers, L., and Wolthoff, R. (2015). Meeting technologies and optimal trading mechanisms in competitive search markets. *Journal of Economic Theory*, 155:1–15.

Lindenlaub, I. (2017). Sorting multidimensional types: Theory and application. *Review of Economic Studies*, 84(2):718–789.

Lise, J., Meghir, C., and Robin, J.-M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19(1):63–87.

Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy*, 126(1):313–346.

McAfee, R. P. (1993). Mechanism design by competing sellers. *Econometrica*, 61(6):pp. 1281–1312.

Milgrom, P. and Segal, I. (2002). Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601.

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

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Shapley, L. and Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, 1(1):111–130.

Shi, S. (2001). Frictional assignment I: Efficiency. *Journal of Economic Theory*, 98:232–260.

Shimer, R. (2005). The assignment of workers to jobs in an economy with coordination frictions. *Journal of Political Economy*, 113(5):996–1025.

Shimer, R. and Smith, L. (2000). Assortative matching and search. *Econometrica*, 68(2):343–369.

Song, J., Price, D., Guvenen, F., Bloom, N., and von Wachter, T. (2018). Firming up inequality. mimeo.

Tinbergen, J. (1956). On the theory of income distribution. *Weltwirtschaftliches Archiv*, 77:155–175.

Wolthoff, R. P. (2018). Applications and interviews: Firms' recruiting decisions in a frictional labor market. *Review of Economic Studies*, 85(2):1314–1351.