Search, Screening and Sorting*

Xiaoming Cai† Pieter Gautier‡ Ronald Wolthoff§

March 25, 2019

Preliminary & Incomplete

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 thank Jim Albrecht, 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.

†Tongji University. E-mail: xiaoming@tongji.edu.cn.

‡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 (2010a) 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 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

[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 (2010a) use buyer/seller terminology, the same idea applies.

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 (2010a) (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 (2010a) are very general in terms of the production technology but only allow firms to meet one worker at a time.We allow firms to conduct simultaneous interviews and show that this creates an additional force that goes against PAM. For 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 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 it makes them more attractive for high-skilled workers. Increasing the hiring probability for high-skilled workers is the cheapest way to give 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. When we *increase* the screening capacity of firms, high-type firms will have less incentives to discourage low-killed

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

workers from applying because they do not have to select them and they do not crowd out high-skilled workers. Moreover, low-type firms have no incentives to discourage high-skilled workers from applying so it will be harder to sustain PAM. It is also more difficult to sustain NAM the better the screening technology is because it gives high-type workers an incentive to spread out and avoid each other's company in the queues of applicants.

We consider two different frictions in the interview process. Besides the one described above that firms can screen a limited amount of workers, we also consider the quality of the screening process. Specifically, we allow for a pre-screening stage where the firm receives noisy signals of their candidates. It then selects the most promising candidate and learns this worker's true type after which this worker starts producing at the promised wage. Those frictions in the interview process do not only affect who gets hired at a particular firm from a realized pool of candidates, they also affect the optimal wage mechanisms and the application strategies of workers in equilibrium. For example, when low types negatively affect the probability that a firm can hire a high type and when productivity differences between types are large, firms will discourage low types from applying by offering them relatively low wages. Similarly when interview frictions are low, firms prefer to screen ex post by posting mechanisms that encourage all workers to apply. This gives incentives to high-skilled workers, on turn to avoid each others company in the application pools. Our model takes all of this into account.

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 discuss how parameters of the production function can be identified for a given meeting technology.[6] 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).

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).[7] Håkanson 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

[5] See Abowd et al. (1999), Gautier and Teulings (2006, 2015) , Eeckhout and Kircher (2011), Lise et al. (2016), Hagedorn et al. (2017), Lopes de Melo (2018), Bartolucci et al. (2018), and Bagger and Lentz (2018).

[6] Gautier and Teulings (2006) and subsequent papers 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.

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

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. This section also derives our main results regarding sorting and shows that the optimal mechanism can be implemented by wage menus. In Section 4, we presents results for general production and recruiting technologies and an arbitrary number of worker types. Finally, section 5 concludes.

2 Model

2.1 Environment

Agents. A static economy is populated by a continuum of risk-neutral firms and workers. Both types of agents are heterogeneous. In particular, each firm is characterized by a type $y \in \mathcal{Y} \equiv [\underline{y}, \overline{y}] \subset \mathbb{R}_+$. The measure of firms with types less or equal to y is denoted by $J(y)$, which we assume to be continuous and strictly increasing on $\mathcal{Y}$. Similarly, each worker is characterized by a type $x \in \mathcal{X} \equiv [\underline{x}, \overline{x}] \subset \mathbb{R}_+$. We initially assume that there are $I = 2$ types of workers, i.e. a low type $x_1 \in \mathcal{X}$ and a high type $x_2 \in \mathcal{X}$.The aggregate measure of workers with type x_i is denoted by $\ell_i > 0$. At the beginning of time, agents' types are private information, but firms can learn workers' types by interviewing them, as we describe in more detail below.

Search. Each firm demands and each worker supplies a single unit of indivisible labor. To attract applicants, each firm commits to a mechanism c from a Borel-measurable mechanism space $\mathcal{C}$. In its most general form, a mechanism specifies an extensive form game that determines i) whom the firm will hire and ii) what transfers will take place, as a function of the firm's applicant pool. We make two key assumptions about $\mathcal{C}$. First, as common in the literature (see e.g. Eeckhout and Kircher, 2010b), we restrict $\mathcal{C}$ by abstracting from mechanisms that are contingent on either workers' identities (as opposed to their types) or other mechanisms that are present in the market.[8] Second, we assume that $\mathcal{C}$ at least includes the set of all wage menus $\boldsymbol{w} = (w_1, \dots, w_I)$, where the firm hires the most profitable interviewee and pays him w_i if his type is x_i.

Workers observe these mechanisms and choose to which one they wish to apply, taking into account that they may face larger competition from other workers at mechanisms that

[8] Mechanisms that condition on identities violate the frictional nature of our environment. Mechanisms that condition on other mechanisms could introduce additional equilibria (see Epstein and Peters, 1999).

offer better terms of trade.[9] We capture the anonymity of the large market with the standard assumption that identical workers must use symmetric strategies (see e.g. Shimer, 2005). All firms and workers attempting to match at a particular mechanism c are said to form a *submarket*.

Applications and Interviews. We initially focus on the following microfoundation of the frictional interaction between workers and firms.[10] Workers and firms *in a given submarket* live on the boundary of a circle, where they are randomly positioned according to a uniform distribution. Workers send their application to the firm that is nearest in a clockwise direction.[11] After receiving all applications, firms start interviewing their applicants in a random order.[12] 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; with complementary probability, the interviewing process stops. This setup allows us to interpret σ as a measure of how easy it is for firms to interview applicants: if $\sigma = 0$, each firm can interview only a single applicant, which is a special case of the bilateral model of Eeckhout and Kircher (2010a), while a firm can interview all its applicants if $\sigma = 1$.

Matching and Production. After the interviews have been concluded, matching takes place and payoffs are realized as specified by the mechanism. We assume that interviewing a worker is a necessary condition for hiring him.[13] A match between a worker of type x and a firm type of y produces output $f(x, y)$, which we assume to be strictly positive, strictly increasing, and twice continuously differentiable for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. For our analysis, a key characteristic of the production function is its *elasticity of complementarity* (Hicks, 1932, 1970), which is defined as

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

with extrema $\overline{\rho} = \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x, y)$ and $\underline{\rho} = \inf_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x, y)$.

This elasticity is closely related to the degree of supermodularity of the production function, which we define in the same way as Eeckhout and Kircher (2010a).

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

[10] We will consider a wide class of alternatives in section 4.

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

[12] In section 4.1, we describe an extension in which firms select interviewees based on noisy signals.

[13] 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.

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

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

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 (1) *is reversed, $f(x,y)$ is said to be* nowhere n-root-supermodular.

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

We will sometimes illustrate our results with a CES production function, as it has a constant elasticity of complementarity $\rho(x,y) = \rho$. That is, for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$,

$$f(x,y) = (x^{1-\rho} + y^{1-\rho})^{\frac{1}{1-\rho}}. \tag{2}$$

It is straightforward to verify that this production function is submodular (i.e. nowhere supermodular) when $\rho \leq 0$, $\frac{1}{1-\rho}$-root-supermodular when $0 < \rho < 1$, and log-supermodular when $\rho \geq 1$.[14]

Queue Lengths and Expected Payoffs. An important role in our analysis is played by the *queue length* $q_i(c)$, which we define as the ratio of the number of workers with type *weakly exceeding* x_i to the number of firms in a given submarket c. To formally derive the queue lengths $\boldsymbol{q}(c) \equiv (q_1(c), q_2(c))$, let $G(C,y)$ denote the probability that a firm with type y offers a mechanism in the set $C \subseteq \mathcal{C}$, where $\mathcal{C}$ is the set of all potential mechanisms. Define $H^f(C)$ to be the measure of firms that post mechanisms in the set C. An accounting identity implies that $H^f(C) = \int_{\mathcal{Y}} G(C,y)\, dJ(y)$ for each $C \subseteq \mathcal{C}$. Similarly, let $H_i^w(C)$ denote the measure of workers of type x_i that visit a mechanism in the set $C \subseteq \mathcal{C}$, satisfying $H_i^w(\mathcal{C}) \leq \ell_i$.[15] To capture the idea that workers cannot visit mechanisms that are not actually offered, H_i^w must be absolutely continuous with respect to H^f. For any mechanism c on the support of

[14] We will define sorting in terms of first-order stochastic dominance. Monotonic transformations of x or y will therefore not change our results. That is, if we measure worker types by a new variable $\alpha(x)$ and firm types by a new variable $\beta(y)$, where α and β are strictly increasing functions, all conclusions remain the same. This simple observation shows that (2) is less restrictive than it may appear at first. For example, letting $\rho \to 0$ gives $f(x,y) = \sqrt{xy}$. Setting $\alpha(x) = x^2$ and $\beta(y) = y^2$ delivers a new production function xy. For our theory of sorting, the two production functions are equivalent.

[15] The inequality reflects the idea that workers may choose not to participate in the market.

H^f, the queue length $q_i(c)$ is then defined by the following Radon-Nikodym derivative:[16]

$$q_i(c) = \frac{d\sum_{j\geq i} H_j^w(c)}{dH^f(c)}. \tag{3}$$

For a mechanism c that is not in the support of H^f, the Radon-Nikodym derivative is arbitrary. Yet, a firm of type y has to form beliefs about the queue $\boldsymbol{q}(c)$ that such a mechanism would attract in order to calculate its expected payoff, which we denote by $\pi(c, \boldsymbol{q}, y)$. 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, 2010a).

To formally state this condition, define $V_i(c, \boldsymbol{q})$ as the expected payoff of a worker with type x_i who applies to a mechanism c with queue $\boldsymbol{q}$. Further, define the *market utility* U_i of a worker of type x_i as the maximum expected payoff that he can obtain in equilibrium, either by visiting one of the submarkets or by being inactive. That is,

$$U_i = \max\left\{\max_{c\in \operatorname{supp} H^f} V_i(c, \boldsymbol{q}(c)), 0\right\},$$

where $\boldsymbol{q}(c)$ is implied by equation (3). Worker optimality then requires that workers only visit submarket(s) that yield(s) them their market utility. Formally, for each $c \in \operatorname{supp} H^f$, we have $V_i(c, \boldsymbol{q}(c)) \leq U_i$, with equality if c is in the support of H_i^w. The market utility condition extends this idea to mechanisms outside the support of H^f by imposing that firms' beliefs regarding the queue that such a mechanism would attract must satisfy

$$V_i(c, \boldsymbol{q}) \leq U_i \quad \text{with equality if } q_i - q_{i+1} > 0 \text{, for any } i, \tag{4}$$

where $q_{I+1} = 0$ by convention. In most cases, there is a unique queue $\boldsymbol{q}$ that satisfies equation (4) for a given mechanism c (see Cai et al., 2018, for a detailed discussion). If the solution to (4) is not unique, then the standard assumption in the literature is that firms are optimistic and expect the solution that maximizes their expected payoff $\pi(c, \boldsymbol{q}, y)$. We follow that approach when necessary.

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

Definition 2. *A directed search equilibrium is a pair $(G, \{H_i^w\})$ of strategies with the following properties:*

[16] On the support of H^f, the Radon-Nikodym derivative is unique almost everywhere. In case of multiplicity, we following Eeckhout and Kircher (2010a) and assume a rule that selects a unique Radon-Nikodym derivative to have well-defined payoffs.

1. *Each mechanism c in the support of $G(\cdot, y)$ maximizes $\pi(c, \boldsymbol{q}(c), y)$, where $\boldsymbol{q}(c)$ is determined by equation* (4)*.*

2. *For each worker type x_i, every c in the support of H_i^w maximizes $V_i(c, \boldsymbol{q}(c))$. Workers with type x_i will choose inactivity if their expected payoff $V_i(c, \boldsymbol{q}(c))$ is negative for any c in the support of H^f.*

3. *Market clearing: $\int_{\mathcal{C}} [q_i(c) - q_{i+1}(c)]\, dH^f(c) \leq \ell_i$, with equality if $U_i > 0$ for each $i = 1, \ldots, I$.*

3 Market Equilibrium

We start our equilibrium analysis by deriving an expression for the surplus created by a firm as a function of its queue length. Subsequently, we consider a relaxed version of firms' problem, in which firms can buy queues of low-type and high-type workers directly in a competitive market at prices equal to their respective market utilities. Afterwards, we show that this is without loss of generality since firms can achieve the same outcome by posting wage menus. Throughout, we apply a change of notation and define $\lambda = q_1$ as the firm's total queue length, and $\mu = q_2$ as its queue length of high-type applicants.

3.1 Surplus

Hiring Rule. To calculate the surplus created by a firm, we need to take a stance on its hiring decisions. We initially assume that hiring decisions are socially optimal, i.e. firms give priority to high-type workers over low-type workers. Later, we will show that this hiring rule is also privately optimal.

Interviewing Probability. Next, we follow the approach of Cai et al. (2018) and calculate $\phi(\mu, \lambda)$, which represents the probability that a firm interviews at least one high-type worker, if it has a queue μ of such workers and a queue $\lambda - \mu$ of low-type workers.

Lemma 1. *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*

$$\phi(\mu, \lambda) = \frac{\mu}{1 + \sigma\mu + (1 - \sigma)\lambda}. \tag{5}$$

Proof. See appendix A.1. $\square$

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 firms' matching probabilities. After all, under the socially optimal hiring rule, 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 (5) 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 proces as it describes how a firm's chances to match with a high-type worker change if the queue of low-type workers gets longer. 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. $\sigma = 1$).[17] In contrast, the partial derivative $\phi_\mu(\mu, \lambda)$ 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 he gets hired and increases surplus because he was the only high-type worker that was interviewed.[18]

Surplus. Given $\phi(\mu, \lambda)$, it is straightforward to calculate the surplus generated by a firm of type y with queues (μ, λ). The following lemma presents the result.

Lemma 2. *Under the socially optimal hiring rule, the surplus generated by a firm of type y with queues (μ, λ) equals*

$$S(\mu, \lambda, y) = m(\lambda) f(x_1, y) + \phi(\mu, \lambda)[f(x_2, y) - f(x_1, y)]. \tag{6}$$

Proof. See below. □

The interpretation of this expression is straightforward. The first term captures the fact that surplus equals at least $f(x_1, y)$ if the firm matches. This is of course a lower bound, because the worker's type may exceed x_1. The second term corrects for that: the firm hires a high-type worker with probability $\phi(\mu, \lambda)$, in which case an additional surplus $f(x_2, y) - f(x_1, y)$ is created.

For given queue lengths, $m(\lambda)$ is independent of σ. The extent to which firms can interview applicants therefore affects $S(\mu, \lambda, y)$ only through $\phi(\mu, \lambda)$, which is a strictly increasing function of σ. If $\sigma = 0$, as in most of the literature, then each applicant has the same probability of being hired, such that a match produces $f(x_i, y)$ with probability equal to the fraction of applicants that has type x_i. As σ increases, a firm becomes more likely to identify

[17] This property is known as *invariance* (see Lester et al., 2015; Cai et al., 2018).

[18] 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.

a high-type worker among its applicants, which shifts probability mass towards producing $f(x_2, y)$.[19]

Concavity. A complication in our analysis is that 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 is given by

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

where we write $f^1 \equiv f(x_1, y)$, $\Delta f = f(x_2, y) - f(x_1, y)$ and omit the arguments of the derivatives of $\phi(\mu, \lambda)$ and $m(\lambda)$ to simplify exposition.

In the bilateral case (i.e. $\sigma = 0$), we have $\phi_{\mu\mu} = 0$, which means that the Hessian is never negative definite and surplus is never strictly concave. In our analysis below, we will therefore focus on cases in which $\sigma > 0$, such that $\phi_{\mu\mu} < 0$; 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. Before we derive the relevant condition, first define a new variable which characterizes productivity dispersion, i.e.

$$\kappa(y) = \frac{\Delta f}{f^1}. \tag{7}$$

We can then establish the following result.

Lemma 3. *The surplus function $S(\mu, \lambda, y)$ is strictly concave at a point (μ, λ) with $0 < \mu < \lambda$ if and only if*

$$\kappa(y) < \frac{-m''}{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}}. \tag{8}$$

Hence, if $S(\mu, \lambda, y)$ is strictly concave at any point (μ, λ) with $0 < \mu < \lambda$, then

$$\kappa(y) < \overline{\kappa} \equiv \sup_{0<\mu<\lambda} \frac{-m''}{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}} = \frac{4\sigma}{(1-\sigma)^2}. \tag{9}$$

Proof. See appendix A.2. □

Since $m(\lambda)$ is always strictly concave in λ, the numerator in (8) is strictly positive. Further, it is easy to verify that the denominator is zero if and only if $\sigma = 1$ and strictly positive otherwise.[20] Hence, the concavity condition in (8) is always satisfied when either

[19] A change in σ will of course affect the equilibrium applicant pool itself, as we analyze in detail below.

[20] See the proof of the lemma for an explicit expression.

$\sigma = 1$ (firms can interview all workers) or $\kappa(y) \to 0$ (i.e. worker heterogeneity disappears). In contrast, the condition is violated for $\sigma < 1$ and large enough $\kappa(y)$; in fact, if $\kappa(y)$ exceeds the threshold $\overline{\kappa}$ defined in (9), concavity is violated for any (μ, λ). As discussed above, in the bilateral limit $\sigma = 0$, no positive $\kappa(y)$ can satisfy the concavity condition.

3.2 Firms' Relaxed Problem

After deriving surplus, we now begin our analysis of firms' relaxed problem, in which they buy queues of low-type and high-type workers directly in a competitive market at prices equal to their respective market utilities.

Firms' Payoff. This formulation of firms' problem implies that the payment of workers is sunk at the time of matching. As a result, firms give priority in hiring to high-type workers over low-type workers, which means that their privately optimal hiring rule coincides with the socially optimal one. Firms' payoff therefore equals the difference between surplus $S(\zeta\lambda, \lambda, y)$ as given by equation (6) and the cost of the queues. Rather than specifying firms' choice of queues in terms of (μ, λ), it will prove convenient to reformulate the problem in the following way: first, firms choose the fraction of high-type workers in their pool of applicants, which we denote by $\zeta \equiv \mu/\lambda \in [0, 1]$, and second, they choose the total queue length λ. In other words, the firm's problem is to select (ζ, λ) rather than (μ, λ). That is,

$$\max_{\zeta,\lambda} \quad \Pi(\zeta, \lambda, y) = S(\zeta\lambda, \lambda, y) - \lambda U_1 - \zeta\lambda (U_2 - U_1). \tag{10}$$

Optimal Queue Length. Working backwards, we first consider the choice of the queue length λ. For a given $\zeta \in [0, 1]$, the payoff $\Pi(\zeta, \lambda, y)$ is *strictly* concave in λ. Thus, assuming an interior solution, the following first-order condition with respect to λ determines a unique optimal queue length as a function of ζ and y, which we denote by $\lambda^o(\zeta, y)$:

$$U_1 + \zeta(U_2 - U_1) = m'(\lambda) f^1 + \frac{\partial \phi(\zeta\lambda, \lambda)}{\partial \lambda} \Delta f. \tag{11}$$

To understand (11), 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. For future use, differentiating equation (11) with respect to y while fixing ζ gives

$$\lambda^o_y(\zeta, y) = \frac{\partial \lambda^o(\zeta, y)}{\partial y} = -\frac{m' f^1_y + \frac{\partial \phi}{\partial \lambda} \Delta f_y}{m'' f^1 + \frac{\partial^2 \phi}{\partial \lambda^2} \Delta f}, \tag{12}$$

where we have suppressed arguments from $m(\lambda^o(\zeta, y))$ and $\phi(\zeta\lambda^o(\zeta, y), \lambda^o(\zeta, y))$. An important observation is that (12) only depends on $f(x, y)$ and $\phi(\mu, \lambda)$, and not on the market utilities. Since $\phi(\zeta\lambda, \lambda)$ is strictly increasing and concave for $\zeta > 0$, we have $\partial\phi(\zeta\lambda, \lambda)/\partial\lambda = \zeta\phi_\mu(\zeta\lambda, \lambda) + \phi_\lambda(\zeta\lambda, \lambda) > 0$ and similarly $\partial^2\phi(\zeta\lambda, \lambda)/\partial\lambda^2 = \zeta^2\phi_{\mu\mu} + 2\zeta\phi_{\mu\lambda} + \phi_{\lambda\lambda} < 0$. It is easy to see that the denominator on the right-hand side of the above is negative. Moreover, the numerator is positive if $\Delta f_y \geq 0$, which then implies $\partial\lambda^o(\zeta, y)/\partial y > 0$. 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.

Optimal Composition. Assuming that firms have solved for the optimal queue length $\lambda^o(\zeta, y)$, we next consider their choice of ζ. Define $\Pi^*(\zeta, y)$ as the payoff of a firm if it chooses ζ, taking into account how it affects $\lambda^o(\zeta, y)$. That is,

$$\Pi^*(\zeta, y) \equiv \Pi(\zeta\lambda^o(\zeta, y), \lambda^o(\zeta, y), y) = \max_{\lambda} \Pi(\zeta\lambda, \lambda, y). \tag{13}$$

In general, $\Pi^*(\zeta, y)$ is not necessarily concave in ζ, so firms' maximization problem may admit multiple solutions of ζ. Moreover, a solution may not be interior, i.e., $\zeta = 0$ or $\zeta = 1$. However, suppose that for a firm of type y, there is an *interior* solution $(\zeta(y), \lambda(y))$ where $0 < \zeta(y) < 1$ and $\lambda(y) = \lambda^o(\zeta(y), y)$. This solution must then satisfy the following first-order condition with respect to ζ.

$$\frac{\partial\Pi^*(\zeta, y)}{\partial\zeta}\Big|_{\zeta=\zeta(y)} = 0 \Leftrightarrow \phi_\mu(\zeta(y)\lambda(y), \lambda(y))\Delta f - (U_2 - U_1) = 0, \tag{14}$$

where we used the envelope theorem and treated the total queue λ as constant in this exercise. Hence, the only effect of an increase in ζ is that the firm's probability of matching with a high-type worker goes up, which increases surplus.[21] Finally, recall that by Lemma 3, an interior solution ζ which satisfies the first-order condition (11) also satisfies the second-order condition if equation (8) holds.

For the corner solutions, i.e. $\zeta(y) = 0$ or $\zeta = 1$, the following first-order conditions must be satisfied:

$$\zeta(y) = 0 \Rightarrow \frac{\partial\Pi^*(\zeta, y)}{\partial\zeta}\Big|_{\zeta=\zeta(y)} \leq 0 \tag{15}$$

$$\zeta(y) = 1 \Rightarrow \frac{\partial\Pi^*(\zeta, y)}{\partial\zeta}\Big|_{\zeta=\zeta(y)} \geq 0 \tag{16}$$

[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 in equation (15), $\lambda(y) = \lambda^o(0, y)$, which is implicitly defined by the equation $m'(\lambda)f^1 = U_1$, and similarly, in equation (16), $\lambda(y) = \lambda^o(1, y)$, which is implicitly defined by the equation $m'(\lambda)f^2 = U_2$,

For future use, notice that differentiating equation (11) with respect to ζ and evaluating the result at $\zeta = \zeta(y)$ gives

$$\lambda^o_\zeta(\zeta(y), y) = \frac{\partial \lambda^o(\zeta, y)}{\partial \zeta}\Big|_{\zeta=\zeta(y)} = -\frac{\lambda(y)\frac{\partial \phi_\mu}{\partial \lambda}\Delta f}{m'' f^1 + \frac{\partial^2 \phi}{\partial \lambda^2}\Delta f}, \quad \text{if } \zeta(y) \in (0, 1) \tag{17}$$

where we have used equation (14) to substitute out $U_2 - U_1$ and suppressed arguments from $m(\lambda(y))$ and $\phi(\zeta(y)\lambda(y), \lambda(y))$. The above equation 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.

Limit Case. In general, the first-order condition—i.e. equation (14), (15) or (16)—is necessary but not sufficient for the optimum. However, in the limit case in which worker heterogeneity disappears, i.e. $x_1, x_2 \to x$, the firms' problem becomes concave and the first-order condition becomes sufficient for the optimum. More precisely, in the proof of proposition 1, we show that there exists a large number $\overline{\lambda}$ such that when $x_2 \to x_1$, the Hessian matrix of all firms will be negative definitive in the set $\{(\mu, \lambda) \,|\, 0 < \mu < \lambda < \overline{\lambda}\}$ and firms' choice of queue will also lie in the set.

Proposition 1. *Fix $x_1 = x$, let $x_2 \to x$, and hold the endowments of workers and firms constant. Then for sufficiently small $x_2 - x_1$, each firm has a unique queue $(\mu(y), \lambda(y))$. Furthermore, both $\mu(y)$ and $\lambda(y)$ are continuous in y, and if for some point y_0, $0 < \mu(y_0) < \lambda(y_0)$, 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.3. □

The above proposition, although simple and intuitive, provides a useful tool for constructing examples of equilibrium allocations which exhibit positive or negative assortative contacting and matching later.

3.3 Sorting

In this subsection, we analyze whether the market equilibrium exhibits sorting. After presenting two inequalities on the production function and introducing two useful elasticities, we first provide our definitions of positive and negative sorting and then derive conditions on the production function that are necessary and sufficient for these outcomes. In the following, we focus on firms' optimal choice of ζ, since for each ζ, there will be a unique optimal queue length λ, which is determined by the first-order condition equation (11). We represent by $\mathrm{Z}(y)$ the set of optimal choices of ζ for a firm of type y. For the moment we assume that $\mathrm{Z}(y)$ contains only one element and as before, denote it by $\zeta(y)$. Later we will prove that with assortative contacting/matching this is the case for all y except at most one.

Elasticity of Complementarity Revisited. The elasticity of complementarity $\rho(x,y)$ measures the elasticity of relative marginal product with respect to relative product. To see this, note that

$$\begin{aligned}\frac{f_y(x+\Delta x,y)}{f_y(x,y)} &\approx 1+\frac{f_{xy}(x,y)}{f_y(x,y)}\Delta x = 1+\rho(x,y)\frac{f_x(x,y)}{f(x,y)}\Delta x \approx \left(1+\frac{f_x(x,y)}{f(x,y)}\Delta x\right)^{\rho(x,y)}\\ &= \left(\frac{f(x+\Delta x,y)}{f(x,y)}\right)^{\rho(x,y)}\end{aligned}$$

where we used the approximation sign "$\approx$" twice because when Δx is small, the first-order approximation is justified. In general, when x is discrete and $\rho(x,y)$ is not necessarily constant, we have the upper and the lower bound estimates for the elasticity of the relative marginal product with respect to the relative product. The result is given by the following.

Proposition 2. *For a given y, $f_y(x,y)/f(x,y)^{\underline{\rho}}$ is increasing in x, and $f(x,y)/f_y(x,y)^{\overline{\rho}}$ is decreasing in x. That is,*

$$\frac{f_y(x_2,y)}{f_y(x_1,y)} \geq \left(\frac{f(x_2,y)}{f(x_1,y)}\right)^{\underline{\rho}} \quad \text{and} \quad \frac{f_y(x_2,y)}{f_y(x_1,y)} \leq \left(\frac{f(x_2,y)}{f(x_1,y)}\right)^{\overline{\rho}} \tag{18}$$

Proof. See appendix A.4. □

From the proof of Proposition 2 we can see that for the two inequalities in (18) to hold, we need $\underline{\rho} = \rho(x,y)$ for all $x \in [x_1, x_2]$ for the former and $\overline{\rho} = \rho(x,y)$ for all $x \in [x_1, x_2]$ for the latter.

Two Key Elasticities. For our analysis, it is useful to define the following two elasticities.

$$\varepsilon_f(\mu,\lambda) \equiv \frac{\partial\phi(\zeta\lambda,\lambda)}{\partial\lambda}\frac{\lambda}{\phi(\zeta\lambda,\lambda)} = \frac{\mu\phi_\mu(\mu,\lambda)+\lambda\phi_\lambda(\mu,\lambda)}{\phi(\mu,\lambda)} \tag{19}$$

$$\varepsilon_w(\mu,\lambda) \equiv \frac{\partial\phi_\mu(\zeta\lambda,\lambda)}{\partial\lambda}\frac{\lambda}{\phi_\mu(\zeta\lambda,\lambda)} = \frac{\mu\phi_{\mu\mu}(\mu,\lambda)+\lambda\phi_{\mu\lambda}(\mu,\lambda)}{\phi_\mu(\mu,\lambda)} \tag{20}$$

where $\zeta \equiv \mu/\lambda$. The first elasticity is from the firm's perspective and measures how the probability that a firm meets a high-type worker is affected by a change in the queue length, holding the fraction of high types ζ constant. The second elasticity is from the workers' perspective and describes how the probability that all other applicants have lower types responds to a change in (a composition preserving) queue length λ.

Definition of Sorting. Much of the literature (see e.g. Becker, 1973; Shi, 2001; Eeckhout and Kircher, 2010a) defines sorting in terms of a monotonic matching function which maps a worker type x to a firm type y.[22] As demonstrated above, this definition is not suitable in our environment, because there is not a unique worker type that firms hire. In other words, 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 in firms' distributions of hires.[23] In the context of our environment with two types of workers, this definition can be expressed in terms of the probability that a firm hires a high-type worker, conditional on hiring someone, which is given by

$$h(\zeta(y), \lambda(y)) \equiv \frac{\phi(\zeta(y)\lambda(y), \lambda(y))}{m(\lambda(y))} \tag{21}$$

We therefore obtain the following definition.

Definition 3. *The market equilibrium exhibits* positive assortative matching *(PAM) if and only if $h(\zeta(y), \lambda(y))$ is (weakly) increasing in y. If $h(\zeta(y), \lambda(y))$ is (weakly) decreasing in y, the equilibrium is said to exhibit* negative assortative matching *(NAM).*

While the literature has traditionally restricted attention to sorting patterns in matches, our environment yields additional predictions. After all, given that firms with multiple candidates select the most desirable one, there is a meaningful distinction between an application or an interview (a "contact") on the one hand and a match on the other hand.[24] Hence, in addition to considering the assortativeness of matches, we can also analyze the assortativeness of contacts, for which we use the following definition.

Definition 4. *The market equilibrium exhibits positive assortative contacting (PAC) if and only if $\zeta(y)$ is weakly increasing in y. If $\zeta(y)$ is weakly decreasing in y, the equilibrium is said to exhibit* negative assortative contacting *(NAC).*

[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 first-order stochastic dominance of 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 first-order stochastic dominance of the distribution of hires as a "more easily testable" alternative.

[24] In our model, applications and interviews necessarily exhibit the same sorting patterns, so we analyze them together.

In words, PAC occurs if we observe weakly more high-type workers in the pool of applicants or interviewees as we increase y. If the reverse occurs, we have NAC.

Relation. To understand the relation between assortative contacting and matching, assume that $\zeta(y)$ is differentiable. Equation (11) and the implicit function theorem then imply that $\lambda(y)$ is differentiable as well. The chain rule of differentiation therefore yields

$$\frac{d}{dy}h(\zeta(y),\lambda(y)) = \zeta'(y)\underbrace{\frac{\partial}{\partial\zeta}h(\zeta(y),\lambda(y))}_{>0} + \lambda'(y)\underbrace{\frac{\partial}{\partial\lambda}h(\zeta(y),\lambda(y))}_{>0}. \tag{22}$$

The term over the first brace is positive because it is simply $\lambda\phi_\mu/m$; the term over the second brace is also positive.[25] We are thus led to consider $\lambda'(y)$. Along the equilibrium path, $\lambda(y) = \lambda^o(\zeta(y), y)$, which implies

$$\lambda'(y) = \zeta'(y)\lambda^o_\zeta + \lambda^o_y. \tag{23}$$

By equation (12), λ^o_y is positive along the equilibrium path if Δf_y is positive, which occurs if the production function f is supermodular.[26] That is, holding constant the fraction of high-type applicants ζ, firms with higher y will demand longer queues. However, from equation (17) , it follows that λ^o_ζ is negative. Thus, it is not a priori clear whether relatively more applications from high-type workers will lead to a higher relative probability of matching with high-type workers, i.e., whether PAC leads to PAM. However, we will show if PAC or PAM are to hold for *any distributions of worker and firm types*, then we obtain the exact same condition. The same is true for NAC and NAM.

Sorting Conditions. To derive necessary conditions for sorting, we first assume that the equilibrium solution $(\zeta(y), \lambda(y))$ is *differentiable* around a point y and $\zeta(y) \in (0,1)$ so that we can use standard tools from calculus to derive the relevant conditions for sorting (again, this assumption will be relaxed later). Differentiating equation (14) with respect to y gives

$$\zeta'(y) = -\frac{\phi_\mu}{\partial\phi_\mu/\partial\zeta}\frac{\Delta f_y}{\Delta f} - \frac{\partial\phi_\mu/\partial\lambda}{\partial\phi_\mu/\partial\zeta}\lambda'(y) \tag{24}$$

When $\zeta(y)$ is differentiable at a point y, PAC implies that $\zeta'(y) \geq 0$. Of course, a necessary condition for NAC is obtained by reversing the inequality. By equation (24), $\zeta'(y) \geq 0$ if and

[25] To see this, note that $\frac{d}{d\lambda}(\log\phi(\zeta\lambda,\lambda) - \log m(\lambda)) = \frac{\partial\phi(\zeta\lambda,\lambda)}{\partial\lambda}\frac{\lambda}{\phi(\zeta\lambda,\lambda)} - m'(\lambda)\frac{\lambda}{m(\lambda)} = (\varepsilon_f(\zeta\lambda,\lambda) - \varepsilon_f(\lambda,\lambda))/\lambda$. Given (5), it is easy to verify that $\varepsilon_f(\mu,\lambda)$ is strictly decreasing in μ.

[26] We will later show that the equilibrium allocation must exhibit both NAC and NAM if f is submodular.

only if

$$\frac{\Delta f_y}{\Delta f} \geq -\frac{\partial \phi_\mu / \partial \lambda}{\phi_\mu} \lambda'(y) = -\varepsilon_w(\zeta(y)\lambda(y), \lambda(y)) \frac{\lambda'(y)}{\lambda(y)} \tag{25}$$

where for the equality we used the definition of ε_w from equation (20). The effect of a higher y is captured by $\Delta f_y / \Delta f$. A firm with a higher y can affect the likelihood of a good match by changing the queue length and or composition. If the inequality in (25) holds, then the firm wants a high fraction of high-type applicants in its pool to increase the expected match value.

The analysis for PAM follows a similar path. Differentiating the function $h(\zeta(y), \lambda(y))$ with respect to y (along the equilibrium path) shows that $dh/dy \geq 0$ if and only if (see equation (22))

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

Next, we use equation (24) to substitute out $\zeta'(y)$ from the above equation, which yields

$$\frac{\Delta f_y}{\Delta f} \geq -\varepsilon_w(\zeta(y)\lambda(y), \lambda(y)) \frac{\lambda'(y)}{\lambda(y)} \cdot \left(1 - \frac{\partial \phi_\mu / \partial \zeta}{\partial \phi_\mu / \partial \lambda} \frac{\partial h / \partial \lambda}{\partial h / \partial \zeta}\right). \tag{26}$$

The interpretation of (26) is similar to that of the PAC case. Firms can invest in the expected match quality by choosing an appropriate expected pool of applicants. If we fix $h(\zeta(y), \lambda(y))$ around some y, then the firm must choose its queue composition and queue length according to, $d\zeta(y) = -\frac{\partial h / \partial \lambda}{\partial h / \partial \zeta} d\lambda(y)$, which induces the following percentage change of ϕ_μ,

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

This explains the additional term appearing in (26) compared to the condition for PAC. Again by equation (14), if (26) holds then a firm with a higher y will take advantage of its type by choosing a higher $h(\zeta(y), \lambda(y))$.

In the above conditions for PAC/PAM, (25) and (26), $\lambda'(y)$ is unknown. We then use equation (24) to substitute out $\zeta'(y)$ from equation (23), which implies

$$\lambda'(y) = \frac{m' f_y^1 + \left(\phi_\lambda - \phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}}\right) \Delta f_y}{-m'' f^1 + \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right) \Delta f}. \tag{27}$$

The derivation of the above equation is quite tedious and relegated to the appendix A.5.

Before presenting the conditions for PAC and PAM, we must first define two new variables which are determined by the search technology only and which enable us to write the conditions for PAC and PAM in a uniform way. Define

$$a^c(\zeta\lambda,\lambda) = \varepsilon_w(\zeta\lambda,\lambda)\frac{m'(\lambda)}{\lambda m''(\lambda)} \tag{28}$$

$$a^m(\zeta\lambda,\lambda) = \varepsilon_w(\zeta\lambda,\lambda)\frac{m'(\lambda)}{\lambda m''(\lambda)}\left(1 - \frac{\partial\phi_\mu/\partial\zeta}{\partial\phi_\mu/\partial\lambda}\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}\right) \tag{29}$$

Furthermore, define

$$\overline{a}^i \equiv \sup_{0\le\mu\le\lambda} a^i(\mu,\lambda), \quad \underline{a}^i \equiv \inf_{0\le\mu\le\lambda} a^i(\mu,\lambda) \tag{30}$$

When $\zeta = 0$, then $\phi(0,\lambda) = 0$ for any λ, which then implies that $\partial(\phi/m)/\partial\lambda = 0$; similarly, when $\zeta = 1$, then by definition $\phi(\lambda,\lambda) = m(\lambda)$, which then implies that $\partial(\phi/m)/\partial\lambda = 0$ since $\phi/m = 1$. Therefore, the more complicated expression for $a^m(\zeta\lambda,\lambda)$ reduces to $a^c(\zeta\lambda,\lambda)$ when $\zeta = 0$, or 1. That is,

$$a^c(0,\lambda) = a^m(0,\lambda) \quad \text{and} \quad a^c(\lambda,\lambda) = a^m(\lambda,\lambda). \tag{31}$$

Lemma 4. *Assume $\zeta(y)$ is both interior and differentiable at a point y. Then $\zeta'(y) \ge 0$ ($dh(\zeta(y),\lambda(y))/dy \ge 0$ respectively) holds at point y if and only if for $i = c$ ($i = m$ respectively)*

$$\frac{f^1\Delta f_y}{f_y^1\Delta f} \ge a^i\frac{1 - \frac{1}{m'}\left(\phi_\mu\frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}} \tag{32}$$

where we have suppressed the arguments from the functions $\phi(\zeta(y)\lambda(y),\lambda(y))$ and $m(\lambda(y))$, and $a^i(\zeta(y)\lambda(y),\lambda(y))$.

Proof. Plugging equation (27) into equations (25) and (26) gives the desired result, where we also used the definitions of $a^c(\zeta\lambda,\lambda)$ and $a^m(\zeta\lambda,\lambda)$ from equations (28) and (29), respectively. □

In general, it is difficult to completely solve the model and characterize the region where $\zeta(y)$ is interior and differentiable. However, the limit case considered by Proposition 1 where worker heterogeneity disappears is particularly simple and it can be used to derive necessary conditions for assortative contacting and matching. The result is given as follows.

Proposition 3. *A necessary condition for PAC/PAM to hold for any firm and worker distributions is*

$$\underline{\rho} \equiv \inf_{x,y} \rho(x,y) \geq \sup_{0 \leq \mu \leq \lambda} a^i(\mu, \lambda) \equiv \overline{a}^i \tag{33}$$

where for PAC we have $i = c$ and for PAM we have $i = m$. Similarly, a necessary condition for NAC/NAM to hold for any firm and worker distribution is that at each x and y

$$\overline{\rho} \equiv \sup_{x,y} \rho(x,y) \leq \inf_{0 \leq \mu \leq \lambda} a^i(\mu, \lambda) \equiv \underline{a}^i \tag{34}$$

where again for NAC we have $i = c$ and for NAM we have $i = m$.

Proof. See appendix A.6. □

So far, our results regarding sorting have not depended on the functional form of $\phi(\mu, \lambda)$ derived in Lemma 1. These results therefore hold for different microfoundations of the recruitment process as well. Our specific microfoundation allows for more progress, as we discuss next. In particular, the following lemma establishes that the right-hand side of equations (33) and (34) takes a simple form when $\phi(\mu, \lambda)$ is given by (5).

Lemma 5. *When $\phi(\mu, \lambda)$ satisfies* (5)*, we have, for $i = c$ and m,*

$$\overline{a}^i = \frac{1+\sigma}{2} \quad \text{and} \quad \underline{a}^i = \frac{1-\sigma}{2}. \tag{35}$$

Proof. See appendix A.7. □

The necessary conditions in Proposition 3 are derived by letting worker heterogeneity disappear. Our result shows that when the condition holds, then it is also sufficient for assortative contacting and matching to hold locally.

Lemma 6. *Assume that* (8)*, the second-order condition of a firm of type y, holds. If $\underline{\rho} \geq (1+\sigma)/2$, then* (32) *holds with the strict inequality "$>$". If $\overline{\rho} \leq (1-\sigma)/2$, then the reverse inequality holds ("$>$" replaced by "$<$").*

Proof. See appendix A.8. □

So far we have showed that for PAC/PAM to hold for any worker and firm endowments, then $\underline{\rho} \geq (1+\sigma)/2$. Similarly, for NAC/NAM we need $\overline{\rho} \leq (1-\sigma)/2$. For the sufficiency side, we proved that if $\zeta(y)$ is interior and differentiable at some point y_0, then the same conditions will lead to assortative contacting and matching to hold locally: $\zeta'(y_0) > 0$ and $\frac{d}{dy} h(\zeta(y_0), \lambda(y_0)) > 0$ in the case of PAC/PAM. If $\mathrm{Z}(y)$ is unique and continuous for all y, then the global version also holds.[27] To see this, consider the PAC/PAM case. Whenever $\zeta(y)$ and $h(\zeta(y), \lambda(y))$ are interior, they must be strictly increasing at the point. So it can never happen that the two functions are decreasing in y. The same logic applies to the NAC/NAM case. The following result shows that multiplicity and discontinuity of $\mathrm{Z}(y)$ do not affect our conclusions on positive or negative sorting.

Proposition 4. *If $\underline{\rho} \geq (1+\sigma)/2$, then in equilibrium $\zeta(y)$ is unique except at most one point y^*. For $y < y^*$, $\zeta(y) = 0$; for $y > y^*$, $\zeta(y)$ is continuous and increasing. Furthermore, if $\zeta(y) \in (0,1)$ at some point y_0, then $\zeta'(y_0) > 0$ and $\frac{d}{dy}h(\zeta(y_0), \lambda(y_0)) > 0$.*

If $\overline{\rho} \leq (1-\sigma)/2$, then in equilibrium $\zeta(y)$ is unique except at most one point y^*. For $y > y^*$, $\zeta(y) = 0$; for $y < y^*$, $\zeta(y)$ is continuous and decreasing. Furthermore, if $\zeta(y) \in (0,1)$ at some point y_0, then $\zeta'(y_0) < 0$ and $\frac{d}{dy}h(\zeta(y_0), \lambda(y_0)) < 0$.

Proof. See Appendix A.9. □

Of course, in equilibrium the point y^* described in the above proposition may not exist: $\zeta(y)$ is unique and strictly positive for all y, in which case we simply set $y^* = \underline{y}$ for the case $\underline{\rho} \geq (1+\sigma)/2$ and $y^* = \overline{y}$ for the case $\overline{\rho} \leq (1-\sigma)/2$.

Since the parameters $\underline{\rho}$ and $\overline{\rho}$ describe the degree of complementaries in production, we can alternatively state the proposition in terms of supermodularity. The following corollary presents this formulation of the result.

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

As the corollary indicates, the degree of complementarity required for PAC/PAM is increasing in the expected number of interviews that a firm can conduct. When $\sigma \to 0$ and meetings are bilateral, PAC/PAM requires square-root-supermodularity, in line with the results in Eeckhout and Kircher (2010a). 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 nowhere square-root-supermodular if $\sigma = 0$ and submodular when $\sigma = 1$. Figure 1 plots these results for the case of CES production.

[27] By the theorem of maximum, $\mathrm{Z}(y)$ is an upper hemiconintuous correspondence. Thus if $\mathrm{Z}(y)$ is unique for each y, it is a continuous function.

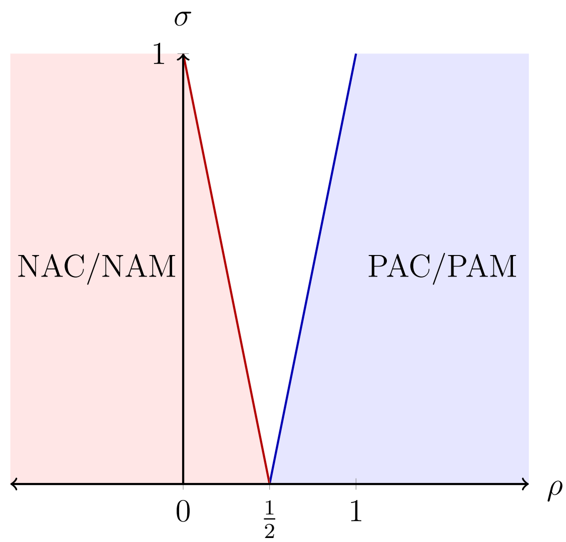

Figure 1: Combinations of ρ and σ that give rise to PAC/PAM (blue) or NAC/NAM (red) for any distribution of types, assuming a CES production function.

3.4 Wage Menus

We now show that firms can attract their most desired queue—i.e. the solution to their relaxed problem—by posting wage menus $\boldsymbol{w} = (w_1, w_2)$. We do so in two steps. First, we assume that firms post extended wage menus $\boldsymbol{w}_+$, which include a commitment to use the socially optimal hiring policy, i.e. a commitment to hire the most *productive* interviewee, even if a less productive interviewee yields a higher payoff. Subsequently, we show that this commitment is redundant, because in any equilibrium, a firm of type y will choose a wage menu satisfying $f(x_2, y) - w_2 > f(x_1, y) - w_1$, meaning that more productive workers are always more profitable.

Payoffs. Given the hiring priority for high-type workers, a firm that posts a wage menu $\boldsymbol{w}_+$ and accordingly attracts queues (μ, λ) has an expected payoff that equals

$$\pi\left(\boldsymbol{w}_+, \zeta, \lambda, y\right) = S(\zeta\lambda, \lambda, y) - \phi\left(\zeta\lambda, \lambda\right) w_2 - \left(m(\lambda) - \phi(\zeta\lambda, \lambda)\right) w_1, \tag{36}$$

where $S(\mu, \lambda, y)$ is given by equation (6). Intuitively, the firm hires a high-type worker if it interviews at least one such worker, which happens with probability $\phi(\zeta\lambda, \lambda)$. Similarly, the firm hires a low-type worker if it interviews no high-type workers but at least one low-type workers, which happens with probability $\phi(\lambda, \lambda) - \phi(\zeta\lambda, \lambda)$.

By an accounting identity, the probability $\psi_i\left(\zeta, \lambda\right)$ that a worker of type x_i is hired by the firm satisfies

$$\psi_1(\zeta,\lambda) = \frac{\phi(\lambda,\lambda) - \phi(\zeta\lambda,\lambda))}{\lambda - \zeta\lambda} \qquad \text{and} \qquad \psi_2(\zeta,\lambda) = \frac{\phi(\zeta\lambda,\lambda)}{\zeta\lambda}.$$

Note that these expressions are also valid for $\zeta = 0$ or 1, in which case we take the corresponding limit. Given these matching probabilities, a worker's expected payoff from applying to the firm equals

$$V_i(\boldsymbol{w}_+, \zeta\lambda, \lambda) = \psi_i(\zeta,\lambda)\, w_i. \tag{37}$$

In order to attract a worker of type x_i, the firm needs to ensure that the worker's expected payoff $V_i(\boldsymbol{w}_+, \zeta, \lambda)$ equals his market utility U_i. This constraint can be used to substitute the wage w_i out of the firm's expected payoff (36), which yields

$$\Pi(\zeta,\lambda,y) = S(\zeta\lambda,\lambda,y) - \lambda U_1 - \zeta\lambda\,(U_2 - U_1)\,, \tag{38}$$

i.e. the payoff in the firm's relaxed problem.[28]

Uniqueness of Queues. Suppose that a solution to the relaxed problem of a given firm of type y is $(\zeta^*\lambda^o, \lambda^o)$. To attract this queue, the firm can post a wage menu $\boldsymbol{w}_+^*$ in which the wage for a worker of type x_i equals $w_i^* = U_i/\psi_i(\zeta^*, \lambda^o)$. However, there is an issue of uniqueness: is it possible that a firm posting $\boldsymbol{w}_+^*$ may end up with queues different from the desired one $(\zeta^*\lambda^o, \lambda^o)$? The answer is not trivial because one could imagine two different queues giving workers the same expected payoffs, e.g. i) a long queue with only low-type workers, where high-type workers do not apply because of negative meeting externalities, and ii) a queue of medium length where both types of workers apply, but the number of low-type workers is limited because they are discouraged by the presence of high-type workers. However, the following proposition shows that such a multiplicity does not actually arise: for any wage menu that firms post, there will be exactly one queue that is compatible with the market utilities. Thus, firms can indeed solve their relaxed problem by posting a wage menu $\boldsymbol{w}_+$.

Proposition 5. *For any wage menu $\boldsymbol{w}_+$, there exists exactly one solution (μ, λ) to the market utility condition.*

Proof. See Appendix A.10. $\square$

[28]We have implicitly assumed that $0 < \zeta < 1$ so that both types of workers are present and both market utility conditions hold with equality. However, it is easy to see that equation (38)—or equivalently, equation (10)—holds for the cases $\zeta = 0$ or $\zeta = 1$ as well.

Productivity vs Profitability. We have showed that firms can achieve the same outcome as their relaxed problem by posting wage menus $\boldsymbol{w}_+$. We now turn to the case in which firms post wage menus $\boldsymbol{w} = \{w_1, w_2\}$ that do not include any commitment to hire the most productive interviewee; instead, firms will hire their most profitable interviewee, which depends on the ranking of $f(x_1, y) - w_1$ versus $f(x_2, y) - w_2$. Note that the cost of attracting queues (μ, λ) remains $\mu U_2 + (\lambda - \mu)U_1$ in this case, since workers must always receive their market utility. Conditional on attracting this queue, the surplus created by a firm without commitment is less or equal to that with commitment. Firms can therefore not obtain higher payoffs without commitment. However, it turns out that they do not obtain lower payoffs either: when a firm of type y chooses the optimal wage menu $\boldsymbol{w}_+^* = \{w_1^*, w_2^*\}$, i.e., the wage menu that attracts the most profitable queue $(\zeta^*\lambda^o, \lambda^o)$, then $f(x_2, y) - w_2^* > f(x_1, y) - w_1^*$, such that the firm will prefer x_2 workers when both types of workers are available. The following proposition formalizes this result.[29]

Lemma 7. *If a wage menu $\boldsymbol{w}_+^*$ corresponds to an optimal solution $(\zeta^*\lambda^o, \lambda^o)$ to the relaxed problem of a firm of type y with $0 < \zeta^* < 1$, then $f(x_2, y) - w_2^* > f(x_1, y) - w_1^*$.*

Proof. See Appendix A.11. □

In sum, we have showed that the directed search equilibrium with wage menus $\boldsymbol{w}$ coincides with the directed search equilibrium with wage menus $\boldsymbol{w}_+$. Moreover, the latter equilibrium is equivalent a competitive price equilibrium in which firms can buy queues directly from a competitive market. Hence, by the first theorem of welfare economics, we obtain the following efficiency result.

Proposition 6. *The directed search equilibrium is constrained efficient.*

4 Generalization

In this section, we show that our main results are robust to various extensions of our environment. In particular, we allow for noisy signals as well as wide class of processes governing workers' applications and interviews.

4.1 Signals

In the setup that we have analyzed so far, firms have absolutely no information about applicants' types when selecting interviewees. In practice, there often exist relatively easy ways

[29] A similar result appears in Shimer (2005) for the case of urn-ball meetings. Our proof of Lemma 7 does not depend on the specific microfoundation of the recruiting process and therefore generalizes his result.

to obtain a signal, e.g. from a quick look at applicants' resumes. Our baseline environment can be extended quite easily to capture this idea.

Environment with Signals. Assume that firms 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.

Isomorphism. It turns out that this 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. The following lemma formalizes this result.

Proposition 7. *In the 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.12. □

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.

4.2 Arbitrary Contact Technologies

So far, we have considered a specific micro-foundation of the *contact technology*, i.e. the process that governs a firm's number of interviews. However, many reasonable alternatives exist; for example, workers may send their applications according to an urn-ball process as in Shimer (2005), after which firms can interview a subset of their applicants.[30] Our analysis can easily be extended to such alternatives as they simply imply a different functional form for $\phi(\mu, \lambda)$ but do not otherwise affect the derivation of our results.

[30] See Lester et al. (2015) and Cai et al. (2017) for other examples.

Bilateral Contact Technologies. Within this broader class of contact technologies, there exist natural generalizations of the cases $\sigma = 0$ and $\sigma = 1$ that we have considered above. First, $\sigma = 0$ describes minimal screening in the sense that each firm interviews only a single applicant. That is, contacts are bilateral (a firm can be at contact with at most one worker) and a firm's probability $\phi(\mu, \lambda)$ of interviewing a high-type worker is simply the product of the probability $\phi(\lambda, \lambda)$ that it has any applicant and the conditional probability μ/λ that this applicant is a high type. Sorting patterns under bilateral contacts have been analyzed in detail by Eeckhout and Kircher (2010a). They define

$$a^{EK}(\lambda) = \frac{m'(\lambda)(m'(\lambda)\lambda - m(\lambda))}{\lambda m(\lambda)m''(\lambda)}, \quad \overline{a}^{EK} = \sup_{\lambda} a^{EK}(\lambda), \quad \underline{a}^{EK} = \inf_{\lambda} a^{EK}(\lambda) \tag{39}$$

It is easy to see that when contacts are bilateral, our definition of $a^c(\mu, \lambda)$ and $a^m(\mu, \lambda)$ are constant in μ and both coincide with $a^{EK}(\lambda)$. Furthermore, the sorting results in (Eeckhout and Kircher, 2010a) (their Theorem 1) can be easily obtained as a limit result by our analysis.

Proposition 8. *[Eeckhout and Kircher, 2010a] Suppose contacts are bilateral, i.e.* $\phi(\mu, \lambda) = \frac{\mu}{\lambda}\phi(\lambda, \lambda)$. *The market equilibrium then exhibits PAM for any distribution of types if and only if* $f(x, y)$ *is n-root-supermodular, where* $n = (1 - \overline{a}^{EK})^{-1}$ *In contract, the market equilibrium exhibits NAM for any distribution of types if and only if* $f(x, y)$ *is nowhere n-root-supermodular, where* $n = (1 - \underline{a}^{EK})^{-1}$.

Proof. To be added. □

Invariant Contact Technologies. In contrast, $\sigma = 1$ means that screening is perfect in the sense that the presence of low-type applicants does not make it harder for a firm to identify a high-type applicant. That is, $\phi_\lambda(\mu, \lambda) = 0$ for all μ and λ. As shown by Cai et al. (2018), this condition is the defining characteristic of the class of *invariant* contact technologies, first introduced by Lester et al. (2015).

For invariant contact technologies, the two elasticities $\varepsilon_f(\mu, \lambda)$ and $\varepsilon_f(\mu, \lambda)$ defined by equations (19) and (20) will depend on the first argument only and hence can be simply written as $\epsilon_f(\mu)$ and $\epsilon_w(\mu)$. The explicit expressions for the two elasticities are now given by

$$\epsilon_f(\mu) = \frac{\mu m'(\mu)}{m(\mu)} \quad \text{and} \quad \epsilon_w(\mu) = \frac{\mu m''(\mu)}{m'(\mu)} \tag{40}$$

Thus $\epsilon_f(\mu)$ is always positive and $\epsilon_w(\mu)$ is always negative since $m(\mu)$ is concave. Furthermore, the two functions $a^c(\mu, \lambda)$ and $a^m(\mu, \lambda)$, which are defined by equations (28) and (29) and are crucial for our sorting results, now can be rewritten as[31]

$$a^c(\mu,\lambda) = \frac{\epsilon_w(\mu)}{\epsilon_w(\lambda)} \quad \text{and} \quad a^m(\mu,\lambda) = \frac{\epsilon_w(\mu)}{\epsilon_w(\lambda)}\frac{\epsilon_f(\lambda)}{\epsilon_f(\mu)} \tag{41}$$

As the previous results suggest, our condition on sorting depends on $\overline{a}^i$ and $\underline{a}^i$ for $i = c$ and m. Our next result shows that for invariant contact technologies, $\underline{a}^c$ and $\underline{a}^m$ are always 0.

Lemma 8. *Suppose contacts are invariant, i.e.* $\phi_\lambda(\mu,\lambda) = 0$. *We have* $\lim_{\mu\to 0} \epsilon_f(\mu) = 1$ *and* $\lim_{\mu\to 0} \epsilon_w(\mu) = 0$. *Thus* $\underline{a}^c = \underline{a}^m = 0$.

Proof. See appendix A.13. □

Next we consider $\overline{a}^c$ and $\overline{a}^m$. By setting $\mu = \lambda$ in (41), we can see that $\overline{a}^c, \overline{a}^m \geq 1$. For common invariant contact technologies such as urn-ball or geometric technologies, one can prove that $\overline{a}^c = \overline{a}^m = 1$. However, this is not always the case.[32]

We have seen from Proposition 4 and 8 that $\underline{\rho} \geq \overline{a}^i$ is necessary and sufficient for positive assortative contacting ($i = c$) and matching ($i = m$), and $\overline{\rho} \leq \overline{a}^i$ is necessary and sufficient for negative assortative contacting ($i = c$) and matching ($i = m$). Our next result shows that this condition continues to hold for invariant contact technologies.

Proposition 9. *Suppose contacts are invariant, i.e.* $\phi_\lambda(\mu,\lambda) = 0$. *The market equilibrium then exhibits PAM (PAC resp.) for any distribution of types if and only if* $\underline{\rho} \geq \overline{a}^m$ ($\underline{\rho} \geq \overline{a}^c$ *resp.). In contrast, the market equilibrium exhibits NAM for any distribution of types if and only if* $f(x,y)$ *is submodular.*

Proof. See appendix A.14. □

Our results so far also suggest that if we are willing to make the following two assumptions, which are satisfied by common invariant contact technologies, the sorting results become particularly simple.

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

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

[31] The expression of $a^m(\mu,\lambda)$ is derived as follows. $\phi_\mu(\zeta\lambda,\lambda) = m'(\zeta\lambda)$ and $h(\zeta,\lambda) = m(\zeta\lambda)/m(\lambda)$. Therefore, the following term in equation (29) can be rewritten as

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

[32] Consider, for example, a mixture between the urn-ball and the geometric contact technologies: $m(\mu) = t(1-e^{-\mu}) + (1-t)(1-1/(1+\mu))$. Numerically one can see that: when $t = 0.2$, $\overline{a}^c > 1$ and $\overline{a}^m = 1$; when $t = 0.98$, both $\overline{a}^c$ and $\overline{a}^m$ are strictly bigger than 1.

With the above assumptions, one can easily see that $\overline{a}^c = \overline{a}^m = 1$. The conditions for PAC/PAM then become $\underline{\rho} \geq 1$. That is, for any x and y we have $\rho(x, y) \geq 1$, which is equivalent to $\frac{\partial^2}{\partial x \partial y} \log f(x, y) \geq 1$, i.e., $f(x, y)$ is log-supermodular.

Corollary 2. *Suppose contacts are invariant, i.e. $\phi_\lambda(\mu, \lambda) = 0$ and satisfy Assumption INV-1 and INV-2. The market equilibrium then exhibits PAC/PAM for any distribution of types if and only if $f(x, y)$ is log-supermodular. In contract, the market equilibrium exhibits NAC/NAM for any distribution of types if and only if $f(x, y)$ is submodular.*

5 Conclusion

[to be completed]

Appendix A Proofs

A.1 Proof of Lemma 1

The firm's *potential* number of interviews n_I follows a geometric distribution with support $\mathbb{N}_1$ and mean $(1-\sigma)^{-1}$. That is: $\mathbb{P}(n_I = n) = (1-\sigma)\sigma^{n-1}$ for $n = 1, 2, \ldots$. However, interviewing might be constrained by the firm's number of applicants n_A, which also follows a geometric distribution but with support $\mathbb{N}_0$ and mean λ, i.e., $\mathbb{P}(n_A = n) = \frac{1}{1+\lambda}(\frac{\lambda}{1+\lambda})^n$ for $n = 0, 1, 2, \ldots$. Hence, the firm's *actual* number of interviews, n, is $\min\{n_I, n_A\} \in \mathbb{N}_0$. Since the potential number of interview n_I is at least one, $P_0(\lambda) = \frac{1}{1+\lambda}$. For $n \geq 1$, things are more complicated and we have the following: for $n \geq 1$,

$$P_n(\lambda) = (1-\sigma)\sigma^{n-1} \sum_{j=n}^{\infty} \frac{1}{1+\lambda} \left(\frac{\lambda}{1+\lambda}\right)^j + \frac{1}{1+\lambda} \left(\frac{\lambda}{1+\lambda}\right)^n \sum_{j=n+1}^{\infty} (1-\sigma)\sigma^{j-1}$$

where the first term on the right-hand side denotes the case where the number of applicants is (weakly) larger than n while the number of interviews equals n; the second term denotes the case where the number of applicants equals n while the number of interviews is strictly larger than n. Hence, the firm's actual number of interviews n has the following probability distribution:

$$P_n(\lambda) \equiv \mathbb{P}\left[\min\{n_I, n_A\} = n | \lambda\right] = \begin{cases} \frac{1}{1+\lambda} & \text{for } n = 0, \\ \sigma^{n-1} \frac{1}{1+\lambda} \left(\frac{\lambda}{1+\lambda}\right)^n (1 + (1-\sigma)\lambda) & \text{for } n \in \mathbb{N}_1. \end{cases} \tag{42}$$

The probability that n interviews reveal at least one high-type worker is $1 - (1 - \mu/\lambda)^n$. Taking expectations with respect to n gives

$$\phi(\mu, \lambda) = 1 - \sum_{n=0}^{\infty} P_n(\lambda) \left(1 - \frac{\mu}{\lambda}\right)^n.$$

Substituting (42) and simplifying the result yields equation (5). □

A.2 Proof of Lemma 3

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 readily gives condition (8).

Since ϕ is given by equation (5), direct computation yields

$$\phi_{\lambda\lambda} - \frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} = \frac{(1-\sigma)^2}{2\sigma(1+(1-\sigma)\lambda)(1+\sigma\mu+(1-\sigma)\lambda)}$$

which is strictly positive if and only if $\sigma < 1$. Note that $m''(\lambda) = -2/(1+\lambda)^3$. Hence,

$$\widehat{\kappa}(\mu, \lambda) \equiv \frac{-m''}{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}} = \frac{4\sigma(1+(1-\sigma)\lambda)(1+\sigma\mu+(1-\sigma)\lambda)}{(1-\sigma)^2(1+\lambda)^3}$$

It is easy to see that for a given λ, the right-hand side is increasing in μ and reaches its maximum at $\mu = \lambda$. Furthermore, we differentiate this maximum with respect to λ, and find it is strictly decreasing in λ. That is,

$$\frac{\partial}{\partial\lambda}\left(\frac{4\sigma(1+(1-\sigma)\lambda)}{(1-\sigma)^2(1+\lambda)^2}\right) = -\frac{4\sigma(1+\sigma+(1-\sigma)\lambda)}{(1-\sigma)^2(1+\lambda)^3} < 0.$$

Therefore, $\overline{\kappa} = \sup_{0<\mu<\lambda} \widehat{\kappa}(\mu, \lambda) = 4\sigma/(1-\sigma)^2$. □

A.3 Proof of Proposition 1

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 $\underline{U} > 0$, which is the market utility of workers if all workers are the same and have 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 of surplus of a worker will be $m'(\lambda)f(x, y)$ for a firm of type y. Define

$\overline{\lambda}$ by $m'(\overline{\lambda})f(x,\overline{y}) = \underline{U}$, where $\overline{y}$ is the supremum of firm types. therefore, we can restrict each firm's choice of queues to be in the convex set $\{(\mu,\lambda)\,|\,0\le\mu\le\lambda\le\overline{\lambda}\}$.

Since the above set is compact, for sufficiently small $x_2 - x_1$, equation (8) will be strictly positive, which implies that for each firm type y, the surplus function will be strictly concave in this set. Therefore, by the theorem of the maximum, firms' 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 first-order conditions: equation (11) and (14). Hence by the implicit function theorem, they are both differentiable at the point. □

A.4 Proof of Proposition 2

Recall that $\overline{\rho} = \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x,y)$ and $\underline{\rho} = \inf_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x,y)$, where $\rho(x,y) = \frac{f(x,y)f_{xy}(x,y)}{f_x(x,y)f_y(x,y)}$.

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

$$\frac{\partial}{\partial x}\big(\log f_y(x,y) - \underline{\rho}\log f(x,y)\big) = \frac{f_{xy}}{f_y} - \underline{\rho}\frac{f_x}{f} = \frac{f_{xy}f - \underline{\rho} f_x f_y}{f f_y} \ge 0$$

where on the right-hand side we suppressed the arguments of $f(x,y)$.

The case with $\overline{\rho}$ follows a similar logic.

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

A.5 Derivation of Equation (27)

Combining equations (23) and (24) gives ,

$$\lambda'(y) = \frac{-\frac{\phi_\mu}{\partial\phi_\mu/\partial\zeta}\frac{\Delta f_y}{\Delta f}\lambda_\zeta^o + \lambda_y^o}{1 + \frac{\partial\phi_\mu/\partial\lambda}{\partial\phi_\mu/\partial\zeta}\lambda_\zeta^o}.$$

Next we plug in the expressions of λ_y^o and λ_ζ^o from equations (12) and (17), respectively. The numerator of the right-hand side of the above equation becomes

$$\begin{aligned}\frac{\phi_\mu}{\lambda\phi_{\mu\mu}}\frac{\Delta f_y}{\Delta f}\frac{\lambda\frac{\partial\phi_\mu}{\partial\lambda}\Delta f}{m''f^1 + \frac{\partial^2\phi}{\partial\lambda^2}\Delta f} - \frac{m'f_y^1 + \frac{\partial\phi}{\partial\lambda}\Delta f_y}{m''f^1 + \frac{\partial^2\phi}{\partial\lambda^2}\Delta f} &= \frac{\frac{\phi_\mu}{\phi_{\mu\mu}}\frac{\partial\phi_\mu}{\partial\lambda}\Delta f_y - \left(m'f_y^1 + \frac{\partial\phi}{\partial\lambda}\right)\Delta f_y}{m''f^1 + \frac{\partial^2\phi}{\partial\lambda^2}\Delta f} \\ &= \frac{-m'f_y^1 - \left(\phi_\lambda - \phi_\mu\frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}}\right)\Delta f_y}{m''f^1 + \frac{\partial^2\phi}{\partial\lambda^2}\Delta f}.\end{aligned}$$

Similarly, the denominator becomes

$$
\begin{aligned}
1-\frac{\partial\phi_\mu/\partial\lambda}{\lambda\phi_{\mu\mu}}\frac{\lambda\frac{\partial\phi_\mu}{\partial\lambda}\Delta f}{m''f^1+\frac{\partial^2\phi}{\partial\lambda^2}\Delta f} &= \frac{m''f^1+\frac{\partial^2\phi}{\partial\lambda^2}\Delta f-\frac{1}{\phi_{\mu\mu}}\left(\frac{\partial\phi_\mu}{\partial\lambda}\right)^2\Delta f}{m''f^1+\frac{\partial^2\phi}{\partial\lambda^2}\Delta f}\\
&= \frac{m''f^1+(\zeta^2\phi_{\mu\mu}+2\zeta\phi_{\mu\lambda}+\phi_{\lambda\lambda})\Delta f-\frac{1}{\phi_{\mu\mu}}\left(\zeta^2\phi_{\mu\mu}^2+2\zeta\phi_{\mu\mu}\phi_{\mu\lambda}+\phi_{\mu\lambda}^2\right)\Delta f}{m''f^1+\frac{\partial^2\phi}{\partial\lambda^2}\Delta f}\\
&= \frac{m''f^1-\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}}-\phi_{\lambda\lambda}\right)\Delta f}{m''f^1+\frac{\partial^2\phi}{\partial\lambda^2}\Delta f}.
\end{aligned}
$$

Combining the expressions for the numerator and the denominator gives equation (27). □

A.6 Proof of Proposition 3

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

Suppose that equation (33) does not hold for $i=c$ and there exists x', y', μ', and λ^o 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 $\epsilon_0>0$ 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 $\lambda\in[\lambda'-\epsilon_0,\lambda'+\epsilon_0]$.

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 $\epsilon_1\to 0$. Then for sufficiently small ϵ_1, by Proposition 1 the equilibrium $\lambda(y)$ is unique, continuous, and will belong to the set $[\lambda'-\epsilon_0,\lambda'+\epsilon_0]$ for all y (note in the proof of Proposition 1, we only let $x_2\to x_1$ and keep the firm value distribution constant, but it is easy to see the whole arguments carry to the case where we let $x_2\to x1$, $\underline{y}\uparrow y'$ and $\overline{y}\downarrow y'$ simultaneously). Furthermore, $\mu(y)$ is continuous and $\int_{\underline{y}}^{\overline{y}}\mu(y)dJ(y)=\mu'$, where $J(y)$ is the distribution of firms types. Therefore, by continuity there exists some y_0 such that $\mu(y_0)=\mu'$. In sum, at point y_0 we have $\mu'=\mu(y_0)<\lambda(y_0)\in(\lambda'-\epsilon_0,\lambda'+\epsilon_0)$. Also by Proposition 1, $\zeta(y)$ is differentiable at point y_0. Hence the assumptions of Lemma 4 are satisfied at point y_0.

It is easy to see that when $\epsilon_1\to 0$ ($x_2\to x_1=x'$), the left-hand side of equation (32) approaches to $\rho(x', y')=f(x', y')f_{xy}(x', y')/f_x(x', y')f_y(x', y')$. 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,

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

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 equation (32) also approaches to 1. Therefore, for equation (32) to hold at point y_0, we need $\rho(x', y') \geq a^c(\mu', \lambda')$. We have thus reached a contradiction, and for PAC to hold for all worker and firm endowments, $\rho(x, y) \geq a^c(\mu, \lambda)$ for all (x, y) and (μ, λ). □

A.7 Proof of Lemma 5

We first consider $a^c(\mu, \lambda)$. Since $\phi(\mu, \lambda)$ is given by equation (5) and $a^c(\mu, \lambda)$ is defined by equation (28), by direct calculation we have

$$a^c(\mu, \lambda) = \frac{1 + \lambda}{2\lambda} \left(1 + \frac{1}{1 + (1 - \sigma)\lambda} - \frac{2}{1 + \sigma\mu + (1 - \sigma)\lambda} \right) \tag{43}$$

It is easy to see that $a^c(\mu, \lambda)$ is strictly increasing in μ. Thus for a given λ, we have

$$\max_\mu a^c(\mu, \lambda) = a^c(\lambda, \lambda) = \frac{1}{2} \frac{1 + \sigma + (1 - \sigma)\lambda}{1 + (1 - \sigma)\lambda} \tag{44}$$

$$\min_\mu a^c(\mu, \lambda) = a^c(0, \lambda) = \frac{1}{2} \frac{(1 - \sigma)(1 + \lambda)}{1 + (1 - \sigma)\lambda} \tag{45}$$

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

$$\frac{da^c(\lambda, \lambda)}{\lambda} = -\frac{\sigma(1 - \sigma)}{2(1 + (1 - \sigma)\lambda)^2}$$

$$\frac{da^c(0, \lambda)}{\lambda} = \frac{\sigma(1 - \sigma)}{2(1 + (1 - \sigma)\lambda)^2}$$

Therefore, $a^c(\lambda, \lambda)$ approaches its supreme when $\lambda \to 0$ and $a^c(0, \lambda)$ approaches its infimum when $\lambda \to 0$. Therefore, $\sup_{0 \leq \mu \leq \lambda} a^i(\mu, \lambda) = \lim_{\lambda \to 0} a^c(\lambda, \lambda) = (1 + \sigma)/2$ and $\inf_{0 \leq \mu \leq \lambda} a^i(\mu, \lambda) = \lim_{\lambda \to 0} a^c(0, \lambda) = (1 - \sigma)/2$. 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

$$a^m(\mu, \lambda) = \frac{\lambda(1 + \lambda)(1 - \sigma) + 2\sigma\mu}{2\lambda 1 + (1 - \sigma)\lambda} \tag{46}$$

It is easy to see that $a^m(\mu, \lambda)$ is strictly increasing in μ. For a given λ, $a^m(\mu, \lambda)$ reaches its maximum at $\mu = \lambda$ and its minimum at $\mu = 0$. Because of equation (31), the rest of the proof is the same as the case of $a^c(\mu, \lambda)$. $\square$

A.8 Proof of Lemma 6

First we consider the case $\underline{\rho} \geq (1+\sigma)/2$. Since f is strictly supermodular, $\Delta f_y > 0$ and the left-hand side of (32) is strictly positive. We will prove a slightly stronger version of (32):

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

where we have replaced $a^i(\mu, \lambda)$ with its supreme $(1+\sigma)/2$. 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. Also note that the supreme of $a^i(\mu, \lambda)$ is never reached as a maximum for $i = c$ and m. So the above weak inequality implies a strong inequality of (32). Since the denominator on the right-hand side is positive because of firms' second-order condition, multiplying it on both sides and rearranging terms gives

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

Since $\phi(\mu, \lambda)$ is given by equation (5), plug it into the above equation and we have

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

It is easy to see that the left-hand side is decreasing in μ and reaches its minimum at $\mu = \lambda$, at which the above inequality becomes

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} + \frac{\Delta f_y}{f_y^1} \frac{(1-\sigma)(2+(1-\sigma)\lambda)(1+\lambda)}{4(1+(1-\sigma)\lambda)} \geq \frac{1+\sigma}{2}$$
$$\Leftrightarrow \frac{f^1 \Delta f_y}{f_y^1 \Delta f} + \frac{\Delta f_y}{f_y^1} \frac{(1-\sigma)}{4}\left(1+\lambda+\frac{1+\lambda}{1+(1-\sigma)\lambda}\right) \geq \frac{1+\sigma}{2}$$

It is easy to see that the left-hand side is increasing in λ and approaches its infimum as $\lambda \to 0$. Therefore, a sufficient condition for the inequality that we want to prove is

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} + \frac{\Delta f_y}{f_y^1} \frac{(1-\sigma)}{2} - \frac{1+\sigma}{2} \geq 0$$

By equation (18), we know that $\Delta f_y / f_y^1 \geq (1+\kappa(y))^{\underline{\rho}} - 1$. We then replace the above inequality with a stronger one:

$$\frac{(1+\kappa)^{\underline{\rho}} - 1}{\kappa} + \left((1+\kappa)^{\underline{\rho}} - 1\right) \frac{(1-\sigma)}{2} - \frac{1+\sigma}{2} \geq 0,$$

where we have suppressed the argument of $\kappa(y)$. If $\underline{\rho} \geq 1$, then the above inequality holds trivially because the first term on the left-hand side is (weakly) larger than 1. If $(1+\sigma)/2 \leq \underline{\rho} < 1$, then note that the left-hand side is decreasing in σ and reaches its maximum at $(1+\sigma)/2 = \underline{\rho}$. Hence we can replace the above inequality by the following.

$$\begin{aligned} &\frac{(1+\kappa)^{\underline{\rho}} - 1}{\kappa} + \left((1+\kappa)^{\underline{\rho}} - 1\right)(1-\underline{\rho}) - \underline{\rho} \geq 0 \\ \Leftrightarrow &\frac{(1+\kappa)^{\underline{\rho}} - 1}{\kappa} + (1+\kappa)^{\underline{\rho}}(1-\underline{\rho}) - 1 \geq 0 \\ &\Leftrightarrow 1 + \kappa(1-\underline{\rho}) - (1+\kappa)^{1-\underline{\rho}} \geq 0 \end{aligned}$$

The last inequality is straightforward since the left-hand side is increasing in κ and its infimum is zero as $\kappa \to 0$. Hence we have proved our claim for the case $\underline{\rho} \geq (1+\sigma)/2$.

Next we consider the case $\overline{\rho} \leq (1-\sigma)/2$. As before, by equation (18) we have

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \leq \frac{(1+\kappa)^{\overline{\rho}} - 1}{\kappa} \leq \overline{\rho} \leq \frac{1-\sigma}{2} \tag{47}$$

The first inequality above holds because when $\overline{\rho} \leq 0$, the left-hand side is negative, and when $0 < \overline{\rho} \leq (1-\sigma)/2$, the left-hand side is decreasing in κ and its supreme is $\overline{\rho}$ when $\kappa = 0$.

Next, we will first prove the following inequality.

$$1 \leq \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right) \frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right) \frac{\Delta f}{f^1}} \tag{48}$$

Note that both $\frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)$ and $\frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)$ are positive and if the form is zero, then the above inequality holds trivially. So in the following we only focus on the case where it is strictly positive. Since the denominator is always positive because of (8), rewriting the above inequality gives

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

where the equality follows from direct computation. It is easy to see that

$$\frac{\Delta f_y}{f_y^1}\frac{f^1}{\Delta f} < \frac{1-\sigma}{2} < \frac{1-\sigma}{2}\frac{1+\lambda}{1+(1-\sigma)\lambda}$$

where the first inequality follows from (47). Thus (48) holds and we have

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

We have thus proved our claim for the case $\overline{\rho} \leq (1-\sigma)/2$. □

A.9 Proof of Proposition 4

For the proof we need the following result from Cai et al. (2019), which analyzed the case where all firms are homogeneous. For given market utility (U_1, U_2), there are at most two optimal ζ which solve a firm's relaxed problem. Furthermore, if there are two ζ for a firm, then one of the two must equal zero. That is, $|\mathrm{Z}(y)| \leq 2$, and if $|\mathrm{Z}(y)| = 2$, then $0 \in \mathrm{Z}(y)$. In the following, we divide our proof into three steps.

Step 1: By the theorem of maximum, $\mathrm{Z}(y)$ is an upper hemicontinuous correspondence. If $\mathrm{Z}(y)$ is unique for all y, then $\mathrm{Z}(y) = \zeta(y)$ is a continuous function. If at some point y_0, $\zeta(y_0) \in (0,1)$, then $\zeta(y_0)$ satisfies equation (14). By the implicit function theorem, $\zeta(y)$ is differentiable at point y_0. Thus as we remarked before Proposition 4, in this case the local result from Lemma 6 implies the global result stated in Proposition 4.

Step 2: Next consider the case where at some point y^*, $\mathrm{Z}(y)$ contains two elements: 0 and $\zeta^* > 0$. Since $\mathrm{Z}(y)$ is hemicontinuous, for firms with type y sufficiently close to y^* there are three scenarios: i) $\mathrm{Z}(y)$ all contain two elements $\zeta^a(y)$ and $\zeta^*(y)$ where $\zeta^a(y)$ is always zero and $\zeta^*(y)$ is continuous; ii) $\mathrm{Z}(y)$ is unique for $y \neq y^*$, $\zeta(y) = 0$ for $y > y^*$ and $\zeta(y)$ is continuous and $\lim_{y \nearrow y^*} \zeta(y) = \zeta^*$ for $y < y^*$; iii) $\mathrm{Z}(y)$ is unique for $y \neq y^*$, $\zeta(y) = 0$ for $y < y^*$ and $\zeta(y)$ is continuous and $\lim_{y \searrow y^*} \zeta(y) = \zeta^*$ for $y > y^*$. We will prove later that when $\underline{\rho} \geq (1+\sigma)/2$, case i) and ii) will not arise, and when when $\overline{\rho} \leq (1-\sigma)/2$, case i) and iii) will not arise.

Step 3: Note that the claim in Step 2 implies that y^* must be the only firm type with multiple optimal ζ's, i.e., y^* is the only y with $|\mathrm{Z}(y)| = 2$. Consider the PAC/PAM case ($\underline{\rho} \geq (1+\sigma)/2$). If not, then let y^{**} be the smallest y which is bigger than y^* and has two optimal ζ's. Then for $y \in (y^*, y^{**})$, $\zeta(y)$ is unique, which, by the argument in Step 1, implies that it is also continuous and increasing by the theorem of maximum. However, the above claim implies that firms with y slightly below y^{**} should pick $\zeta = 0$, which leads to a contradiction. Similarly, if there exists a largest y^{***} which is smaller than y^* and has two optimal ζ's, then again we have a contradiction. The NAC/NAM case ($\overline{\rho} \leq (1-\sigma)/2$) follows the same logic.

Step 4: Therefore, $\mathrm{Z}(y)$ is unique for firms with either $y < y^*$ or $y > y^*$ and hence continuous by the argument in Step 1. If $\underline{\rho} \geq (1+\sigma)/2$, then $\zeta(y)$ and $h(\zeta(y), \lambda(y))$ are increasing whenever $\zeta(y)$ is interior. Therefore, Proposition 4 follows readily.

Proof Step 2. Recall that $\Pi^*(\zeta, y)$ is defined by equation (13) and $\lambda^o(\zeta, y)$ is given by equation (11). Therefore, $\Pi^*(\zeta, y)$ can be rewritten as

$$\Pi^*(\zeta, y) = (m(\lambda^o) - \lambda^o m'(\lambda^o))\, f^1 + \left(\phi(\zeta\lambda^o, \lambda^o) - \lambda^o \frac{\partial \phi(\zeta\lambda^o, \lambda^o)}{\partial \lambda^o}\right) \Delta f \tag{49}$$

where we suppressed the arguments of $\lambda^o(\zeta, y)$.

Consider a firm with type y^*: $\max_\zeta \Pi^*(\zeta, y) = \Pi^*(0, y^*) = \Pi^*(\zeta^*, y^*)$. If $\underline{\rho} \geq (1+\sigma)/2$, we will show that for firms with y slightly above y^*, it is strictly better for them to choose ζ close to ζ^* instead of 0, and for firms with with y slightly below y^*, it is strictly better for them to choose $\zeta = 0$. This corresponds to case iii) in Step 2. In case i) mentioned in Step 2, firms with y close to y^* will be indifferent between $\zeta = 0$ and restricting their choice of ζ close to ζ^*. In case ii), it is the other way around, firms with y slightly above y^* will choose $\zeta = 0$ and firms with y slightly below y^* will choose a ζ close to ζ^*.

For firms with y around y^*, if we restrict their choice of ζ to be close to ζ^* then we denote the maximum value of $\Pi^*(\zeta, y)$ by $\widetilde{\Pi}(y)$. Our claim is equivalent to the following: if $\underline{\rho} \geq (1+\sigma)/2$, we have

$$\frac{d}{dy}\Pi^*(0, y) < \frac{d}{dy}\widetilde{\Pi}(y)$$

By an envelope theorem, Theorem 3 of Milgrom and Segal (2002), the above inequality is equivalent to

$$m(\lambda')f_y^1 < (m(\lambda^*) - \lambda^* m'(\lambda^*))\, f_y^1 + \left(\phi(\zeta^*\lambda^*, \lambda^*) - \lambda^o \frac{\partial \phi(\zeta^*\lambda^*, \lambda^*)}{\partial \lambda}\right) \Delta f_y \tag{50}$$

where $\lambda' = \lambda^o(0, y^*)$ and $\lambda^* = \lambda^o(\zeta^*, y^*)$. Note that to apply the envelope theorem we need $\Pi^*(\zeta, y)$ be equidifferentiable in y, which is trivial because $\phi(\zeta\lambda, \lambda) - \lambda\frac{\partial\phi(\zeta\lambda,\lambda)}{\partial\lambda}$ is always between 0 and 1, which follows from the fact that for a given ζ, $\phi(\zeta\lambda, \lambda)$ is strictly concave in λ.

Note that $m(\lambda) = \lambda/(1+\lambda)$. By equation (49), $\Pi^*(0, y^*) = \left(\frac{\lambda'}{1+\lambda'}\right)^2 f^1$. Therefore, equation (50) can be rewritten as

$$\sqrt{\frac{\Pi^*(\zeta^*, y^*)}{f^1}} f_y^1 < \Pi_y^*(\zeta^*, y^*) \tag{51}$$

where we have used the fact $\Pi^*(0, y^*) = \Pi^*(\zeta^*, y^*)$ and $\Pi_y^*(\zeta^*, y^*) \equiv \frac{\partial}{\partial y}\Pi^*(\zeta^*, y^*)$ is the right-hand side of equation (50). Plugging the expression of $\phi(\mu, \lambda)$ into the above inequality shows that it is equivalent to

$$\mathcal{T}(\zeta^*) = -\Delta f (f_y^1)^2 (1 - (1-\zeta^*)\sigma) + \Delta f_y^2 \zeta f^1 + 2\Delta f_y f^1 f_y^1 (1 - (1-\zeta^*)\sigma\frac{\lambda^*}{1+\lambda^*}) > 0$$

Note that the above expression is linear in ζ^*. To prove it is strictly positive or negative we just need to prove that $\mathcal{T}(\zeta^* = 0)$ and $\mathcal{T}(\zeta^* = 1)$ are strictly positive or negative. For this, we have

$$\begin{aligned}
\mathcal{T}(\zeta^* = 1) &= \Delta f_y^2 f^1 + 2\Delta f_y f^1 f_y^1 - \Delta f (f_y^1)^2 \\
\mathcal{T}(\zeta^* = 0) &= f_y^1 (2\Delta f_y f^1 (1 - \sigma\frac{\lambda^*}{1+\lambda^*}) - \Delta f f_y^1 (1-\sigma))
\end{aligned}$$

In the following we will prove that if $\underline{\rho} \geq (1+\sigma)/2$, $\mathcal{T}(\zeta^* = 0)$ and $\mathcal{T}(\zeta^* = 1)$ are strictly positive, and if $\overline{\rho} \leq (1-\sigma)/2$, then both are strictly negative. This then finished our proof of Step 2 and hence Proposition 4.

We first consider $\mathcal{T}(\zeta^* = 1)$. It is easy to see that

$$\Delta f_y^2 f^1 + 2\Delta f_y f^1 f_y^1 - \Delta f (f_y^1)^2 = f^1 (f_y^1)^2 \left(\left(\frac{\Delta f_y}{f_y^1} + 1 \right)^2 - \frac{\Delta f}{f^1} - 1 \right)$$

If $\underline{\rho} \geq (1+\sigma)/2$, then by (18) we have

$$\mathcal{T}(\zeta^* = 1) \geq f^1 (f_y^1)^2 ((1+\kappa)^{2\underline{\rho}} - 1 - \kappa) > 0$$

where $\kappa = \kappa(y^*)$ and the last inequality holds because $\underline{\rho} > 1/2$.

If $\overline{\rho} \leq (1-\sigma)/2$, then

$$\mathcal{T}(\zeta^* = 1) \leq f^1 (f_y^1)^2 ((1+\kappa)^{2\overline{\rho}} - 1 - \kappa) < 0$$

where the last inequality holds because $\overline{\rho} < 1/2$.

Next, consider $\mathcal{T}(\zeta^* = 0)$. Since f is increasing in x and y, $\mathcal{T}(\zeta^* = 0) > 0$ if and only if

$$\frac{2f^1 \Delta f_y}{\Delta f f_y^1} > \frac{1-\sigma}{1-\sigma\frac{\lambda^*}{1+\lambda^*}}$$

Note that the supreme of the right-hand side is 1 (when $\lambda^* \to 0$) and the infimum is $1-\sigma$ (when $\lambda^* \to \infty$).

If $\underline{\rho} \geq (1+\sigma)/2$, then by (18) a sufficient condition for the above inequality is

$$2\frac{(1+\kappa)^{\frac{1+\sigma}{2}}-1}{\kappa} > 1$$

Since the left hand side of the above inequality is decreasing in κ and by equation (8), $\kappa < 4\sigma/(1-\sigma)^2$. Then by setting $\kappa = 4\sigma/(1-\sigma)^2$, the above inequality becomes

$$(1+\sigma)^{1+\sigma}(1-\sigma)^{1-\sigma} \geq (1+\sigma^2)$$

At $\sigma = 0$, both sides equal 2. For $\sigma > 0$: the derivative of the logarithm of the left-hand side is $2 + \log(1+\sigma) - \log(1-\sigma)$, which is strictly bigger than $2\sigma/(1+\sigma^2)$, the derivative of the logarithm of the right-hand side.

If $\overline{\rho} \leq (1-\sigma)/2$, then by similar logic, a sufficient condition for $\mathcal{T}(\zeta^* = 0) < 0$ is

$$2\frac{(1+\kappa)^{\frac{1-\sigma}{2}}-1}{\kappa} \leq 1-\sigma$$

As before, the left-hand side is decreasing in κ, and at $\kappa = 0$ reaches its supreme $1-\sigma$. Thus the above inequality holds. $\square$

A.10 Proof of Proposition 5

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 (5) and (37), the market utility condition can be written as

$$\frac{1+(1-\sigma)\lambda}{(1+\lambda)(1+(1-\sigma+\sigma\zeta)\lambda)}w_1 = U_1 \tag{52}$$

$$\frac{1}{1+(1-\sigma+\sigma\zeta)\lambda}w_2 = U_2 \tag{53}$$

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

$$\frac{w_1}{U_1} = \frac{w_2}{U_2}\left(1+\frac{\left(\frac{w_2}{U_2}-1\right)\sigma}{(1-\sigma)\frac{w_2}{U_2}+\zeta\sigma}\right) \tag{54}$$

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

$$\frac{(\frac{w_2}{U_2})^2}{(1-\sigma)\frac{w_2}{U_2}+\sigma} < \frac{w_1}{U_1} < \frac{\frac{w_2}{U_2}-\sigma}{1-\sigma} \tag{55}$$

where the term on the left (right) is obtained by setting $\zeta = 1$ ($\zeta = 0$) on the right-hand side of equation (54).

For case ii), we have $\zeta = 0$. Setting $\zeta = 0$ in equations (52) and (53) 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 (53) with "≤" 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

$$\frac{\frac{w_2}{U_2}-\sigma}{1-\sigma} \leq \frac{w_1}{U_1} \tag{56}$$

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

$$\frac{(\frac{w_2}{U_2})^2}{(1-\sigma)\frac{w_2}{U_2}+\sigma} \geq \frac{w_1}{U_1} \tag{57}$$

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.11 Proof of Lemma 7

Our proof does not rely on the specific meeting technology that we consider. Below, we consider a general meeting technology $\phi(\mu, \lambda)$. The idea for this proof is from Shimer (2005) and it can be easily extended to the case with any number of worker types.

Because $\phi(\mu, \lambda)$ is concave in μ, we have

$$\phi_\mu(\lambda^o, \lambda^o) < \psi_1(\zeta^*, \lambda^o) < \phi_\mu(\zeta^*\lambda^o, \lambda^o) < \psi_2(\zeta^*, \lambda^o) < \phi_\mu(0, \lambda^o).$$

As a result, the wages must satisfy

$$\begin{aligned} w_1^* &= \frac{U_1}{\psi_1(\zeta^*, \lambda^o)} \geq \frac{U_1}{\phi_\mu(\zeta^*\lambda^o, \lambda^o)} \\ w_2^* &= \frac{U_2}{\psi_2(\zeta^*, \lambda^o)} < \frac{U_2}{\phi_\mu(\zeta^*\lambda^o, \lambda^o)}. \end{aligned}$$

Moreover, by equation (14), we have

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

Therefore,

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

□

A.12 Proof of Lemma 7

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 Lemma 1, the probability that firms hire a type-type worker is

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

which is exactly the equation given in proposition 7.

Alternatively, we can prove the same claim by using equation 42. In this case, we have

$$\phi(\mu, \lambda) = 1 - \sum_{n=0}^{\infty} P_n(\widetilde{\lambda}) \left(1 - \frac{\mu}{\widetilde{\lambda}}\right)^n,$$

where $P_n(\widetilde{\lambda})$ is given by equation (42). □

A.13 Proof of Lemma 8

Since $m(\mu)$ is concave and increasing, $\epsilon_f(\mu)$ is always positive and $\epsilon_w(\mu)$ is always negative, which implies that both $a^c(\mu, \lambda)$ and $a^m(\mu, \lambda)$ are positive.

We first prove that $m'(0)$ is strictly positive but finite. Since $m(\mu)$ is concave and increasing, $m'(\mu)$ is positive and decreasing and reaches the maximum at $\mu = 0$. If $m'(0) = 0$, then $m'(\mu) = 0$ for all μ and $m(\mu) = 0$ for all μ. We have a contradiction. Next, $m'(0) = 1 - Q_0(0) \leq 1$. So $0 < m'(0) \leq 1$. Since $m(0) = 0$, $\lim_{\mu \to 0} \mu/m(\mu) = 1/m'(0)$ by L'Hopital's rule. Thus, $\lim_{\mu \to 0} \epsilon_f(\mu) = \lim_{\mu \to 0} \mu m'(\mu)/m(\mu) = \lim_{\mu \to 0} \mu/m(\mu) \cdot \lim_{\mu \to 0} m'(\mu) = 1$.

Before we prove that $\lim_{\mu \to 0} \epsilon_w(\mu) = 0$, we need the following simple mathematical result: $(n+1)x(1-x)^n \leq 1$ for any $x \in [0,1]$ and $n \in \mathbb{N}_0$. When $n = 0$ and 1, the result is trivially true. For $n \geq 2$, FOC implies that for a given n, the maximum is reached when $x = 1/(n+1)$. The corresponding maximum is $(1 - \frac{1}{n+1})^n \leq 1$.

Next we prove that for general contact technologies (not necessarily invariant), $\lim_{\mu \to 0} \mu \phi_{\mu\mu}(\mu, \lambda) = 0$. Since for a given λ, $\phi(\mu, \lambda)$ is analytic in μ, by equation (**??**) we have $\phi_{\mu\mu}(\mu, \lambda) = -\sum_{n=2}^{\infty} Q_n(\lambda)(n-1)\frac{1}{\lambda}(1 - \frac{\mu}{\lambda})^{n-2}$. Therefore, $\mu \phi_{\mu\mu}(\mu, \lambda) = -\sum_{n=2}^{\infty} Q_n(\lambda)(n-1)\frac{\mu}{\lambda}(1 - \frac{\mu}{\lambda})^{n-2}$. By the above result, $Q_n(\lambda)(n-1)\frac{\mu}{\lambda}(1 - \frac{\mu}{\lambda})^{n-2} \leq Q_n(\lambda)$. Since $\sum_2^{\infty} Q_n(\lambda) \leq 1$ and as $\mu \to 0$, $(n-1)\frac{\mu}{\lambda}(1 - \frac{\mu}{\lambda})^{n-2} \to 0$ for any given n and λ, $\lim_{\mu \to 0} \mu \phi_{\mu\mu}(\mu, \lambda) = 0$ by the dominated convergence theorem.

Since for invariant contact technologies, $m(\mu) = \phi(\mu, \lambda)$, we have $\lim_{\mu \to 0} -\mu m''(\mu) = 0$. As shown above, $m'(0) \in (0, 1]$. Thus we have $\lim_{\mu \to 0} \epsilon_w(\mu) = \lim_{\mu \to 0} \mu m''(\mu)/m'(\mu) = 0$. □

A.14 Proof of Proposition 9

The necessity follows directly from Proposition 3. We only need to consider the sufficiency part.

When the contact technology is invariant, $\phi(\mu, \lambda)$ depends only on μ, which implies that $S(\mu, \lambda, y)$ defined in equation (6) is strictly concave in (μ, λ). Thus for each firm y, the optimal choice of (μ, λ) must be unique. Therefore, $(\mu(y), \lambda(y))$ or equivalently $(\zeta(y), \lambda(y))$ must be continuous and whenever $\zeta(y)$ is interior, they must be differentiable too. Thus the local condition in (32) is also sufficient. Since for invariant contact technologies, $\phi_\lambda(\mu, \lambda) = 0$ for any μ and λ, (32) becomes

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq a^i$$

We first consider the PAC (PAM resp.) case, where by assumption $\underline{\rho} \geq \overline{a}^c$ ($\overline{a}^m$ resp.). Note that for invariant contact technologies, $\overline{a}^c, \overline{a}^m \geq 1$ by setting $\mu = \lambda$ in equation (41). By (18), we have

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

where the second inequality is because of $\underline{\rho} \geq 1$.

Next, consider the NAC/NAM case, where we have assumed that $\overline{\rho} \leq 0 = \underline{a}^c = \underline{a}^m$. Again by (18), we have

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

Thus we have proved our claim. $\square$

References

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

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.

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.

Cai, X., Gautier, P., and Wolthoff, R. (2019). Screening frictions and labor market segmentation. 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. (2010a). Sorting and decentralized price competition. *Econometrica*, 78:539–574.

Eeckhout, J. and Kircher, P. (2010b). Sorting versus screening - search frictions and competing mechanisms. *Journal of Economic Theory*, 145:1354–1385.

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

Epstein, L. and Peters, M. (1999). A revelation principle for competing mechanisms. *Journal of Economic Theory*, 88(1):119–161.

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.

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.