I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14501

Search, Screening and Sorting

Xiaoming Cai
Pieter Gautier
Ronald Wolthoff

JUNE 2021

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14501

Search, Screening and Sorting

Xiaoming Cai
Peking University

Pieter Gautier
VU University Amsterdam, Tinbergen Institute and IZA

Ronald Wolthoff
University of Toronto and IZA

JUNE 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Search, Screening and Sorting*

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 sorting in applications as well as matches. We show that positive sorting is obtained when production complementarities outweigh a force against sorting measured by a *quality-quantity elasticity*. Interestingly, we find that the required degree of production complementarity for positive sorting is *increasing* in the number of interviews: it ranges from square-root-supermodularity if each firm can interview a single applicant to log-supermodularity if each firm can interview all its applicants.

JEL Classification: C78, D82, D83, E24

Keywords: sorting, complementarity, search frictions, information frictions, heterogeneity

Corresponding author:
Pieter Gautier
Department of Economics
VU Amsterdam
De Boelelaan 1105
NL-1081 HV Amsterdam
The Netherlands

E-mail: p.a.gautier@vu.nl

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

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. That is, does the extent to which firms can interview workers affect whether 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. The earliest work on assignment problems (Tinbergen, 1956; Shapley and Shubik, 1971; Becker, 1973; Rosen, 1974) considers frictionless environments with no role for screening because there is full information about types. More recent work by Shimer and Smith (2000), Shi (2001, 2002), Shimer (2005) and Eeckhout and Kircher (2010a) allows for frictions but makes particular assumptions about the available information in the matching process and does not explore how outcomes depend on these assumptions.[3]

To answer our question, we therefore 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 interview multiple (but not necessarily all) applicants before making a job offer to the most profitable candidate. We show how the equilibrium allocation of workers to firms in this environment depends on the degree of production complementarities on the one hand and the extent to which firms can interview applicants on the other hand. Perhaps surprisingly, we find that reducing frictions by allowing firms to interview more workers is a force *against* sorting.

To explain this result, we must first describe our setup in more detail. We consider a static environment in which heterogeneous firms compete for heterogeneous workers by posting menus of type-contingent wages. Workers direct their search to the menu that maximizes their expected payoff. This choice determines the expected number of

[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 subject all applicants to an online job test. Based on their answers, every applicant is assigned a score, 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.

applicants (the 'queue length') at each firm, although the actual number is stochastic due to coordination frictions. As mentioned, the key novelty in our setup is that we allow firms to interview a subset of applicants, which reveals their types. Firms hire the most profitable candidate among their interviewees and subsequently produce output according to a general production function.

Firms in this environment face a trade-off. Attracting low-type applicants can be beneficial because the search frictions imply that it is always possible that no high type applies, in which case hiring a low type is better than remaining unmatched. However, this kind of insurance comes at a cost, because the presence of low types makes it harder for the firm to identify the high types in the applicant pool. Clearly, the magnitude of the cost is smaller if firms can screen more, so firms' decision what applicant pool to attract ex ante depends on the extent to which they can screen workers ex post.

We start our analysis by establishing that a firm's problem can be rewritten as one in which it purchases queues of applicants at prices equal to workers' expected payoffs. This reformulation implies that the market equilibrium is constrained efficient and simplifies exposition. Firms will purchase queues of applicants such that each worker's marginal contribution to surplus equals his marginal cost. An applicant directly contributes to surplus if no other applicant with the same or better type is being interviewed. However, when firms cannot screen everyone, an applicant also affects surplus by making it harder for other applicants to be interviewed.

We then turn to sorting. Given the meaningful distinction between applicants and hires in our environment, we analyze sorting along both dimensions. We define PAM as first-order stochastic dominance in the distribution of hires, and introduce *positive assortative contacting* (PAC) as the corresponding concept for the distribution of applicants.[4] We first show that when worker types are sufficiently close the conditions for PAC and PAM boil down to a simple comparison between two elasticities: along the equilibrium path, the *elasticity of complementarity* of the production function must exceed a novel elasticity which we label the *quality-quantity elasticity* (and which differs between PAC and PAM). It therefore follows that we need to consider the bounds on the two elasticities. In particular, the supremum of the quality-quantity elasticity should not exceed the infimum of the elasticity of complementarity. We

[4] We also provide results for negative assortative contacting (NAC) and NAM. We omit intuition for those results here as it mirrors the intuition for PAC and PAM.

prove that this condition is not only necessary but also sufficient.

The elasticity of complementarity is a scalar measure, which can be used to describe varying degrees of supermodularity. The relevance of complementarity in production for sorting has been known since Becker (1973) showed that supermodularity of the production function (i.e. the elasticity of complementarity being positive) is necessary and sufficient for PAM in a frictionless economy. It is also well known that this condition is no longer sufficient in the presence of search frictions. The reason is that the opportunity cost of remaining unmatched is larger for high types, which makes them more eager to match with a low type rather than run the risk to not match at all. To undo this effect, the production function must exhibit stronger complementarities. For example, Shimer and Smith (2000) derive a set of conditions for PAM under random search that are even stronger than log-supermodularity (i.e. the elasticity of complementarity being larger than 1).

Most related to our work, Eeckhout and Kircher (2010a) show that under directed search (but with a single interview per firm) PAM requires that the elasticity of complementarity exceeds the elasticity of substitution of the aggregate meeting function. This latter elasticity is positive but bounded above by $\frac{1}{2}$ for common meeting technologies, making square-root-supermodularity of the production function (i.e. an elasticity of complementarity equal to $\frac{1}{2}$) sufficient for PAM.[5] Intuitively, while frictions still exist under directed search, they are weaker than under random search, because high types can avoid meeting low types.

In our environment with simultaneous interviews, the relevant threshold for sorting is the *quality-quantity elasticity*. Like the threshold in Eeckhout and Kircher (2010a), this elasticity depends on the properties of the meeting technology only. However, a crucial difference is that the quality-quantity elasticity depends not only on the queue length but also on the queue composition and the degree of screening. To better understand this elasticity, note that when worker types are sufficiently close, equilibrium requires more-productive firms to have longer queue lengths. This longer queue length reduces the probability that a high-type applicant creates surplus, which discourages more-productive firms from attracting such applicants and therefore forms a force against sorting. The quality-quantity elasticity measures the magnitude of this drop: the larger it is, the stronger the force against sorting and the larger production

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

complementarities therefore need to be to offset this force and induce positive sorting.

To see the dependence of the quality-quantity elasticity on the queue length and queue composition, consider the case where high-type workers are abundant in the economy. The probability that a high-type worker creates surplus decreases then more quickly along the equilibrium path, as productive firms with long queue lengths are increasingly likely to interview multiple high-type applicants. Hence, the force against sorting, as measured by the quality-quantity elasticity, is highest in this case.

Subsequently, we show that the supremum of the quality-quantity elasticity increases with the degree of screening. Viewing increased screening as a relaxation of the frictions in the environment, one may have expected from the literature cited above that it must facilitate sorting. We show that this intuition is wrong. The logic again follows from the above scenario in which the probability that a high-type worker creates surplus decreases along the equilibrium path. This force against is amplified by increased screening as that further increases the probability that more-productive firms interview multiple high-type applicants. The elasticity of complementarity that is necessary and sufficient for PAC and PAM is *increasing* in the expected number of interviews that firms can conduct, ranging from $\frac{1}{2}$ (square-root-supermodularity) with a single interview to 1 (log-supermodularity) when firms can interview all their applicants.

The paper is organized as follows. The remainder of this section discusses related literature. Section 2 introduces the model. Section 3 considers the market equilibrium and establishes that it is constrained efficient. Section 4 derives our main sorting results. In Section 5, we consider various extensions, including noisy signals for every applicant and endogenous choice of screening capacity. For the latter extension, we show that firms in the middle of the productivity distribution (rather than the most productive ones) have the strongest incentives to invest in screening. Finally, Section 6 concludes, while proofs and additional results can be found in the (online) appendix.

Related Literature. Our results do not only contribute to the theoretical literature referenced above, but also have important implications for the empirical literature that deals with both the sign and the strength of sorting (Gautier and Teulings, 2006; Eeckhout and Kircher, 2011; Gautier and Teulings, 2015; Lise et al., 2016; Hagedorn et al., 2017; Lopes de Melo, 2018; Bartolucci et al., 2018; Bagger and Lentz, 2018; Borovičková and Shimer, 2020). An important aim of this literature is to identify the shape of the production function from observed matching patterns.

In general, a particular meeting technology is assumed and then the strength and sign of sorting are used to identify key parameters of the production function.[6] Our findings imply however that such assumptions are not innocuous and that the meeting technology needs to be identified alongside the production function. Progress along this dimension is facilitated by our theoretical results on PAC/PAM combined with recent empirical work by Banfi et al. (2020) who document evidence for PAC as well as PAM using data from a Chilean online job board.

The strength of sorting is often used to estimate how far an economy is from the frontier. Our results show that stronger sorting patterns do not necessary imply lower frictions.

Some papers have argued that increased sorting of high-type workers at high-wage firms has contributed to the observed increased inequality from the mid-nineties onwards (see e.g. Card et al., 2013; Song et al., 2019).[7] Håkanson et al. (2018) argue that the increased sorting patterns are mainly due to increasing complementarities in production. Our results suggest that if during the same period, new technologies like automated resume screening made it easier to screen workers, then this would require even stronger complementarities in the production technology.

Finally, although our focus is on the labor market, our results are also important for other markets with matching between heterogeneous agents and a role for screening, such as the housing market or the 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).

2 Model

2.1 Environment

Agents. A static economy is populated by a continuum of risk-neutral firms and workers. Each firm demands and each worker supplies a single unit of indivisible labor. Both types of agents are heterogeneous. In particular, each firm is characterized by a type $y \in \mathcal{Y} = [\underline{y}, \overline{y}] \subset \mathbb{R}_+$. The measure of firms with types less or equal to y is

[6] Since wages for a given worker type are typically non-monotonic in firm types, the methodology by Abowd et al. (1999) of detecting sorting patterns from simply correlating worker and firm fixed effects fails; the cited papers propose various ways to deal with this.

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

denoted by $J(y)$, where the total measure $J(\overline{y})$ is normalized to one. Similarly, each worker is characterized by a type $x \in \mathcal{X} = [\underline{x}, \overline{x}] \subset \mathbb{R}_+$, which initially is private information, but can be learned by a firm during an interview, as we describe in more detail below. There are two types of workers: a low type x_1 and a high type x_2, with $0 < x_1 < x_2$.[8] The measure of workers with type x_i is denoted by $\ell_i > 0$. The resource set or the *endowment of agents* in the economy is thus given by $(x_1, x_2, \ell_1, \ell_2, J(y))$.

Wage Menus and Search. Each firm commits to a wage menu $\boldsymbol{w} = (w_1, w_2)$, where w_i is the wage for a hire of type x_i. Workers observe all wage menus and apply to one, taking into account that there will be more competition at high wages.[9] We initially assume that workers also observe firm types, but then show that this assumption is redundant because workers only care about their expected payoff, which in equilibrium they can infer from the wage alone. We capture the anonymity of the large market with the standard assumption that identical workers must use symmetric strategies (see e.g. Shimer, 2005).

A *submarket* $(\boldsymbol{w}, y)$ consists of the firms of type y that post a wage menu $\boldsymbol{w}$ and all workers who apply to such a menu. For each submarket, we denote the ratio of the number of high-type applicants to the number of firms by $\mu(\boldsymbol{w}, y)$, and the ratio of the total number of applicants (regardless of their type) to the number of firms by $\lambda(\boldsymbol{w}, y)$. Naturally, these ratios—or *queue lengths*—satisfy $0 \leq \mu(\boldsymbol{w}, y) \leq \lambda(\boldsymbol{w}, y)$ for all $(\boldsymbol{w}, y)$.[10] For future reference, define $\zeta(\boldsymbol{w}, y) = \mu(\boldsymbol{w}, y)/\lambda(\boldsymbol{w}, y)$ as the fraction of high-type applicants in submarket $(\boldsymbol{w}, y)$.

Benchmark Frictions. The matching process within a submarket is frictional and exhibits constant returns to scale, in the sense that outcomes only depend on queue lengths rather the absolute measures of workers and firms. Within those boundaries, we can allow for a fairly wide class of matching processes, but we initially focus on a specific benchmark with two stages (applying and screening) to simplify exposition.[11]

To introduce the benchmark, consider a particular submarket with queues (μ, λ).

[8] In Appendix B.10 we show how our results can be generalized to N worker types for a widely used class of meeting technologies, that includes the urn-ball and geometric.

[9] A single chance to match (per period) is standard and captures the idea that (opportunity) costs are associated with applying. The 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; Albrecht et al., 2019). An exception is Auster et al. (2020).

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

[11] In section 5, we consider various generalizations.

Workers and firms in the submarket are randomly located on the circumference of a circle according to a uniform distribution. Workers apply clockwise to the nearest firm.[12] A firm therefore receives n applications with probability $\frac{1}{1+\lambda}(\frac{\lambda}{1+\lambda})^n$ for $n = 0, 1, 2, \ldots,$ which is a geometric distribution with mean λ.[13] In the screening stage, firms interview their applicants in a random order. An interview reveals the type of the applicant, which is x_2 with probability μ/λ. After every interview, and conditional on applicants remaining, there is an exogenous probability $\sigma \in [0, 1]$ that the firm can conduct another interview, while interviewing stops with complementary probability.

Our setup nests two common but extreme specifications of the meeting technology as special cases. If $\sigma = 0$, each firm can interview only a single applicant, as in the bilateral model of Eeckhout and Kircher (2010a). In this case, the presence of low-type applicants makes it harder for firms to identify a high type in their applicant pool. Increasing σ reduces this meeting externality. It disappears entirely when σ reaches 1 and firms can interview all their applicants. As in the urn-ball setup of Shimer (2005), firms' chances of finding a high type in their applicant pool then become independent of the number of low-type applicants—a property known in the literature as *invariance* (see Lester et al., 2015; Cai et al., 2017).

Matching and Production. After the interviews have been conducted, matches are formed. Firms can only hire a worker which they have interviewed.[14] If a firm has interviewed multiple applicants, it hires the most profitable one. A match between a worker of type x and a firm type of y produces output $f(x, y) > 0$, which is twice continuously differentiable. The partial derivatives $f_x(x, y)$ and $f_y(x, y)$ are strictly positive for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and the cross-partial is denoted by $f_{xy}(x, y)$. From the produced output, the firm pays the worker the promised wage w_i and keeps the rest. Firms and workers which fail to match obtain a zero payoff.

Elasticity of Complementarity. For our analysis, a key characteristic of the production function is its *elasticity of complementarity* (Hicks, 1932, 1970), which is usually defined for constant-returns-to-scale production functions, and is the inverse

[12] When workers cannot keep track of the distance they travel, this is merely a tie-breaking rule.

[13] Note the subtle difference compared to an equidistant positioning of firms, which yields a Poisson number of applicants with mean λ, as in an urn-ball technology. We discuss this case in Section 5.1.

[14] 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.

of the elasticity of substitution. For general production functions, it is defined as

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

with extrema $\overline{\rho} \equiv \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x,y)$ and $\underline{\rho} \equiv \inf_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \rho(x,y)$. For future reference, note that $\rho(x,y)$ is the relative percentage change in marginal output by firms, $f_y(x,y)$ in response to a marginal change in $f(x,y)$ (elasticity) caused by increasing the worker type to $x+\Delta x$. That is, for sufficiently small $\Delta x > 0$, we have

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

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

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

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

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

Proof. See Appendix A.1. □

Supermodularity. The elasticity of complementarity $\rho(x,y)$ is closely related to the notion of n-root-supermodularity, as defined in Eeckhout and Kircher (2010a).[15]

Definition 1. *The function $f(x,y)$ is* n-root-supermodular *on $\mathcal{X}\times\mathcal{Y}$ if and only if $\rho(x,y) \geq 1-1/n$, for all $(x,y)\in\mathcal{X}\times\mathcal{Y}$; special cases include supermodularity ($n=1$) and log-supermodularity ($n\to\infty$). When $\rho(x,y) \leq 1-1/n$, $f(x,y)$ is said to be* n-root-submodular *on $\mathcal{X}\times\mathcal{Y}$.*

[15] Eeckhout and Kircher (2010a) define $f(x,y)$ to be n-root-supermodular if $\sqrt[n]{f(x,y)}$ is supermodular. Since $\frac{1}{\partial x\partial y}\sqrt[n]{f} = n^{-2}f^{1/n-2}\left(ff_{xy} - (1-\frac{1}{n})f_xf_y\right)$, our definition is equivalent.

In other words, n-root-supermodularity is equivalent to $\underline{\rho} \geq 1 - 1/n$ and n-root-submodularity is equivalent to $\overline{\rho} \leq 1 - 1/n$.[16]

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

Beliefs. A firm of type y posting a wage menu $\boldsymbol{w}$ has to form beliefs about its queues $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$. Following the standard approach in the literature, we restrict these beliefs in the spirit of subgame perfection through what is known as the *market utility condition* (see e.g. Eeckhout and Kircher, 2010b). To state this condition, consider a worker of type x_i. Define $V_i(\boldsymbol{w}, \mu, \lambda, y)$ as his expected payoff in a submarket $(\boldsymbol{w}, y)$ with queues (μ, λ), and his *market utility* U_i as the maximum expected payoff that he can obtain in equilibrium, either by visiting one of the submarkets or by remaining inactive. Firms' beliefs $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ must then satisfy

$$\begin{cases} V_1(\boldsymbol{w}, \mu, \lambda, y) \leq U_1, \text{ with equality if } \lambda - \mu > 0, \\ V_2(\boldsymbol{w}, \mu, \lambda, y) \leq U_2, \text{ with equality if } \mu > 0. \end{cases} \tag{3}$$

For common meeting technologies, including our benchmark as we will show in Lemma 3 below, (3) admits a unique solution (μ, λ), which is then the firm's belief. For other technologies, there can be multiple solutions to (3). The standard assumption is then that firms are optimistic and expect the solution that maximizes their expected payoff. We denote this expected payoff by $\pi(\boldsymbol{w}, \mu, \lambda, y)$.[17]

Strategies. Let $G(\boldsymbol{w} \,|\, y)$ denote the (conditional) probability that a firm of type y offers a wage menu $\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}$, where $\widetilde{\boldsymbol{w}} = (\widetilde{w}_1, \widetilde{w}_2)$, $\boldsymbol{w} = (w_1, w_2)$, $\widetilde{w}_1 \leq w_1$ and $\widetilde{w}_2 \leq w_2$. Given market utilities (U_1, U_2), firm optimality means that $G(\boldsymbol{w} \,|\, y)$ must maximize $\pi(\boldsymbol{w}, \mu, \lambda, y)$ subject to the constraint (3).

Similarly, let $H_i(\boldsymbol{w}, y)$ denote the probability that workers of type x_i apply to a firm with $\widetilde{\boldsymbol{w}} \leq \boldsymbol{w}$ and $\widetilde{y} \leq y$. The following accounting identities then link workers'

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

[17] Explicit expressions for π and V_i will be provided in Section 3.1.

strategies $H_1(\boldsymbol{w}, y)$ and $H_2(\boldsymbol{w}, y)$ to the queues in different submarkets.

$$H_2(\boldsymbol{w}, y) = \frac{1}{\ell_2} \int_{\widetilde{y} \le y} \int_{\widetilde{\boldsymbol{w}} \le \boldsymbol{w}} \mu(\widetilde{\boldsymbol{w}}, \widetilde{y}) \, dG(\widetilde{\boldsymbol{w}} \,|\, \widetilde{y}) \, dJ(\widetilde{y}), \tag{4}$$

$$H_1(\boldsymbol{w}, y) = \frac{1}{\ell_1} \int_{\widetilde{y} \le y} \int_{\widetilde{\boldsymbol{w}} \le \boldsymbol{w}} [\lambda(\widetilde{\boldsymbol{w}}, \widetilde{y}) - \mu(\widetilde{\boldsymbol{w}}, \widetilde{y})] \, dG(\widetilde{\boldsymbol{w}} \,|\, \widetilde{y}) \, dJ(\widetilde{y}). \tag{5}$$

Worker optimality requires that workers must obtain exactly U_i at any firm to which they apply with positive probability, and weakly less at other firms i.e. (3) must hold. Further, note that no firm will post a wage menu $\boldsymbol{w} \ge \overline{\boldsymbol{w}} \equiv (f(x_1, \overline{y}), f(x_2, \overline{y}))$. Thus, $H_i(\overline{\boldsymbol{w}}, \overline{y})$ is the probability that workers of type x_i apply, which must equal 1 if $U_i > 0$, as the payoff from not sending an application is zero. This condition can be interpreted as "market clearing": in equilibrium, demand for each type of applicant must equal supply, which determines the "market prices" U_1 and U_2.

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

Definition 2. *A (directed search) equilibrium is a triple $(G, \{H_1, H_2\}, \{U_1, U_2\})$ satisfying ...*

(i) Firm Optimality. *Given (U_1, U_2), every wage menu $\boldsymbol{w}$ in the support of $G(\cdot \,|\, y)$ maximizes $\pi(\boldsymbol{w}, \mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y), y)$ for each firm type y, where the queue lengths $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ are determined by (3).*

(ii) Worker Optimality. *Given (U_1, U_2), the application strategy of high-type and low-type workers satisfies (4) and (5), respectively, where the queue lengths $(\mu(\boldsymbol{w}, y), \lambda(\boldsymbol{w}, y))$ are determined by (3). Further, $H_i(\overline{\boldsymbol{w}}, \overline{y}) = 1$ if $U_i > 0$.*

3 Equilibrium

We start our equilibrium analysis by establishing an equivalence: the problem of a firm in our environment is equivalent to the problem of a firm that can buy queues of applicants directly in a competitive market.[18] The consequence of this result is that the equilibrium in our environment is constrained efficient.

[18]Hence, the difference with a "conventional" competitive market is that the firm buys a distribution of applicants rather than directly hiring a particular type of worker.

3.1 Efficiency

Ranking. A caveat in the analysis is that firms' ranking of workers is endogenous. To simplify exposition, we initially assume that firms post wage menus satisfying

$$f(x_2, y) - w_2 > f(x_1, y) - w_1, \tag{6}$$

i.e. more productive workers are more profitable and are therefore preferred by firms. Later, in Lemma 4, we will show that this assumption is without loss of generality.

Interviewing Probability. Given (6), a firm will hire a high-type worker if and only if it interviews *at least* one such worker. The following lemma, which we borrow from Cai et al. (2020), derives the probability of this event.[19]

Lemma 2 (Cai et al., 2020)**.** *In a submarket with queues* (μ, λ)*, the probability that a firm interviews at least one high-type worker equals*

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

Proof. See Appendix B.1. □

The key insight of Cai et al. (2020) is that $\phi(\mu, \lambda)$ is useful for multiple reasons. First, $\phi(\mu, \lambda)$ is sufficient to summarize the meeting process within a submarket. Given (6), it not only describes the probability that the firm will hire a high-type worker, but—upon evaluation in $\mu = \lambda$—also the firm's overall matching probability (regardless of the hire's type), which we denote by $m(\lambda) \equiv \phi(\lambda, \lambda)$.

Second, the partial derivatives of $\phi(\mu, \lambda)$ have economically meaningful interpretations. The partial derivative $\phi_\lambda(\mu, \lambda) \leq 0$ captures externalities in the recruiting process as it describes how a firm's chances to hire a high-type worker change if the queue of low-type workers gets longer. As discussed before, these externalities are absent, i.e. $\phi_\lambda(\mu, \lambda) = 0$, if and only if all applicants are interviewed (i.e. $\sigma = 1$).

In contrast, $\phi_\mu(\mu, \lambda)$ describes how a firm's probability of hiring a high-type worker changes if the queue of such workers increases, while the total queue remains constant (i.e. changing the composition of the applicant pool). From the perspective of a high-

[19] Cai et al. (2020) study market segmentation in a world with homogeneous firms. Our focus is quite different, so we provide a derivation of $\phi(\mu, \lambda)$ for completeness.

type applicant, this partial derivative represents the probability that he or she is hired and increases surplus because no other high-type worker was interviewed.[20]

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

A0. $\phi(\mu, \lambda)$ is strictly increasing and concave in μ, i.e. replacing low-type workers with high-type workers in a submarket increases a firm's probability of interviewing at least one high-type worker, but at a decreasing rate;

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

A2. for any given $\zeta \in (0, 1]$, $\phi_\mu(\lambda\zeta, \lambda)$ is strictly decreasing in λ, i.e. holding the fraction of high-type workers constant, adding more workers to the submarket reduces the probability that a high-type worker creates surplus.

Surplus and Payoffs. To derive expected surplus, consider a firm of type y facing a queue (μ, λ). With probability $m(\lambda) \equiv \phi(\lambda, \lambda)$, the firm receives at least one application, hence generating at least a surplus $f(x_1, y)$; with probability $\phi(\mu, \lambda)$, the firm interviews at least one high-type worker, hence generating an additional surplus $f(x_2, y) - f(x_1, y)$. The expected surplus is therefore

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

By the same logic, the expected payoff of the firm equals

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

Finally, the expected payoff of applicants of type x_i is $V_i(\boldsymbol{w}, \mu, \lambda, y) = \psi_i(\mu, \lambda) w_i$, where, by a simple accounting identity, their matching probability $\psi_i(\mu, \lambda)$ equals

$$\psi_1(\mu, \lambda) = \frac{m(\lambda) - \phi(\mu, \lambda)}{\lambda - \mu} \qquad \text{or} \qquad \psi_2(\mu, \lambda) = \frac{\phi(\mu, \lambda)}{\mu}. \tag{10}$$

[20] To see this, note that $\phi_\mu(\mu, \lambda)\Delta\mu = \phi(\mu + \Delta\mu, \lambda) - \phi(\mu, \lambda)$ represents the probability that replacing $\Delta\mu$ low-type workers with high types generates additional surplus. Naturally, this is the case if and only if these $\Delta\mu$ workers are the only high types that are interviewed.

Uniqueness of Queues. In a submarket $(\boldsymbol{w}, y)$, the queues (μ, λ) are determined by the market utility condition (3). Since this is a system of non-linear equations, it is not immediate that there is a unique solution. Lemma 3 establishes this result.

Lemma 3. *If ϕ is given by* (7)*, there exists exactly one solution (μ, λ) to the market utility condition for any wage menu $\boldsymbol{w}$.*

Proof. See Appendix B.2. □

Competitive Market for Queues. As is standard in the literature, we can use the market utility condition (3) to substitute the wages w_1 and w_2 out of (9) and rewrite the firm's problem with queue lengths as choice variables. This yields

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

where, for use in Section 3.2, the arguments of the firm profit function Π are the fraction of high-type applicants μ/λ and the queue length λ.[21] Equation (11) has a straightforward interpretation: it is the payoff of a firm buying queues of low-type and high-type workers in a competitive market at prices equal to their respective market utilities. This formulation will be the starting point for our sorting analysis below.

Productivity versus Profitability. We have only considered wage menus satisfying (6). To see that this restriction is without loss of generality, suppose that a firm posts a wage menu where low-type workers yield a higher profit ex post, i.e. $f(x_2, y) - w_2 < f(x_1, y) - w_1$, and attracts a queue (μ, λ). Workers must again obtain their market utility, so that the expected transfer from the firm to the workers must be $\mu U_2 + (\lambda - \mu)U_1$. However, expected surplus must be strictly smaller than $S(\mu, \lambda, y)$ in (8). Accordingly, the firm's expected profit must be strictly smaller than the maximum profit in (11). The following lemma shows that any interior solution to (11) satisfies (6), so there is no need to consider wage menus that give priority to low-type workers because those wage menus always generate lower expected profit.[22] We omit the case in which a firm attracts only one type of workers; these corner solutions are trivial, since the firm can always prevent a certain type of workers from applying by offering them a zero wage.

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

[22] A similar result appears in Shimer (2005) for the case of urn-ball meetings. Our proof of Lemma 4 generalizes his result to arbitrary meeting technologies.

Lemma 4. *Let (μ^*, λ^*) with $0 < \mu^* < \lambda^*$ be an interior solution to the firm's problem (11). The corresponding wage menu $(w_1^*, w_2^*) = (U_1/\psi_1(\mu^*, \lambda^*),\ U_2/\psi_2(\mu^*, \lambda^*))$ then satisfies (6).*

Proof. See Appendix A.2. □

Observability of Firm Productivity. By Lemma 4, all firms will post wage contracts such that high-type workers are more profitable. Given the wage contract, the market utility condition then determines the queue length and composition. Since workers only care about their hiring probability and the wage, this then means that all our results carry through if they do not observe firm types.

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

Proposition 1. *The market equilibrium is constrained efficient.*

Marginal Contributions. Since the equilibrium is constrained efficient, the expected payoffs of firms and workers equal their marginal contribution to surplus. Adding more low-type workers to a submarket only increases λ, while adding more high-type workers increases both μ and λ. Thus, the marginal contribution of low-type and high-type workers at a firm of type y with queues (μ, λ) are $S_\lambda(\mu, \lambda, y)$ and $S_\mu(\mu, \lambda, y) + S_\lambda(\mu, \lambda, y)$, respectively. Because of constant returns to scale, the firm's marginal contribution is the difference between total surplus and the sum of the marginal contributions of its applicants, i.e. $S(\mu, \lambda, y) - \mu S_\mu(\mu, \lambda, y) - \lambda S_\lambda(\mu, \lambda, y)$.[23] Using $S(\mu, \lambda, y)$ from (8), $f^1 \equiv f(x_1, y)$ and $\Delta f = f(x_2, y) - f(x_1, y)$, we get

$$T_1(\mu, \lambda, y) = m'(\lambda) f^1 + \phi_\lambda(\mu, \lambda)\Delta f, \tag{12}$$

$$T_2(\mu, \lambda, y) = m'(\lambda) f^1 + \left(\phi_\mu(\mu, \lambda) + \phi_\lambda(\mu, \lambda)\right)\Delta f, \tag{13}$$

$$R(\mu, \lambda, y) = \left(m(\lambda) - \lambda m'(\lambda)\right) f^1 + \left(\phi(\mu, \lambda) - \mu\phi_\mu(\mu, \lambda) - \lambda\phi_\lambda(\mu, \lambda)\right)\Delta f, \tag{14}$$

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

[23] Alternatively, increase the number of firms by a factor $1 + \Delta s$. The additional surplus is then $(1 + \Delta s)S(\mu/(1 + \Delta s), \lambda/(1 + \Delta s), y) - S(\mu, \lambda, y)$, which yields the same result when $\Delta s \to 0$.

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

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

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

In the bilateral case $\sigma = 0$, we have $\phi(\mu, \lambda) = m(\lambda)\mu/\lambda$ such that $\phi_{\mu\mu} = 0$, which means that the Hessian is never negative definite and surplus is never concave at points (μ, λ) with $0 < \mu < \lambda$. Hence, firms will then attract only one type of workers. Below, we will therefore focus on the case $\phi_{\mu\mu} < 0$, i.e. $\sigma > 0$; the results will extend to the bilateral case by continuity.[24] Given $\phi_{\mu\mu} < 0$, the Hessian is negative definite if and only if its determinant is positive. Let $\kappa(y)$ be a measure of output dispersion, defined as the relative gain in output for a firm of type y from hiring a high- rather than a low-type worker, i.e.

$$\kappa(y) \equiv \frac{f(x_2, y) - f(x_1, y)}{f(x_1, y)} > 0. \tag{15}$$

Then, we have the following result.

Lemma 5. *Surplus $S(\mu, \lambda, y)$ is strictly concave at a point (μ, λ) with $0 < \mu < \lambda$ if*

$$\frac{1}{\kappa(y)} > \frac{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}}{-m''}, \tag{16}$$

Proof. See Appendix A.3. □

Note that $\phi_{\mu\mu}\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2$ is the determinant of the Hessian matrix of $\phi(\mu, \lambda)$. For our benchmark meeting process, it is zero if $\sigma = 1$ and strictly negative otherwise, making the right-hand side of (16) (weakly) positive.[25] Consequently, the concavity condition in (16) is satisfied when $\kappa(y)$ is sufficiently small and always holds when worker heterogeneity disappears $(x_1 \to x_2)$.

[24] We will revisit the bilateral case in Section 5.1.

[25] For other meeting processes, the determinant can be positive for any μ and λ, making the right-hand side of (16) negative, so that the firm's second-order condition is always satisfied. This is the case if and only if $\phi(\mu, \lambda)$ is jointly concave in (μ, λ). However, as discussed in Cai et al. (2017), such meeting processes feature positive meeting externalities, which are empirically less relevant for the labor market, so we abstract from them here.

3.2 Optimal Queue Length and Composition

In our sorting analysis below, the fraction of high-type applicants across firms will play an important role. It is therefore convenient to reformulate the firm's problem as follows: first, firms choose the fraction of high-type workers in their pool of applicants, which we denote by $\zeta \equiv \mu/\lambda \in [0,1]$; second, they choose the total queue length λ. In other words, the firms' problem is given by $\max_{\zeta\in[0,1]} \Pi^*(\zeta, y)$, where $\Pi^*(\zeta, y) \equiv \max_{\lambda\geq 0} \Pi(\zeta, \lambda, y)$ and $\Pi(\zeta, \lambda, y) = S(\zeta\lambda, \lambda, y) - \lambda U_1 - \zeta\lambda\,(U_2 - U_1)$, as defined in (11).

Optimal Queue Length. Working backwards, we first consider the choice of the queue length λ for a given $\zeta \in [0,1]$. Since $\phi(\zeta\lambda, \lambda)$ is strictly concave in λ for all $\zeta > 0$ and $m(\lambda) \equiv \phi(\lambda, \lambda)$, the payoff $\Pi(\zeta, \lambda, y)$ is *strictly* concave in λ for a given $\zeta \in [0,1]$. Thus, assuming that firms of type y are active in hiring, their optimal queue is unique and determined by the first-order condition (FOC)

$$U_1 + \zeta(U_2 - U_1) = m'(\lambda)\, f^1 + \frac{\partial\phi\,(\zeta\lambda, \lambda)}{\partial\lambda}\Delta f, \tag{17}$$

where $\partial\phi\,(\zeta\lambda, \lambda)\,/\partial\lambda \equiv \zeta\phi_\mu\,(\zeta\lambda, \lambda) + \phi_\lambda\,(\zeta\lambda, \lambda)$. Denote the optimal queue length by $\lambda^o(\zeta, y)$. To understand (17), note that the first term denotes the marginal contribution to surplus of a low-type applicant when all applicants are of a low type. The second term corrects for the fact that a fraction ζ of applicants has high productivity.

To understand how the optimal queue length varies with firm type, we can differentiate equation (17) with respect to y (holding ζ constant). This yields

$$\lambda^o_y(\zeta, y) = \frac{\partial\lambda^o(\zeta, y)}{\partial y} = -\frac{m' f^1_y + \frac{\partial\phi}{\partial\lambda}\Delta f_y}{m'' f^1 + \frac{\partial^2\phi}{\partial\lambda^2}\Delta f}, \tag{18}$$

where we have suppressed arguments from $m(\lambda^o(\zeta, y))$ and $\phi(\zeta\lambda^o(\zeta, y), \lambda^o(\zeta, y))$. Since $\phi(\zeta\lambda, \lambda)$ is strictly increasing and concave in λ for $\zeta > 0$, the numerator in (18) is positive if $\Delta f_y \geq 0$ and the denominator is negative. In other words, when the opportunity costs of remaining unmatched are larger for more productive firms (i.e. supermodularity of the production function), those firms are more willing to invest in longer queues (holding ζ constant). Another special case is when x_1 and x_2 are close, which implies that Δf_y is small, so that the sign of the numerator is dominated by $m' f^1_y$. In this case, we again have $\lambda^o_y(\zeta, y) > 0$: more productive firms prefer longer queues (keeping ζ constant).

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

Suppose first that $\zeta(y)$ is interior. With a slight abuse of notation, denote by $\lambda(y)$ the corresponding optimal queue length $\lambda^o(\zeta(y), y)$. Then $\zeta(y)$ and $\lambda(y)$ must satisfy,

$$\frac{\partial \Pi^*(\zeta, y)}{\partial \zeta}\Big|_{\zeta=\zeta(y)} = 0 \Leftrightarrow \phi_\mu(\zeta(y)\lambda(y), \lambda(y))\Delta f = U_2 - U_1, \tag{19}$$

where we used the envelope theorem and treated the total queue λ as constant in this exercise. The left-hand side of the above equation is exactly the difference between the marginal contribution to surplus of high-type and low-type workers—given by equations (13) and (12), respectively—while the right-hand side is the difference in their cost. Intuitively, a larger ζ increases the firm's probability of matching with a high-type worker, but comes at a cost as these workers are more expensive.[26] Finally, recall that, by Lemma 5, an interior solution ζ which satisfies the FOC (19) also satisfies the second-order condition (SOC) if equation (16) holds.

For future use, note that differentiating equation (17) with respect to ζ and evaluating the result at $\zeta = \zeta(y)$ gives

$$\lambda^o_\zeta(\zeta(y), y) = \frac{\partial \lambda^o(\zeta, y)}{\partial \zeta}\Big|_{\zeta=\zeta(y)} = -\frac{\lambda(y)\frac{\partial \phi_\mu}{\partial \lambda}\Delta f}{m'' f^1 + \frac{\partial^2 \phi}{\partial \lambda^2}\Delta f}, \quad \text{if } \zeta(y) \in (0, 1) \tag{20}$$

where we have used equation (19) to substitute out $U_2 - U_1$ and suppressed arguments from $m(\lambda(y))$ and $\phi(\zeta(y)\lambda(y), \lambda(y))$. For a given firm type y, this equation denotes the effect of a higher ζ on the optimal queue length, evaluating ζ at its equilibrium value $\zeta(y)$. It shows that along the equilibrium path, $\lambda^o_\zeta(\zeta(y), y)$ is negative since with a higher fraction of high-type workers, firms will reduce the total queue length to reduce negative hiring spillovers from low-productivity workers (recall $\partial \phi_\mu/\partial \lambda < 0$).

The optimal ζ does not need to be interior. In case of corner solutions, i.e.,

[26] 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 $\lambda\Delta\zeta$.

$\zeta(y) = 0$ or $\zeta(y) = 1$, the following FOCs must be satisfied:

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

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

where $\lambda(y) = \lambda^o(0, y)$ in equation (21), implicitly defined by $m'(\lambda)f(x_1, y) = U_1$, while $\lambda(y) = \lambda^o(1, y)$ in equation (22), implicitly defined by $m'(\lambda)f(x_2, y) = U_2$.

Multiplicity. As mentioned, a firm's optimal choice of the fraction of high-type workers ζ may not be unique, which generally renders the analysis intractable. However, Cai et al. (2020) show that under a single-crossing condition, which we present in Appendix B.3 and which is satisfied by our benchmark meeting technology, the optimal queue composition takes the following simple form.

Lemma 6 (Cai et al., 2020)**.** *If ϕ is given by (7), then for any given firm type y, $Z(y)$ contains at most two elements, and when it contains two elements, one of the two must be zero.*

Proof. See Appendix B.3. □

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

Limit Case. In general, the FOCs—i.e. equation (19), (21) or (22)—are necessary but not sufficient for the optimum. However, when worker heterogeneity is sufficiently small, the difference between the two strategies (quality or quantity) becomes negligible. That is, the firm's problem is concave, the FOCs are sufficient, and the solution is both unique and continuous. The following proposition formalizes this idea.

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

Proof. See Appendix A.4. □

Inspection of the proof reveals that Proposition 2 does not rely on the particular functional form of $\phi(\mu, \lambda)$; it only requires Property A0 and a weaker version of A1, namely $m(\lambda)$ is strictly increasing and concave in λ. This proposition, although simple and intuitive, will prove useful in the following section for constructing (counter)examples of equilibrium allocations that exhibit sorting.

4 Sorting

In this section, we analyze under what conditions the market equilibrium exhibits sorting. We first provide our definitions of positive and negative sorting. Subsequently, we show that the necessary and sufficient condition for sorting relates the elasticity of complementarity of the production function to a *quality-quantity elasticity*.

4.1 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 each worker type x to their employer type y.[27] This definition is not suitable in our environment, because firms do not necessarily hire a unique worker type. Instead, we require a set-based notion of sorting. Following Shimer and Smith (2000) and Shimer (2005), we therefore define sorting as first-order stochastic dominance (FOSD) in firms' distributions of hires.[28] In our environment, this definition can be expressed in terms of the probability that a firm hires a high-type worker, conditional on hiring someone,

$$h(\zeta(y), \lambda(y)) \equiv \frac{\phi(\zeta(y)\lambda(y), \lambda(y))}{m(\lambda(y))}. \tag{23}$$

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

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

[28]Strictly speaking, Shimer and Smith (2000) use a *weaker* notion of sorting which is based on the bounds of the support of the distribution of hires; however, their definition is equivalent to FOSD of this distribution in the random-search environment that they consider. In contrast, Shimer (2005) proves a *stronger* sorting result (high-type workers are more likely to be employed in high-type jobs than in low-type jobs) for a special case (multiplicatively separable production function and urn-ball meetings); however, he acknowledges that the data demands to test this result "may be unrealistic" and suggests FOSD of the distribution of hires as a "more easily testable" alternative.

Definition 3. *An equilibrium exhibits PAM (resp. NAM) if and only if* $h(\zeta(y), \lambda(y))$ *is weakly increasing (resp. decreasing) in* y*.*

We will occasionally distinguish between three types of sorting: i) local sorting, where the definition is satisfied in (the neighborhood of) a particular point y for a given endowment of agents; ii) global sorting, where it is satisfied along the equilibrium path for a given endowment; and iii) robust sorting, where it is satisfied along the equilibrium path for any endowment of agents.

While the literature has traditionally restricted attention to sorting patterns in matches, our environment yields additional predictions. After all, given that firms may interview multiple applicants and subsequently select the most desirable one, there is a meaningful distinction between an application on the one hand and a match on the other hand. Hence, in addition to assortativeness of matches, we can also analyze the assortativeness of applications, i.e. whether the fraction of high-type applicants $\zeta(y)$ increases or decreases in y.

Definition 4. *An equilibrium exhibits PAC (resp. NAC) if and only if* $\zeta(y)$ *is weakly increasing (resp. decreasing) in* y*.*

4.2 Quantity versus Quality

The problem of a firm has two margins: the queue length λ, determining the total number of matches $m(\lambda)$ (quantity), and the fraction of high-type applicants ζ, determining the number of high-type matches $\phi(\lambda\zeta, \lambda)$ (quality). To describe the relation between quality and quantity, we define two elasticities that depend on the meeting technology only and enable us to write our sorting conditions in a uniform way.

Contact Quality-Quantity Elasticity. The marginal effect of an additional applicant (varying λ) on the total number of matches is given by $m'(\lambda)$, while the marginal effect of replacing a low-type applicant with a high-type applicant (varying ζ) on the number of high-type matches is $\phi_\mu(\lambda\zeta, \lambda)$. Keeping ζ constant, we then consider the relative percentage changes of $\phi_\mu(\lambda\zeta, \lambda)$ and $m'(\lambda)$, i.e. the elasticity of $\phi_\mu(\zeta\lambda, \lambda)$ with respect to $m'(\lambda)$, which is given by

$$a^c(\zeta, \lambda) \equiv \frac{\partial \log \phi_\mu(\zeta\lambda, \lambda)}{\partial \log m'(\lambda)} = \frac{\zeta\lambda\phi_{\mu\mu}(\zeta\lambda, \lambda) + \lambda\phi_{\mu\lambda}(\zeta\lambda, \lambda)}{\phi_\mu(\zeta\lambda, \lambda)} \frac{m'(\lambda)}{\lambda m''(\lambda)} > 0, \tag{24}$$

with extrema $\overline{a}^c \equiv \sup_{\zeta,\lambda} a^c(\zeta,\lambda)$ and $\underline{a}^c \equiv \inf_{\zeta,\lambda} a^c(\zeta,\lambda)$.[29] In other words, $a^c(\zeta,\lambda)$ measures the relative percentage changes in marginal changes in match quality and matching probability due to a longer queue while holding the queue composition fixed. It is strictly positive because $m(\lambda)$ is strictly concave and $\phi_\mu(\lambda\zeta,\lambda)$ is strictly decreasing in λ. We call $a^c(\zeta,\lambda)$ the *contact* quality-quantity elasticity because it holds constant the fraction of high-type applicants, i.e. the high-type workers who contacted the firm.

Match Quality-Quantity Elasticity. In equation (24), we kept ζ fixed. If, instead, we keep $h(\zeta,\lambda)$ fixed while varying both ζ and λ, then the firm must choose its queue composition and queue length according to $d\zeta = -\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}d\lambda$, which induces a rate of change of ϕ_μ equal to

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

while the rate of change of $m'(\lambda)$ is $\Delta \log m'(\lambda) = \frac{d\log m'(\lambda)}{d\lambda}d\lambda$. Thus, by equation (24), the elasticity that measures the relative percentage changes of $\phi_\mu(\lambda\zeta,\lambda)$ and $m'(\lambda)$, i.e. $\Delta \log \phi_\mu/\Delta \log m'(\lambda)$, while fixing $h(\zeta,\lambda)$ equals

$$a^m(\zeta,\lambda) \equiv \left.\frac{\partial \log \phi_\mu(\zeta\lambda,\lambda)}{\partial \log m'(\lambda)}\right|_{h(\zeta,\lambda)=\overline{h}} = a^c(\zeta,\lambda)\left(1 - \frac{\partial \phi_\mu/\partial\zeta}{\partial \phi_\mu/\partial\lambda}\frac{\partial h/\partial\lambda}{\partial h/\partial\zeta}\right) > 0, \tag{25}$$

with extrema $\overline{a}^m \equiv \sup_{\zeta,\lambda} a^m(\zeta,\lambda)$ and $\underline{a}^m \equiv \inf_{\zeta,\lambda} a^m(\zeta,\lambda)$.[30] We refer to $a^m(\zeta,\lambda)$ as the *match* quality-quantity elasticity, because it holds constant the probability that the firm matches with a high-type worker.

Homogeneous Queues. It is worth highlighting that $a^m(\zeta,\lambda)$ reduces to $a^c(\zeta,\lambda)$ when the queue is homogeneous, i.e. $a^m(\zeta,\lambda) = a^c(\zeta,\lambda)$ when $\zeta = 0$ or $\zeta = 1$.[31] As we will see below, the infimum (resp. supremum) of a^c and a^m can be reached or approached with $\zeta = 0$ (resp. $\zeta = 1$) for our benchmark and various other meeting technologies. It is therefore not surprising that $\overline{a}^c = \overline{a}^m$ and $\underline{a}^c = \underline{a}^m$ in those cases.

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

[30]Later, in Lemma 8, we prove that the term in the parenthesis in (25) is always between 0 and 1.

[31]To see this, note that $\phi(0,\lambda) = 0$ and $\phi(\lambda,\lambda) = m(\lambda)$ for any λ. Both imply $\partial h/\partial\lambda = 0$.

4.3 Conditions for Sorting

We now analyze the conditions for the market equilibrium to exhibit PAC/PAM; the analysis for NAC/NAM is similar with reversal of the relevant inequalities and therefore omitted. To simplify exposition, we focus on the case in which the queue composition $\zeta(y)$ is unique for all y, e.g. because worker types are close ($x_1 \to x_2$). This case provides our main sorting condition. To prove that this condition is sufficient for any degree of worker heterogeneity, we need to solve the more complicated case where the equilibrium contains multiplicity points; we do this in Appendix B.5.

Condition for Local Sorting. First, assume that $\zeta(y)$ is interior. This allows us to differentiate the FOCs (17) and (19) with respect to y, giving us two equations that jointly determine $\lambda'(y)$ and $\zeta'(y)$, i.e. how the optimal queue length and composition vary with y. In this case, local PAC/PAM is defined as follows.

Definition 5. *If $\zeta(y)$ is unique and interior, then PAC holds locally at point y if $\zeta'(y) \geq 0$, while PAM holds locally if $\frac{d}{dy}h(\zeta(y), \lambda(y)) \geq 0$.*

Recall that (19) is the FOC with respect to ζ. Differentiating (19) with respect to y along the equilibrium path yields

$$-\frac{1}{\phi_\mu}\left(\frac{\partial \phi_\mu}{\partial \zeta}\zeta'(y) + \frac{\partial \phi_\mu}{\partial \lambda}\lambda'(y)\right) = \frac{\Delta f_y}{\Delta f}, \tag{26}$$

which states that the decrease in ϕ_μ must equal the increase in Δf. Similarly, (17) is the FOC with respect to queue length λ. As we show in the proof of Lemma 7, differentiating (17) with respect to y along the equilibrium path and combining the result with (26) yields the percentage change of $m'(\lambda)$ across firm types, i.e.

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

When there is no worker heterogeneity or the meeting technology exhibits no congestion externalities (i.e. it is is invariant), the second factor on the right-hand side reduces to 1.[32] That is, when we move towards more productive jobs, the percentage

[32]In the former case, the firms' optimization problem is simply $\max_\lambda m(\lambda)f(x_1, y) - \lambda U_1$, and in the latter case, it is $\max_{\mu,\lambda} m(\lambda)f(x_1, y) - \lambda U_1 + m(\mu)\Delta f - \mu \Delta U$, where we used the fact that

decrease in $m'(\lambda)$ (as a result of a longer queue) is independent of ζ and simply equals the percentage increase in $f(x_1, y)$.

When there are congestion externalities between heterogeneous workers, however, the optimal queue involves a trade-off between quantity and quality, and more of one affects the marginal contribution of the other. The second factor on the right-hand side of (27) represents this complex interplay between quality and quantity. Under the sufficient condition for robust NAC/NAM (see Proposition 4 below), the second factor exceeds 1, since high-type firms respond to congestion by hiring less high-type workers and more low-type workers, implying that the percentage decrease in $m'(\lambda)$ must exceed $f(x_1, y)$. When the sufficient condition for robust PAC/PAM is satisfied and worker types are relatively close, the second factor is less than 1. That is, the optimal choice for high-type firms is to attract short queues with a high fraction of high-type workers.[33] Dividing both sides of (26) by the corresponding side of (27) gives

$$\frac{\frac{1}{\phi_\mu}\left(\frac{\partial \phi_\mu}{\partial \zeta}\zeta'(y) + \frac{\partial \phi_\mu}{\partial \lambda}\lambda'(y)\right)}{\frac{m''}{m'}\lambda'(y)} = \frac{f^1 \Delta f_y}{f_y^1 \Delta f} \frac{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}}{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f_y^1}}. \tag{28}$$

The left-hand side reflects the relative change in ϕ_μ and $m'(\lambda)$ across firm types. The right-hand side reduces to the elasticity of complementarity in production $\rho(x, y)$ when worker heterogeneity disapears $(x_1, x_2 \to x)$.

Recall that $a^c(\zeta, \lambda)$ measures the relative change in ϕ_μ and $m'(\lambda)$, while fixing ζ. Thus if the right-hand side of (28) is larger than $a^c(\zeta, \lambda)$, then it must be the case that $\zeta'(y) \geq 0$, i.e. PAC holds locally. Similarly, if the right-hand side of (28) is larger than $a^m(\zeta, \lambda)$, then it must be the case that PAM holds locally. We can summarize this in the following Lemma.

Lemma 7. *If $\zeta(y)$ is both unique and interior, then PAC (for $i = c$) and PAM (for*

$m(\mu) = \phi(\mu, \mu) = \phi(\mu, \lambda)$. In both cases, the optimal λ is determined by: $\max_\lambda m(\lambda) f(x_1, y) - \lambda U_1$, where match quality plays no role.

[33] See Appendix B.4 for a formal proof of the statements regarding the magnitude of the second term on the right-hand side of equation (27).

$i = m$) hold locally at y if and only if

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq a^i \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}}, \tag{29}$$

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

Proof. See Appendix A.5. □

Necessary Condition for Robust Sorting. In the limit case considered by Proposition 2 where worker heterogeneity disappears, all firms have a unique optimal ζ. Furthermore, equation (29), the condition for PAC/PAM in this case, reduces to something particularly simple: the left-hand side reduces to $\rho(x, y)$, and the right-hand side becomes $a^i(\zeta(y), \lambda(y))$. Our next result provides a simple necessary condition for robust sorting based on this limit case.

Proposition 3. *A necessary condition for robust PAC (resp. PAM) is that for $i = c$ (resp. $i = m$), we have*

$$\underline{\rho} \equiv \inf_{x,y} \rho(x, y) \geq \sup_{\zeta,\lambda} a^i(\zeta, \lambda) \equiv \overline{a}^i. \tag{30}$$

Similarly, a necessary condition for robust NAC (resp. NAM) is that for $i = c$ (resp. $i = m$), we have

$$\overline{\rho} \equiv \sup_{x,y} \rho(x, y) \leq \inf_{0 \leq \mu \leq \lambda} a^i(\zeta, \lambda) \equiv \underline{a}^i. \tag{31}$$

Proof. See Appendix A.6. □

Proposition 3 relies on the exact same mild assumptions on the meeting technology as Proposition 2 (i.e. Property A0 and a weaker version of A1).

Intuition. Since the market equilibrium is constrained efficient, we describe the intuition for Proposition 3 from the planner's perspective. As discussed in section 3.1, the planner chooses a queue length $\lambda(y)$ and composition $\zeta(y)$ for every firm type to maximize total surplus, subject to the frictions.

When x_1 and x_2 are sufficiently close, a firm's matching probability is of first-order importance: firms with higher productivity are assigned a longer queue length, such that the marginal contribution of a low-type applicant $m'(\lambda)f(x_1, y)$ is (approximately) constant across firm types. This implies that $-d\log m'(\lambda) = d\log f(x_1, y)$, i.e. the increase in queue length is such that the percentage decrease in $m'(\lambda)$ equals the percentage increase in $f(x_1, y)$.

By equation (19), the marginal contribution of a high type replacing a low type in the queue, $\phi_\mu \Delta f$, must also be constant across firm types. This implies that $-d\log\phi_\mu = d\log\Delta f$, i.e. the percentage decrease in ϕ_μ equals the percentage increase in Δf. Therefore, $d\log\phi_\mu / d\log m'(\lambda) = d\log\Delta f / d\log f^1$ across firm types. Making the dependence on y explicit and taking the limit $x_1, x_2 \to x$, we obtain

$$\frac{\frac{d}{dy}\log\phi_\mu(\lambda(y)\zeta(y), \lambda(y))}{\frac{d}{dy}\log m'(\lambda(y))} = \rho(x, y). \tag{32}$$

The left-hand side of this equation measures the relative change of ϕ_μ and m' across firm types; it differs from the quality-quantity elasticity $a^c(\zeta(y), \lambda(y))$, because it does not fix ζ. In other words, a^c describes how large the left-hand side of (32) would be if $\zeta(y)$ were independent of y. A large value for a^c means that the longer queue at firms with higher productivity results in a relatively large drop in the probability ϕ_μ that a high-type worker creates surplus; naturally, this constitutes a force against PAC. To nevertheless obtain PAC locally, this force must therefore be dominated by the complementarities in production, i.e. $a^c(\zeta(y), \lambda(y)) \leq \rho(x, y)$.

This local condition depends not only on x and y, but also on the population measures of workers and firms, through $\lambda(y)$ and $\zeta(y)$. For example, the force against PAC measured by a^c is largest when the relative supply of high-type workers is large such that $\zeta(y) \to 1$. To obtain robust PAC, we therefore need that the supremum of a^c does not exceed the infimum of ρ, i.e. the condition in (30). NAC, PAM and NAM follow a similar logic.[34]

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

[34]This intuition suggests that the necessary conditions (30) and (31) continue to hold with an arbitrary number of worker types. The difficulty in analyzing this case rather lies in proving sufficiency, as dealing with the multiplicity points where the optimal ζ is not unique is manageable with two types but quickly becomes intractable when the number of types increases.

4.4 How Screening Affects Sorting

Our sorting analysis so far has been quite general and has not made use of the specific functional form of $\phi(\mu, \lambda)$ given by (7), except that it needs to satisfy regularity conditions A0 and A1. The following lemma establishes that this functional form yields very simple expressions for the right-hand side of equations (30) and (31).

Lemma 8. *If ϕ is given by (7), then $a^i(\zeta, \lambda)$ is strictly increasing in ζ for $i = c$ or m, and*

$$\overline{a}^c = \overline{a}^m = \frac{1+\sigma}{2} \quad \text{and} \quad \underline{a}^c = \underline{a}^m = \frac{1-\sigma}{2}. \tag{33}$$

Proof. See Appendix A.7. □

Together with Proposition 3, this lemma implies that $\underline{\rho} \geq (1+\sigma)/2$ is necessary for robust PAC/PAM, while $\overline{\rho} \leq (1-\sigma)/2$ is necessary for robust NAC/NAM. The following proposition shows that these two conditions are also sufficient.

Proposition 4. *Assume that ϕ is given by (7) with $\sigma > 0$. The equilibrium then exhibits robust PAC/PAM (resp. NAC/NAM) if and only if $\underline{\rho} \geq (1+\sigma)/2$ (resp. $\overline{\rho} \leq (1-\sigma)/2$).*

Proof. See Appendix A.8. □

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

Corollary 1. *Assume that ϕ is given by (7) with $\sigma > 0$. The equilibrium then exhibits robust PAC/PAM (resp. NAC/NAM) if and only if $f(x, y)$ is $2/(1-\sigma)$-root-supermodular (resp. $2/(1+\sigma)$-root-submodular) on $\mathcal{X} \times \mathcal{Y}$.*

Two special cases are worth highlighting. 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, a stronger degree of substitutability is required for NAC/NAM as the expected number of interviews goes up: the production function should be square-root-submodular if $\sigma = 0$ and submodular when $\sigma = 1$. Figure 1 illustrates these results.

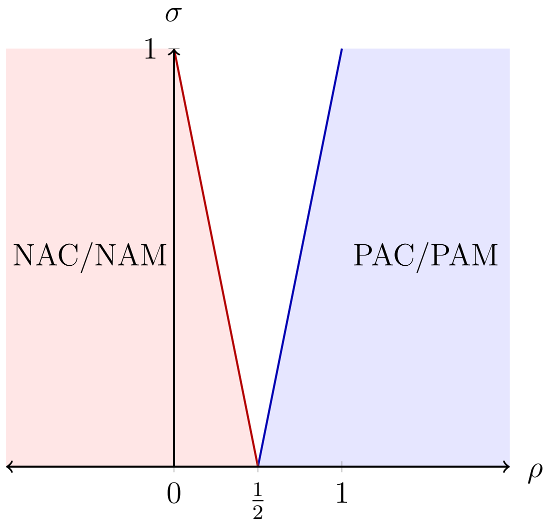

Figure 1: Combinations of ρ and σ that give rise to robust PAC/PAM (blue) or robust NAC/NAM (red), assuming a CES production function.

Intuition. To understand why an increase in σ is a force against positive sorting, consider again the case where $x_1 \to x_2$, such that firms with higher productivity have a longer queue, i.e. $\lambda'(y) > 0$. As discussed above, for given ζ, this longer queue unambiguously decreases the probability $\phi_\mu(\zeta\lambda(y), \lambda(y))$ that a high-type applicant will create surplus, which is a force against sorting. The magnitude of this force, depends on 1) whether primarily low types or high types are being added (as measured by ζ), and 2) whether types can easily be distinguished (as measured by σ). When the queue primarily consists of low types, i.e. ζ is low, the force against positive sorting is small and it is mitigated by increasing σ, because that increases the chance that the firm will interview at least one high type. In contrast, when the queue primarily consists of high types, i.e. ζ is high, the force against positive sorting is large and it is exacerbated by increasing σ, because that increases the chance that the firm will interview multiple high types (making the marginal contribution to surplus by any one of the high types zero, as the firm will still be matched with a high type in their absence). This latter case forms the strongest restriction given our objective to derive a condition for robust sorting.

5 Extensions

In this section, we explore various extensions of our environment.

5.1 Alternative Meeting Technologies

We have derived our main result, Proposition 4, for a specific micro-foundation of the meeting technology, such that $\phi(\mu, \lambda)$ was given by (7). The search literature offers various alternatives.[35] Since most of our analysis is presented in a general way, such alternatives can be analyzed by updating $\phi(\mu, \lambda)$. In particular, the necessary conditions in Proposition 3 continue to apply under very mild restrictions on $\phi(\mu, \lambda)$, as discussed there. Proving that these conditions are also sufficient is more involved because one needs to verify that $\phi(\mu, \lambda)$ satisfies the regularity conditions specified in the proofs of Lemma 6 and Proposition 4. We briefly discuss the two most common classes of meeting technologies in the literature: bilateral and invariant technologies.[36]

Bilateral Technologies. Bilateral technologies can be interpreted as firms being able to interview only a single applicant.[37] In that case, $\phi(\mu, \lambda) = m(\lambda)\mu/\lambda$, where, as before, $m(\lambda)$ is the probability that a firm receives at least one applicant. The following lemma establishes that $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$ then become independent of ζ and both reduce to the elasticity of substitution of the total number of matches in a submarket, which is precisely the object that Eeckhout and Kircher (2010a) show to be important in their study of sorting patterns for bilateral technologies.

Lemma 9. *When the meeting technology is bilateral, we have*

$$a^c(\zeta, \lambda) = a^m(\zeta, \lambda) = \frac{m'(\lambda)(\lambda m'(\lambda) - m(\lambda))}{\lambda m(\lambda) m''(\lambda)}. \tag{34}$$

Proof. See Appendix B.6. □

With bilateral technologies, firms always find it optimal to attract either only low-type or only high-type workers. Hence, if $\zeta(y)$ is not unique at some point y, then we must have $Z(y) = \{0, 1\}$. In this case, PAC/PAM simply requires that $Z(y') = \{0\}$ for all $y' < y$ and $Z(y') = \{1\}$ for all $y' > y$, while the reverse must hold for NAC/NAM. Appendix B.7 demonstrates for our case with two worker types that we recover the sorting results of Eeckhout and Kircher (2010a), who assume a continuum of worker types.

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

[36] For the meeting technology of Wolthoff (2018), in which workers send their applications according to an urn-ball process while screening remains geometric, we can show sufficieny numerically.

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

Invariant Technologies. Invariant technologies, such as urn-ball or geometric, exhibit perfect screening in the sense that the presence of low types does not make it harder (or easier) for a firm to identify a high-type applicant. That is, $\phi_\lambda(\mu, \lambda) = 0$ for all μ and λ, or equivalently, $\phi(\mu, \lambda) = \phi(\mu, \mu) \equiv m(\mu)$, where $m(\mu)$ is always assumed to be strictly concave (see Cai et al., 2017). Furthermore, one can show that $\lim_{\mu\to 0} \mu m''(\mu) = 0$.[38] To analyze this case, we first introduce two elasticities:

$$\varepsilon_0(\mu) = \frac{\mu m'(\mu)}{m(\mu)} \quad \text{and} \quad \varepsilon_1(\mu) = \frac{\mu m''(\mu)}{m'(\mu)}. \tag{35}$$

The following lemma then presents $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$ for invariant technologies in terms of $\varepsilon_0(\cdot)$ and $\varepsilon_1(\cdot)$.

Lemma 10. *When the meeting technology is invariant, we have*

$$a^c(\zeta, \lambda) = \frac{\varepsilon_1(\lambda\zeta)}{\varepsilon_1(\lambda)} \quad \textit{and} \quad a^m(\zeta, \lambda) = \frac{\varepsilon_1(\lambda\zeta)}{\varepsilon_1(\lambda)} \frac{\varepsilon_0(\lambda)}{\varepsilon_0(\lambda\zeta)}, \tag{36}$$

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

Proof. See Appendix B.8. □

Recall that the contact quality-quantity elasticity $a^c(\zeta, \lambda)$ measures the relative percentage changes of $\phi_\mu(\lambda\zeta, \lambda)$ and $m'(\lambda)$ while holding ζ constant. For invariant technologies, $\phi_\mu(\lambda\zeta, \lambda) = m'(\lambda\zeta)$; thus, $a^c(\zeta, \lambda)$ is simply $\varepsilon_1(\lambda\zeta)/\varepsilon_1(\lambda)$. Next, note that $a^m(\zeta, \lambda)$ measures the same relative percentage changes while holding $m(\lambda\zeta)/m(\lambda)$ constant. The latter requires the percentage changes of $m(\lambda\zeta)$ and $m(\lambda)$ to be equal, that is, the percentage change of $\lambda\zeta$ equals $\varepsilon_0(\lambda)/\varepsilon_0(\lambda\zeta)$ times the percentage change of λ; thus, $a^m(\zeta, \lambda) = a^m(\zeta, \lambda)\varepsilon_0(\lambda)/\varepsilon_0(\lambda\zeta)$.

When the meeting technology is invariant, the surplus function $S(\mu, \lambda, y)$ defined in (8) simplifies to $m(\lambda)f(x_1, y) + m(\mu)[f(x_2, y) - f(x_1, y)]$. The firms' problem becomes strictly concave in (μ, λ), which implies that for each y, there exists exactly one optimal queue $(\mu(y), \lambda(y))$. That is, the complication that $Z(y)$ may not be unique never occurs, which greatly simplifies the proof that the necessary conditions from Proposition 3 are also sufficient.

[38] For common invariant technologies such as urn-ball and geometric, this is trivially satisfied because $m''(0)$ is finite. A general proof—based on the observation that $m(\mu)$ is a Bernstein function for invariant technologies, so it admits the Lévy–Khintchine representation (see Theorem 3.2 in Schilling et al., 2012)—is available upon request.

Proposition 5. *Suppose the meeting technology is invariant. The equilibrium then exhibits robust PAC (resp. PAM) if and only if $\underline{\rho} \geq \overline{a}^c$ (resp. $\underline{\rho} \geq \overline{a}^m$). In contrast, The equilibrium exhibits robust NAC/NAM if and only if $f(x, y)$ is submodular.*

Proof. See Appendix B.9. □

With two mild assumptions, which are satisfied by both the urn-ball and the geometric technology, the sorting results become particularly simple.[39]

Assumption INV-1. *$\varepsilon_1(\mu)$ is decreasing in μ.*

Assumption INV-2. *$\varepsilon_1(\mu)/\varepsilon_0(\mu)$ is decreasing in μ.*

Assumption INV-1 (resp. INV-2) implies that $a^c(\zeta, \lambda)$ (resp. $a^m(\zeta, \lambda)$) is increasing in ζ so that $\overline{a}^c = 1$ (resp. $\overline{a}^m = 1$). This yields the following result.

Corollary 2. *When the meeting technology is invariant and satisfies Assumption INV-1 (resp. INV-2), the equilibrium exhibits robust PAC (resp. PAM) if and only if $f(x, y)$ is log-supermodular.*

Because invariance implies that the optimal queue composition is always unique, these sorting results can be generalized to an arbitrary number of worker types in a straightforward manner. We demonstrate this in Appendix B.10.

5.2 Signals

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

Environment with Signals. Consider an environment like our benchmark model, except that firms can costlessly observe a signal for every applicant. For high-type applicants, the signal is positive with certainty. In contrast, a low-type applicant generates a negative signal with probability $\tau \in [0, 1]$ and a positive signal with complementary probability. Hence, the signal is perfect if $\tau = 1$, but pure noise if $\tau = 0$. Using this information, firms will first interview applicants with positive signals

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

and only interview applicants with negative signals if interview capacity remains. As before, an interview reveals the applicant's true type.

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

Proposition 6. *In our environment with signals, consider a firm with queues* (μ, λ)*. Let* $\widehat{\sigma} = 1 - (1-\tau)(1-\sigma) \in [0,1]$*, then the probability that the firm interviews at least one high-type worker equals*

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

Proof. See Appendix B.11. □

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 to account for the fact that the signal precision τ is a substitute for the screening intensity σ.

5.3 Endogenous Screening

The screening intensity σ was exogenous in our baseline model. However, firms can generally influence the number of applicants that they interview. In this section, we therefore endogenize σ and discuss how it affects our results. We keep the discussion concise and refer to the proof of Proposition 7 for details.

Environment with Endogenous Screening. Consider an environment which is like our benchmark model, except that firms additionally choose (and post) their recruiting intensity $\sigma \in [0,1]$ at a linear cost $c\sigma$, where $c \geq 0$.[40] That is, they solve

$$\max_{\sigma, \mu, \lambda} \quad \frac{\lambda}{1+\lambda} f^1 + \frac{\mu}{1 + \sigma\mu + (1-\sigma)\lambda} \Delta f - \lambda U_1 - \mu \Delta U - c\sigma. \tag{37}$$

Since the second term above is convex in σ and $c\sigma$ is linear, the above profit function is *convex* in σ. The maximum is therefore reached at a corner, i.e. when $\sigma = 0$ or 1.

[40] Posting contracts that include σ in addition to wages is necessary for constrained efficiency in this environment. More restrictive contract spaces and more general cost functions are left for future research. Wolthoff (2018) endogenizes σ in a similar way as us, but with a cost function that is sufficiently convex (in an otherwise quite different model). In the random search model of Birinci et al. (2020), firms have the option to learn all their applicants' types after paying a fixed cost.

To determine firms' choice, we compare the profits from the two options.

Profits with No Screening. Consider a firm of type y choosing $\sigma = 0$. This firm's optimal queue then consists of either low-type workers or high-type workers, but not both. Suppose the firm attracts workers of type x_i. Equation (37) then reduces to $\max_{\lambda_i} m(\lambda_i)f(x_i, y) - \lambda_i U_i$. Because $m(\lambda)$ is strictly concave, the FOC of this problem is both necessary and sufficient. Assuming that $f(x_i, y) > U_i$, the optimal queue length is $\lambda_i = \sqrt{f(x_i, y)/U_i} - 1$, which yields an expected payoff of

$$\pi_i(y) = \left(\sqrt{f(x_i, y)} - \sqrt{U_i}\right)^2. \tag{38}$$

Naturally, the firm chooses the type of workers it wishes to attract based on whether $\pi_1(y)$ or $\pi_2(y)$ is higher, which requires comparing $\sqrt{f(x_2, y)} - \sqrt{f(x_1, y)}$ with $\sqrt{U_2} - \sqrt{U_1}$. If the former is strictly increasing in y, i.e. f is strictly square-root supermodular, then there exists a unique y^{EK} such that $\pi_2(y) > \pi_1(y)$ if $y > y^{EK}$ and vice versa. This result is a special case of Section 5.1 or Eeckhout and Kircher (2010a).

Profits with Perfect Screening. When the firm chooses $\sigma = 1$, (37) reduces to

$$\overline{\pi}(y) \equiv \max_{0 \leq \mu \leq \lambda} \frac{\lambda}{1+\lambda} f^1 + \frac{\mu}{1+\mu} \Delta f - \lambda U_1 - \mu \Delta U. \tag{39}$$

This problem is strictly concave in (μ, λ), so that the FOCs are both necessary and sufficient. The only complexity lies in the constraint $0 \leq \mu \leq \lambda$, which implies that there are four possibilities with respect to the optimal applicant pool: a firm choosing $\sigma = 1$ may attract (i) no applicants, such that $\overline{\pi}(y) = 0$; (ii) only low-type applicants, such that $\overline{\pi}(y) = \pi_1(y)$; (iii) only high-type applicants, such that $\overline{\pi}(y) = \pi_2(y)$; or (iv) both types of applicants, in which case the FOCs imply $\mu = \sqrt{\Delta f / \Delta U} - 1$ and $\lambda = \sqrt{f^1 / U_1} - 1$, such that

$$\overline{\pi}(y) = \left(\sqrt{f^1} - \sqrt{U_1}\right)^2 + \left(\sqrt{\Delta f} - \sqrt{\Delta U}\right)^2. \tag{40}$$

Choice of Screening Intensity. The characterization of $\pi_1(y)$, $\pi_2(y)$, and $\overline{\pi}(y)$ completes the analysis of the firm's problem (37): the firm chooses $\sigma = 1$ if $\overline{\pi}(y) - c \geq \max\{\pi_1(y), \pi_2(y)\}$ and $\sigma = 0$ otherwise.

An Auxiliary Function. To simplify exposition, we introduce a transformation $\Omega(\cdot)$ of $\kappa(y)$, the output dispersion parameter defined by equation (15), and establish that this transformation is strictly decreasing:

$$\Omega(\kappa) \equiv \frac{1}{2} + \frac{\ln(\sqrt{\kappa} + \sqrt{1+\kappa})}{\ln(1+\kappa)}. \tag{41}$$

Lemma 11. *$\Omega(\kappa)$ is strictly decreasing with $\lim_{\kappa \to 0} \Omega(\kappa) = \infty$ and $\lim_{\kappa \to \infty} \Omega(\kappa) = 1$.*

Proof. See Appendix B.12. □

Sorting. In our main analysis, we showed that—relative to a bilateral world—allowing firms to interview multiple applicants makes robust sorting harder; in particular, we found that log-supermodularity is necessary and sufficient to obtain robust PAC/PAM for any screening intensity σ, while submodularity is the corresponding condition for NAC/NAM.

We now analyze whether these conditions carry over to an environment with endogenous screening in the sense that they are also necessary and sufficient to obtain robust sorting for any screening cost c. Necessity is immediate: when $c = 0$, all firms choose $\sigma = 1$ and the results of Proposition 5 and Corollary 2 apply. For NAC/NAM, sufficiency is also relatively straightforward: we find that strict submodularity is sufficient for robust NAC/NAM for any screening cost c. PAC/PAM is more complicated. The following proposition summarizes the results.

Proposition 7. *In our environment with endogenous screening, the following holds:*

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

(ii) Given any log-supermodular function f, we can find an endowment of agents and a screening cost c such that PAC/PAM fails in equilibrium. However, given an endowment of agents and c, PAC/PAM holds in equilibrium as long as

$$\underline{\rho} \geq \Omega(\kappa(\underline{y})). \tag{42}$$

Proof. See Appendix B.13. □

The sufficient condition (42) depends only on $\underline{\rho}$, the lower bound of the production complementarities, and $\kappa(\underline{y})$, the lower bound of the output dispersion.[41] Hence,

[41] Since f is log-supermodular, $\kappa(y)$ is smallest at $y = \underline{y}$. Also, by Lemma 11, (42) requires $\underline{\rho} > 1$.

PAC/PAM requires that either production complementarity or output dispersion is sufficiently large. Note that condition (42) is quite sharp: in the proof of Proposition 7, we show that with CES production we can construct counterexamples where PAC/PAM fails in equilibrium whenever $\rho < \Omega(\kappa(\underline{y}))$.

At first, our result regarding PAC/PAM may seem puzzling. One may have expected that, with strong complementarities, firms' incentives to invest in (ex post) screening are increasing in their productivity and that since the least-productive firms can only afford to attract low-type workers, PAC/PAM arises. This intuition turns out to be wrong. When x_1 and x_2 are sufficiently close, the most-productive firms find it optimal to attract high-type applicants only. They are not willing to provide low-type workers with their market utility, because compensating them for the low matching probability that results from the presence of many high-type workers requires a very high wage. Therefore, firms in the middle of the productivity distribution have the strongest incentives to screen ex post. Although those firms also prefer to hire high-type workers, they can only afford to offer modest wages to them and therefore they attract relatively few of them. As a consequence, they can attract low-type workers for a relatively low wage (since they offer them a high hiring probability). However, some firm types below those screening firms are not productive enough to be willing to pay the screening cost as an insurance device (since the opportunity costs of remaining unmatched is lower for those firms), but conditional on not screening ex-post, they are productive enough to target high-type applicants. When this happens, PAC/PAM fails in the middle. In the limit where $x_1 \to x_2$, we can always find an endowment of agents $(x_1, x_2, \ell_1, \ell_2, J(y))$ and a screening cost c such that PAC/PAM fails in equilibrium, even for log-supermodular production functions.

This scenario does not arise when either the degree of complementarity $\underline{\rho}$ or output dispersion $\kappa(\underline{y})$ is large, i.e. (42) holds. In that case, the incentive to attract low-type workers as insurance against failing to hire is decreasing in firms' type. Then, the most-productive firms attract only high-type workers, firms in the middle attract both types and screen ex post, and the least-productive firms attract only low-type workers. In other words, the gains from ex-post screening are first increasing in y, reach their maximum at y^{EK} and from then onwards are decreasing in y.

6 Conclusion

A firm with a vacancy typically has multiple instruments to screen applicants. By announcing the terms of trade ex ante, it can discourage certain types of workers from applying, while ex post—after receiving applications—it can interview applicants in an attempt to identify the most profitable hire. In this paper, we show how these instruments jointly determine equilibrium outcomes, including sorting patterns. Perhaps surprisingly, we find that if ex post screening is easier (firms can screen more applicants), this makes sorting harder. That is, stronger complementarities in the production technology are necessary to get positive assortative matching. The more workers a firm can screen, the stronger the incentives for high-type workers are to avoid ending up in the same pool of applicants and this is a force against sorting which is by itself efficient (a social planner also wants to reduce the probability that valuable resources are wasted because they end up in the same pool).

There are several promising avenues for future research. On the theoretical side, in markets with a long hiring cycle, like the academic job market, workers may have relatively strong incentives to send multiple applications simultaneously. This would reduce the cost for high-type workers to end up in the same queue as other high-type workers. However, even in those markets, high-type workers have incentives to diversify and not only apply to the top places.

On the empirical side, an important implication of our model is that sorting patterns are driven both by the production function and the meeting process. In order to identify complementarities in production, we may need—besides data on matches—additional information on the entire pool of applicants. This way, we can first identify the parameters of the meeting technology (i.e. how many workers apply, which workers apply and how many are screened) and then, conditional on the meeting technology, matching patterns are informative on production complementarities.

References

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

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

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

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

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

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

Banfi, S., Choi, S., and Villena-Roldán, B. (2020). Sorting on-line and on-time. mimeo.

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.

Birinci, S., See, K., and Wee, S. L. (2020). Job applications and labor market flows. mimeo.

Borovičková, K. and Shimer, R. (2020). High wage workers work for high wage firms. mimeo.

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. (2020). Meetings and mechanisms. mimeo.

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

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

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

Eeckhout, J. and Kircher, P. (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.

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.

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

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

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

Schilling, R., Song, R., and Vondracek, Z. (2012). *Bernstein Functions: Theory and Applications*. De Gruyter Studies in Mathematics. De Gruyter.

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.

Shi, S. (2002). A directed search model of inequality with heterogeneous skills and skill-biased technology. *Review of Economic Studies*, 69(2):467–491.

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. (2019). Firming up inequality. *Quarterly Journal of Economics*, 134(1):1–50.

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.

Appendix A Proofs

A.1 Proof of Lemma 1

For a given ρ_0, the derivative of $\log f_y(x, y) - \rho_0 \log f(x, y)$ with respect to x equals

$$\frac{\partial}{\partial x}(\log f_y - \rho_0 \log f) = \frac{f_{xy}}{f_y} - \rho_0 \frac{f_x}{f} = \frac{f_{xy} f - \rho_0 f_x f_y}{f f_y},$$

where we suppress the arguments of $f(x, y)$ and its partial derivatives for simplicity. The right-hand side is weakly positive (resp. negative) if $\rho_0 = \underline{\rho}$ (resp. $\rho_0 = \overline{\rho}$), which means that $\log f_y(x_2, y) - \underline{\rho} \log f(x_2, y) \geq \log f_y(x_1, y) - \underline{\rho} \log f(x_1, y)$, and $\log f_y(x_2, y) - \overline{\rho} \log f(x_2, y) \geq \log f_y(x_1, y) - \overline{\rho} \log f(x_1, y)$, which jointly imply (2). □

A.2 Proof of Lemma 4

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

$$\psi_1(\mu^*, \lambda^*) \leq \phi_\mu(\mu^*, \lambda^*) \leq \psi_2(\mu^*, \lambda^*), \tag{43}$$

where ψ_1 and ψ_2 are defined by equation (10). Consequently, the wages must satisfy

$$w_1^* = \frac{U_1}{\psi_1(\mu^*, \lambda^*)} \geq \frac{U_1}{\phi_\mu(\mu^*, \lambda^*)} \quad \text{and} \quad w_2^* = \frac{U_2}{\psi_2(\mu^*, \lambda^*)} \leq \frac{U_2}{\phi_\mu(\mu^*, \lambda^*)}. \tag{44}$$

Moreover, the FOC of (11) with respect to μ implies $\phi_\mu(\mu^*, \lambda^*)(f(x_2, y) - f(x_1, y)) = U_2 - U_1$. Combining this FOC with (44) implies

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

The strict inequality in $f(x_2, y) - w_2^* > f(x_1, y) - w_1^*$ then follows because the two inequalities in (43) cannot hold simultaneously; that would imply that $\phi(\mu, \lambda^*)$ is linear for $\mu \in [0, \lambda^*]$, in which case the firm's problem never has an interior solution (see Section 5.1 for an extensive discussion of this case). □

A.3 Proof of Lemma 5

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

A.4 Proof of Proposition 2

As $x_2 \to x_1$, U_2 and U_1 approach a common value $\underline{U}$ and total surplus (8) converges to $m(\lambda) f(x_1, y)$, which implies that firms' expected profit tends to $m(\lambda) f(x_1, y) - \lambda \underline{U}$. Since $0 \leq m(\lambda) f(x_1, y) - \lambda \underline{U} < f(x_1, y) - \lambda \underline{U}$, this means that—when x_2 is sufficiently close to x_1—firms' choice of queue length λ is bounded from above; in particular $\lambda < \overline{\lambda}(y) \equiv \underline{U}/f(x_1, \underline{y})$, such that we can restrict each firm's choice of queues to be in the convex set $\Delta(y) \equiv \{(\mu, \lambda) \,|\, 0 \leq \mu \leq \lambda \leq \overline{\lambda}(y)\}$.

On this set, the right-hand side of the firm's SOC (16) is bounded due to continuity. Hence, (16) will hold for all (μ, λ) in $\Delta(y)$ when $\kappa(y)$, or equivalently $x_2 - x_1$, is sufficiently small. That is, for each firm type y, the surplus function is concave on the set $\Delta(y)$. Therefore, each firm's solution $(\mu(y), \lambda(y))$ is unique, and by the Theorem of the Maximum, it is also continuous. Furthermore, when $\mu(y)$ and $\lambda(y)$ satisfy $0 < \mu(y) < \lambda(y)$, they are jointly determined by the FOCs (17) and (19). Hence, by the implicit function theorem, they are both continuously differentiable around that point. The same then applies to $\zeta(y) = \mu(y)/\lambda(y)$. □

A.5 Proof of Lemma 7

Differentiating (17) along the equilibrium path yields

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

where we have suppressed the arguments $\mu(y)$ and $\lambda(y)$ from the functions m and ϕ. By (19), we can substitute $\phi_\mu \Delta f$ for $U_2 - U_1$ on the left-hand side. The resulting equation and equation (26) are two linear equations in $\zeta'(y)$ and $\lambda'(y)$. A simple but tedious calculation then yields (27).

Rearranging equation (28) gives

$$-\frac{1}{\phi_\mu}\frac{\partial \phi_\mu}{\partial \zeta}\zeta'(y) = \frac{f_y^1}{f^1}\left(\frac{f^1 \Delta f_y}{f_y^1 \Delta f} - \frac{\frac{1}{\phi_\mu}\frac{\partial \phi_\mu}{\partial \lambda}}{\frac{m''}{m'}} \frac{1 - \frac{1}{m'}\left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda\right)\frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''}\left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda}\right)\frac{\Delta f}{f^1}}\right) \tag{45}$$

where we used equation (27) to substitute out $\lambda'(y)$. Since $\phi(\mu, \lambda)$ is strictly concave, $\frac{\partial \phi_\mu}{\partial \zeta} = \lambda \phi_{\mu\mu} < 0$, which implies that $\zeta'(y) \geq 0$ if and only if the term in the parenthesis on the right-hand side is positive, i.e. (29) holds with $i = c$.

By definition, PAM is equivalent to $\frac{\partial h}{\partial \zeta}\zeta'(y) + \frac{\partial h}{\partial \lambda}\lambda'(y) \geq 0$. Combining (27) and (45) then shows that PAM is obtained if and only if (29) holds with $i = m$. □

A.6 Proof of Proposition 3

First consider PAC. We show that if the necessary condition fails, we can construct a counterexample in which worker heterogeneity is small and PAC fails. The other cases (PAM, NAC and NAM) follow the same logic.

Suppose that (30) does not hold for $i = c$, so that there exist x', y', μ', and λ' such that $\rho(x', y') < a^c(\mu', \lambda')$. By continuity, we can then assume that $0 < \mu' < \lambda'$ (note the strict inequality), and that there exists a small $\epsilon_0 \in (0, (\lambda' - \mu')/2)$ such that the above inequality holds for all $x \in [x' - \epsilon_0, x' + \epsilon_0]$, $y \in [y' - \epsilon_0, y' + \epsilon_0]$, $\mu \in [\mu' - \epsilon_0, \mu' + \epsilon_0]$, and $\lambda \in [\lambda' - \epsilon_0, \lambda' + \epsilon_0]$. Fix ϵ_0 from now on.

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, pick x_2, $\underline{y}$, and $\overline{y}$ such that $x_2 - x' = y' - \underline{y} = \overline{y} - y'$ (i.e. y' is the midpoint of $\underline{y}$ and $\overline{y}$). We denote this difference by ϵ_1 and let $\epsilon_1 \to 0$. Note that when

$\epsilon_1 = 0$, both firm and worker heterogeneity disappear. For sufficiently small ϵ_1, the equilibrium $\lambda(y)$ is unique and continuous by Proposition 2. Since the aggregate ratio of workers and firms is λ', for sufficiently small ϵ_1, we have $\lambda(y) \in [\lambda' - \epsilon_0, \lambda' + \epsilon_0]$ for all y. Furthermore, $\mu(y)$ is continuous and the average $\mu(y)$ is μ', i.e. $\int_{\underline{y}}^{\overline{y}} \mu(y) dJ(y) = \mu'$. Therefore, by continuity there exists some y_0 such that $\mu(y_0) = \mu'$. To sum up, at point y_0 we have $\mu' = \mu(y_0) < \lambda(y_0) \in (\lambda' - \epsilon_0, \lambda' + \epsilon_0)$. Since $\zeta(y_0)$ is both unique and interior, $\zeta(y)$ is differentiable at point y_0 (see Proposition 2). Hence, the assumptions of Lemma 7 are satisfied at point y_0.

When $\epsilon_1 \to 0$, the left-hand side of (29) approaches $\rho(x', y')$. On the right-hand side, $a^c(\mu(y_0), \lambda(y_0)) \to a^c(\mu', \lambda')$ (note that the choice of y_0 depends on the value of ϵ_1). Furthermore, at point y_0, we have

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

where we suppress the arguments of $m(\lambda')$ and $\phi(\mu', \lambda')$. Similarly, the denominator on the right-hand side of (29) also goes to 1. Therefore, for (29) to hold at point y_0, we need $\rho(x', y') \geq a^c(\mu', \lambda')$, which yields a contradiction. $\square$

A.7 Proof of Lemma 8

We first consider $a^c(\zeta, \lambda)$. Since $\phi(\mu, \lambda)$ is given by equation (7) and $a^c(\zeta, \lambda)$ is defined by equation (24), direct calculation yields

$$a^c(\zeta, \lambda) = \frac{1 + \lambda}{2\lambda} \left(1 + \frac{1}{1 + (1 - \sigma)\lambda} - \frac{2}{1 + \sigma\zeta\lambda + (1 - \sigma)\lambda} \right). \tag{46}$$

Note that $a^c(\zeta, \lambda)$ is strictly increasing in ζ. Thus, we have $\max_\zeta a^c(\zeta, \lambda) = a^c(1, \lambda)$ and $\min_\zeta a^c(\zeta, \lambda) = a^c(0, \lambda)$. Moreover, (46) reveals that $a^c(0, \lambda) + a^c(1, \lambda) = 1$ and $\frac{da^c(1,\lambda)}{\lambda} = -\frac{\sigma(1-\sigma)}{2(1+(1-\sigma)\lambda)^2} < 0$. Therefore, $a^c(1, \lambda)$ approaches its supremum when $\lambda \to 0$ and $a^c(0, \lambda)$ approaches its infimum when $\lambda \to 0$. Hence, we have $\sup_{\zeta,\lambda} a^c(\zeta, \lambda) = \lim_{\lambda \to 0} a^c(1, \lambda) = (1 + \sigma)/2$ and $\inf_{\zeta,\lambda} a^c(\zeta, \lambda) = 1 - \sup_{\zeta,\lambda} a^c(\zeta, \lambda) = (1 - \sigma)/2$, where neither the infimum nor the supremum can be reached because we require $\lambda > 0$.

Next, we consider $a^m(\mu, \lambda)$. Analogous to above, direct computation yields

$$a^m(\zeta, \lambda) = \frac{\lambda(1 + \lambda)(1 - \sigma) + 2\sigma\zeta\lambda}{2\lambda\,(1 + (1 - \sigma)\lambda)}. \tag{47}$$

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

Finally, note that

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

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

A.8 Proof of Proposition 4

We focus here on the case where $\zeta(y)$ is unique for all y. The more complicated case where $Z(y)$ may contain multiple elements is analyzed in Appendix B.5. First, we present a useful technical lemma; then, we show that the necessary conditions in Proposition 3 are also sufficient, under some regularity conditions; finally, we show that our benchmark technology satisfies these regularity conditions.

A.8.1 A Technical Lemma

The first two parts of the following lemma are trivial, whereas the third part is non-trivial and critical for our results.

Lemma 12. *(i) If $\rho > 1$, then $\frac{1}{\kappa}((1+\kappa)^\rho - 1)$ is strictly increasing for $\kappa > 0$; (ii) if $\rho \in (0,1)$, then $\frac{1}{\kappa}((1+\kappa)^\rho - 1)$ is strictly decreasing for $\kappa > 0$; and (iii) if $\rho \in (0,1)$, then $\left(\frac{1}{\kappa} + \frac{1-\rho}{2}\right)((1+\kappa)^\rho - 1)$ is strictly increasing for $\kappa > 0$.*

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

For (iii), direct computation gives

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

The numerator on the right-hand side equals zero for $\kappa = 0$. Moreover, its derivative is $\frac{d}{d\kappa}[2(1+\kappa)^{1-\rho} - 2 - \kappa(1-\rho)(2-\kappa\rho)] = 2(1-\rho)[(1+\kappa)^{-\rho} - (1-\kappa\rho)] > 0$, because convexity of $(1+\kappa)^{-\rho}$ implies $(1+\kappa)^{-\rho} - (1-\kappa\rho) > 0$. Hence, the numerator on the right-hand side is strictly positive for $\kappa > 0$, which proves (iii). □

A.8.2 Sufficiency of the Necessary Conditions

Since we assumed $\zeta(y)$ is unique for all y, it is also continuous by the Theorem of the Maximum. If PAC/PAM holds locally at all interior points, then by continuity, it also holds globally along the equilibrium path. The following lemma establishes that—subject to regularity conditions—the necessary conditions (30) and (31) imply that PAC/PAM holds locally at all interior points, so they are also sufficient.

Lemma 13. *Let $i = c$ or $i = m$. If, for any μ and λ, we have*

$$\overline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \geq \max(1 - \overline{a}^i, 0), \tag{48}$$

then (30) implies that (29) always holds in equilibrium (i.e. PAC/PAM).

If (i) ϕ_λ/ϕ_μ is weakly decreasing in μ, (ii) $0 \leq \underline{a}^i < 1$, and (iii) for any μ and λ,

$$\underline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \leq 0, \tag{49}$$

then (31) implies that (29) holds in equilibrium with $\leq$ instead of $\geq$ (i.e. NAC/NAM).

Proof. Recall $\kappa(y) \equiv \Delta f / f^1$. Throughout this proof, we will then use the following inequalities which result from rewriting (2):

$$\frac{(1 + \kappa(y))^{\underline{\rho}} - 1}{\kappa(y)} \leq \frac{f^1 \Delta f_y}{f_y^1 \Delta f} \leq \frac{(1 + \kappa(y))^{\overline{\rho}} - 1}{\kappa(y)}. \tag{50}$$

First, consider PAC/PAM. Assume the necessary condition (30) holds, i.e. $\underline{\rho} \geq \overline{a}^i$. Since $\overline{a}^i \geq 0$, this implies that $\Delta f_y \geq 0$ (i.e. f is supermodular) such that the left-hand side of (29) is positive. We now prove a stronger version of (29), i.e.

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

where $a^i(\zeta, \lambda)$ is replaced by its supremum $\overline{a}^i$. This is justified because if the second factor 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. Firms' SOC implies that the denominator of this factor is positive. Rearranging terms therefore gives

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} + \frac{\Delta f_y}{f_y^1} \left[\overline{a}^i \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \right] \geq \overline{a}^i. \tag{51}$$

Consider now two subcases, determined by the value of $\underline{\rho}$. If $\underline{\rho} \geq 1$, then

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

where inequalities follow from of Lemma 1, $\underline{\rho} \geq 1$ (see part (i) of Lemma 12), and the necessary condition (30), respectively. Since (48) implies that the term in square brackets in (51) is positive, (51) holds, which then implies (29).

In contrast, if $\underline{\rho} \in (0, 1)$, then (48) implies that the term in the square brackets in (51) is greater than $1 - \overline{a}^i$, which is greater than $1 - \underline{\rho} \geq 0$, because of the necessary condition (30). Hence, the left-hand side of (51) is greater than

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

which reaches its minimum value $\underline{\rho}$ at $\kappa(y) = 0$, by part (iii) of Lemma 12. Hence, (51) holds (recall we assume $\underline{\rho} \geq \overline{a}^i$), which subsequently implies (29).

Next, consider NAC/NAM. Note that condition (i), i.e. ϕ_λ/ϕ_μ being weakly decreasing in μ, is equivalent to $\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \geq 0$, as can be seen by taking the derivative with respect to μ. We now distinguish two subcases based on the sign of Δf_y.

If $\Delta f_y \leq 0$, then the left-hand side of (29) is negative. The numerator on the right-hand side is positive because $\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \geq 0$, while the denominator is positive because of the firm's SOC. Thus, it follows immediately that (29) holds with $\leq$.

In contrast, if $\Delta f_y \geq 0$, we prove the inequality

$$1 \leq \frac{1 - \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) \frac{\Delta f_y}{f_y^1}}{1 - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \frac{\Delta f}{f^1}}, \tag{52}$$

which is equivalent to

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \frac{1}{m'} \left(\phi_\mu \frac{\phi_{\mu\lambda}}{\phi_{\mu\mu}} - \phi_\lambda \right) - \frac{1}{m''} \left(\frac{\phi_{\mu\lambda}^2}{\phi_{\mu\mu}} - \phi_{\lambda\lambda} \right) \leq 0. \tag{53}$$

To do so, note that $\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \leq \frac{(1+\kappa(y))^{\overline{\rho}}-1}{\kappa(y)} \leq \overline{\rho} \leq \underline{a}^i$, where the three inequalities follow from (50), part ii) of Lemma 12, and our assumption $\overline{\rho} \leq \underline{a}^i$, respectively. Hence, (53) and (52) follow from (49). Therefore,

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

Hence, we have proved the case of NAC/NAM. □

A.8.3 Verification of the Regularity Conditions in Lemma 13

If $\phi(\mu, \lambda)$ satisfies (7), $\overline{a}^c = \overline{a}^m = (1+\sigma)/2$ and $\underline{a}^c = \underline{a}^m = (1-\sigma)/2$, by Lemma 8. Plugging (7) into the left-hand side of (48) yields

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

where the first inequality is because the denominator reaches its maximum at $\mu = \lambda$, and the second one is because $1 + \lambda \geq 1 + (1-\sigma)\lambda$. This proves (48).

The regularity conditions for NAC/NAM in Lemma 13 are satisfied because (i) $\phi_\lambda/\phi_\mu = -(1-\sigma)\mu/(1+(1-\sigma)\lambda)$ which is decreasing in μ, (ii) $\underline{a}^i = (1-\sigma)/2 \in [0, 1)$, and (iii) plugging (7) into the left-hand side of (49) yields

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

Hence the necessary conditions in Proposition 3 are also sufficient. □

Appendix B Online Appendix

B.1 Proof of Lemma 2

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

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

where the first equality uses the definition of invariance and the fact that the probability that an applicant is high-type is μ/λ and is independent across applicants, while the second and the third equality follow from summation by parts.

A firm's *potential* number of interviews, n_C, follows a geometric distribution with support $\mathbb{N}_1$ and mean $(1-\sigma)^{-1}$. That is, $\mathbb{P}[n_C \geq n \mid \sigma] = \sigma^{n-1}$ for $n = 1, 2, \ldots$. Since interviewing might be constrained by the number of applications, the firm's *actual* number of interviews is $n_I = \min\{n_A, n_C\} \in \mathbb{N}_0$, distributed according to $\mathbb{P}[n_I \geq n \mid \lambda, \sigma] = \mathbb{P}[n_A \geq n \mid \lambda] \sigma^{n-1}$. An interview reveals a high-type worker with probability μ/λ, independently across applicants. The firm therefore interviews at least one high-type worker with probability

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

where the second equality follows from summation by parts, analogous to (54). Substituting $\mathbb{P}[n_I \geq n \,|\, \lambda, \sigma] = \mathbb{P}[n_A \geq n \,|\, \lambda] \, \sigma^{n-1}$ yields

$$\begin{aligned}\phi(\mu, \lambda) &= \frac{\mu}{\sigma\mu + (1-\sigma)\lambda} \sum_{n=1}^{\infty} \mathbb{P}[n_A \geq n \,|\, \lambda] \frac{\sigma\mu + (1-\sigma)\lambda}{\lambda} \left(1 - \frac{\sigma\mu + (1-\sigma)\lambda}{\lambda}\right)^{n-1} \\ &= \frac{\mu}{\sigma\mu + (1-\sigma)\lambda} \mathbb{P}[n_A \geq 1 \,|\, \sigma\mu + (1-\sigma)\lambda] \,. \end{aligned} \tag{55}$$

Since $\mathbb{P}[n_A \geq 1 \,|\, \sigma\mu + (1-\sigma)\lambda] = \frac{\sigma\mu + (1-\sigma)\lambda}{1 + \sigma\mu + (1-\sigma)\lambda}$ in our baseline model, equation (7) then follows from (55). □

B.2 Proof of Lemma 3

Given U_1/w_1 and U_2/w_2, consider then the level curves $\psi_2(\lambda\zeta, \lambda) = U_2/w_2$ and $\psi_1(\lambda\zeta, \lambda) = U_1/w_1$ in the λ-ζ space. Note that

$$\psi_1(\lambda\zeta, \lambda) = \frac{1 + (1-\sigma)\lambda}{(1+\lambda)(1 + (1 - \sigma + \sigma\zeta)\lambda)} \quad \text{and} \quad \psi_2(\lambda\zeta, \lambda) = \frac{1}{1 + (1 - \sigma + \sigma\zeta)\lambda},$$

both of which are strictly decreasing in ζ. We now show that the two curves intersect at most once so that there exists exactly one solution (μ, λ). At any intersection point, the difference between the slopes of the two level curves is

$$-\frac{\partial\psi_1(\lambda\zeta, \lambda)/\partial\lambda}{\partial\psi_1(\lambda\zeta, \lambda)/\partial\zeta} + \frac{\partial\psi_2(\lambda\zeta, \lambda)/\partial\lambda}{\partial\psi_2(\lambda\zeta, \lambda)/\partial\zeta} = \frac{1 + (1 - \sigma + \sigma\zeta)\lambda}{\lambda(\lambda+1)(1 + (1-\sigma)\lambda)} > 0.$$

Hence, by a standard single-crossing argument, the two level curves cross each other at most once. Note that we can also derive the solution (μ, λ) explicitly. However, with this approach we need to discuss the conditions under which we have a corner solution ($\mu = 0$ or $\mu = \lambda$) or an interior solution ($0 < \mu < \lambda$). □

B.3 Proof of Lemma 6

We prove this result and discuss it extensively in Cai et al. (2020). Here, we state the single-crossing condition and briefly argue why it leads to Lemma 6. To do so, we define $H(\mu, \lambda)$ as the right-hand side of (16), i.e.

$$H(\mu, \lambda) \equiv \frac{\phi_{\lambda\lambda} - \phi_{\mu\lambda}^2/\phi_{\mu\mu}}{-m''}. \tag{56}$$

Cai et al. (2020) then show that Lemma 6 holds whenever a meeting technology satisfies Property A0, A1, A2 and the following A3.

A3. (single-crossing condition) At any point (ζ, λ) where $H(\lambda\zeta, \lambda) > 0$, we have $\partial H(\lambda\zeta, \lambda)/\partial\lambda > 0$ and

$$-\frac{\partial\phi_\mu(\lambda\zeta, \lambda)/\partial\zeta}{\partial\phi_\mu(\lambda\zeta, \lambda)/\partial\lambda} < -\frac{\partial H(\lambda\zeta, \lambda)/\partial\zeta}{\partial H(\lambda\zeta, \lambda)/\partial\lambda}. \tag{57}$$

Note that Property A0 states that $\partial\phi_\mu(\lambda\zeta, \lambda)/\partial\zeta < 0$, while Property A2 states that $\partial\phi_\mu(\lambda\zeta, \lambda)/\partial\lambda < 0$, making the left-hand side of (57) strictly negative. When $\phi(\mu, \lambda)$ is given by (7), direct computation reveals that both $H(\lambda\zeta, \lambda)$ and the right-hand side of (57) are strictly positive. Thus, Property A3 is trivially satisfied in this case.

The idea of the proof of Cai et al. (2020) is then as follows. Suppose that the equilibrium payoff—or equivalently the marginal contribution to surplus—of a given firm y is $R^*(y)$. Property A3 then implies that level curve $R(\lambda\zeta, \lambda, y) = R^*(y)$ crosses the level curve $H(\lambda\zeta, \lambda) = 1/\kappa(y)$ at most once and from the left, as illustrated in Figure 1 of Cai et al. (2020). If the intersection exists, denote it by (λ^*, ζ^*). Along the level curve $R(\lambda\zeta, \lambda, y) = R^*(y)$, the second-order condition (16) is then satisfied for $\zeta > \zeta^*$ and violated for $\zeta < \zeta^*$. The only feasible submarket when $\zeta < \zeta^*$ is therefore the corner solution $\zeta = 0$. Furthermore, along the level curve $R(\lambda\zeta, \lambda, y) = R^*(y)$, $T_2(\lambda\zeta, \lambda, y)$ is monotonically decreasing in ζ for $\zeta \geq \zeta^*$. Since the marginal contribution of high-type workers must be the same among all submarkets containing such workers, there can exist only one submarket with $\zeta \geq \zeta^*$. Hence, there exist at most two submarkets: one with $\zeta = 0$ and the other with $\zeta \geq \zeta^*$. □

B.4 About the Second Term on the Right-Hand Side of (27)

Equation (27) characterizes how firms of different types take advantage of production complementarity by adjusting their queue length and queue composition. The second term on the right-hand side reflects the complicated interaction between meeting externalities and production complementarity and is also present in the sorting condition (29). Below we analyze whether it is greater than 1.

First, consider the case $\underline{\rho} \geq (1 + \sigma)/2$, under which condition the equilibrium exhibits PAC/PAM by Proposition 4. Note that the second term on the right-hand side of equation (27) is less than 1 if and only if

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

where the right-hand side is smaller than $(1+\sigma)/2$ by equation (48). When worker types are close, the left-hand side of the above inequality is approximately $\rho(x, y)$, which is greater than $\underline{\rho}$ and hence greater than the right-hand side. Thus, in this case, the term under consideration is smaller than 1.

Next, consider the case $\overline{\rho} \leq (1-\sigma)/2$, under which condition the equilibrium exhibits NAC/NAM by Proposition 4. Equation (52) in the proof of Proposition 4 states that the term under consideration is greater than 1. □

B.5 General Analysis with Multiplicity Points

For our sorting analysis, we need to distinguish between the optimal queue composition $\zeta(y)$ being (i) unique and interior, and (ii) not unique. The first case has been analyzed in Section 4.3 and 4.4; here, we complete the analysis by discussing the second case. We do not discuss the case in which $\zeta(y)$ is unique but a corner solution, i.e. $Z(y) = \{0\}$ or $Z(y) = \{1\}$, because $\zeta(\cdot)$ is not necessarily differentiable at such points. However, as Lemma 14 below shows, we do not need to consider such points for our global sorting analysis because of continuity. We focus on the conditions for PAC/PAM; the analysis for NAC/NAM is similar with reversal of the relevant inequalities and is spelled out only when necessary.

Local PAC/PAM. Consider a point y_m where $Z(y_m)$ is not a singleton. By Lemma 6, it then has two elements, with one of them equal to zero. Denote these elements by ζ_0 and ζ_1, satisfying $0 = \zeta_0 < \zeta_1$, and their corresponding queue lengths by λ_0 and λ_1, respectively. By the Theorem of the Maximum, $Z(y)$ is an upper hemi-continuous correspondence. Therefore, for firm types y close to y_m, $Z(y)$ is either the corner solution 0, or some ζ close to ζ_1, or both. We now define when PAC/PAM holds locally at a multiplicity point y_m. Note that the definition is the same for PAC and PAM.

Definition 6. *Consider a point y_m for which $Z(y_m)$ is not unique. PAC/PAM then holds locally at y_m if, for y sufficiently close to y_m, the corner solution 0 is the only solution for $y < y_m$, but does not belong to $Z(y)$ for $y > y_m$.*

In other words, local PAC/PAM for a multiplicity point means that firms with y slightly below y_m have a unique optimal ζ equal to zero and firms with y slightly above y_m have a unique optimal ζ close to ζ_1. This definition is local in the sense that we only require that the previous statement holds for y sufficiently close to y_m. Note that the definition does *not* require that, for y slightly above y_m, we have $\zeta(y) \geq \zeta_1$ (PAC) or $h(\zeta(y), \lambda(y)) \geq h(\zeta_1, \lambda_1)$ (PAM); these inequalities are implied the fact that we also require PAC/PAM to hold locally for $y > y_m$ (see Lemma 14 below for the detailed proof).

From Local to Global. Our next result then shows that if PAC/PAM holds locally according to our Definition 5 and 6, then the equilibrium exhibits PAC/PAM; alternatively, we say that PAC/PAM holds globally in that case.

Lemma 14. *If PAC/PAM (resp. NAC/NAM) holds locally at all points where $Z(y)$ is either unique and interior or contains multiple points, then PAC/PAM (resp. NAC/NAM) holds globally.*

Proof. By the Theorem of Maximum, $Z(y)$ is an upper hemi-continuous correspondence: if $y_k \to y^*$ and $\zeta_k \to \zeta^*$ with $\zeta_k \in Z(y_k)$, then $\zeta^* \in Z(y^*)$. As we discussed after Definition 6, when PAC/PAM holds locally at some multiplicity point y_m, then all points sufficiently close to y_m, but not equal to y_m, have a unique optimal ζ. Thus, if PAC/PAM holds locally at all multiplicity points, then multiplicity points are isolated from each other.

Next, we prove that there exists at most one multiplicity point. Suppose otherwise. Because multiplicity points are isolated, consider two consecutive multiplicity points y'_m and y''_m with $Z(y'_m) = \{0, \zeta'_1\}$ and $Z(y''_m) = \{0, \zeta''_1\}$. By construction, $Z(y)$ is then unique for all y between y'_m and y''_m and hence continuous. Furthermore, since PAC/PAM holds locally at the two points y'_m and y''_m, we have $\lim_{y \downarrow y'_m} = \zeta'_1 > 0$ (left limit) and $\lim_{y \uparrow y''_m} = 0$ (right limit).

If $\zeta(y)$ is interior for all $y \in (y'_m, y''_m)$, then by assumption $\zeta'(y) \geq 0$ and hence $\zeta(y)$ is continuously increasing, which contradicts with $\lim_{y \downarrow y'_m} = \zeta'_1 > 0$ and $\lim_{y \uparrow y''_m} = 0$. More generally, if $\zeta(y)$ is a corner solution for some $y \in (y'_m, y''_m)$, then $\zeta(y)$ is piecewise differentiable, and whenever it is differentiable, it is (weakly) increasing. We again have a contradiction. Similar logic applies for the NAC/NAM case. □

Proof of Proposition 4 with Multiplicity Points. We now extend the proof of Proposition 4 to allow for multiplicity points. By Lemma 14, we only need to show that $\underline{\rho} \geq (1+\sigma)/2$ and $\overline{\rho} \leq (1-\sigma)/2$ are sufficient for PAC/PAM and NAC/NAM to hold locally at multiplicity points, respectively. For the former, below we show that ($\underline{\rho} > 1/2$) is already sufficient.

Suppose that $Z(y_m)$ is not a singleton and denote the two optimal queues again by (ζ_0, λ_0) and (ζ_1, λ_1), where $0 = \zeta_0 < \zeta_1$. Since firm y_m must be indifferent, the expected payoff—or, equivalently, the marginal contribution to surplus—must be the same for the two queues. By (14), we have

$$m(\lambda_0) - \lambda_0 m'(\lambda_0) = m(\lambda_1) - \lambda_1 m'(\lambda_1) + \left(\phi(\zeta_1\lambda_1, \lambda_1) - \lambda_1 \frac{d\phi(\zeta_1\lambda_1, \lambda_1)}{d\lambda}\right) \frac{\Delta f}{f^1}, \quad (58)$$

where $\Delta f = f(x_2, y_m) - f(x_1, y_m)$ and $f^1 = f(x_1, y_m)$. The left-hand side is the firm's marginal contribution to surplus with a queue $(0, \lambda_0)$, divided by $f(x_1, y_m)$, and the right-hand side is the corresponding value with a queue (ζ_1, λ_1).

If $\zeta_1 \in (0, 1)$, then low-type workers are present in both queues and their marginal contribution to surplus must be the same. Equation (12) then yields

$$m'(\lambda_0) = m'(\lambda_1) + \phi_\lambda(\zeta_1\lambda_1, \lambda_1)\frac{\Delta f}{f^1} \quad \text{if } \zeta_1 \in (0, 1). \quad (59)$$

Low-type workers are not present in the shorter queue if $\zeta_1 = 1$. In this special case, optimality requires that the left-hand side of (59) is larger than the right-hand side.

The requirement in Definition 6 can be characterized by the envelope theorem: a *sufficient* condition is $\Pi_y(0, \lambda_0, y_m) < \Pi_y(\zeta_1, \lambda_1, y_m)$, where $\Pi(\zeta, \lambda, y)$ is the expected profit of a firm y with queue (ζ, λ) as defined in equation (11).[42] The condition $\Pi_y(0, \lambda_0, y_m) < \Pi_y(\zeta_1, \lambda_1, y_m)$ can be written as

$$m(\lambda_0) < m(\lambda_1) + \phi(\zeta_1\lambda_1, \lambda_1)\frac{\Delta f_y}{f_y^1}, \quad (60)$$

where $\Delta f_y = f(x_2, y_m) - f(x_1, y_m)$ and $f_y^1 = f_y(x_1, y_m)$. If the reverse inequality

[42] The envelope theorem states that if a firm with type y close to y_m is constrained to choose only ζ close to ζ_1, then its maximum expected profit is approximately (first-order) $\Pi(\zeta_1, \lambda_1, y_m) + \Pi_y(\zeta_1, \lambda_1, y_m)\Delta y$ where $\Delta y = y - y_m$. Similarly, if the firm is constrained to choose $\zeta = 0$, then its maximum expected profit is approximately $\Pi(0, \lambda_0, y_m) + \Pi_y(0, \lambda_0, y_m)\Delta y$. Recall that $\Pi(\zeta_1, \lambda_1, y_m) = \Pi(0, \lambda_0, y_m)$. When $\Pi_y(\zeta_1, \lambda_1, y_m) > \Pi_y(0, \lambda_0, y_m)$, then a firm type $y > y_m$ strictly prefers to choose ζ around ζ_1 instead of zero, and a firm type $y < y_m$ strictly prefers zero. As mentioned before, by continuity, it is without loss of generality to constrain the firm to choose between zero and all ζ close to ζ_1. See Milgrom and Segal (2002) for a discussion of envelope theorems.

$\Pi_y(0, \lambda_0, y_m) > \Pi_y(\zeta_1, \lambda_1, y_m)$ holds, then the opposite is true (NAC/NAM holds locally at point y_m), i.e. for y slightly above y_m, $Z(y)$ is unique and equals zero, and for y slightly below y_m, $Z(y)$ is unique and equals some ζ around ζ_1.

Consider first the case in which $\zeta_1 = 1$. The sorting conditions for this case are equivalent to the ones in Eeckhout and Kircher (2010a), even though the meeting technology is not bilateral. The proof of Proposition 8 in Appendix B.7 shows that equation (58) implies (60) when $\underline{\rho} > 1/2$, while equation (58) implies (60) with reverse inequality when $\underline{\rho} < 1/2$. Hence, assortative sorting holds locally at y_m under the sufficient conditions in Proposition 4.

Next, consider the case in which $\zeta_1 < 1$, such that (59) holds with equality. From (58) and (59), we can solve for $\kappa(y_m)$ and λ_0 in terms of ζ_1 and λ_1. This yields

$$\kappa(y_m) = \frac{4\sigma(1+\lambda_1-\lambda_1\sigma(1-\zeta_1))^2}{(1+\lambda_1)(\lambda_1-\sigma-\lambda_1\sigma(1-\zeta_1)+1)^2}, \tag{61}$$

$$\lambda_0 = \frac{\lambda_1(\lambda_1+\sigma(-\lambda_1+(\lambda_1+2)\zeta_1-1)+1)}{1-\sigma-\lambda_1(1-\sigma-\sigma\zeta_1)}. \tag{62}$$

Rewrite (60) as

$$1 + \frac{m(\lambda_0)-m(\lambda_1)}{\phi(\zeta_1\lambda_1, \lambda_1)} < \frac{f_y(x_2, y_m)}{f_y(x_1, y_m)}. \tag{63}$$

Consider PAC/PAM first. If $\underline{\rho} > 1/2$, then $f_y(x_2, y_m)/f_y(x_1, y_m) > (1+\kappa(y_m))^{1/2}$ by (2). Note that

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

hence (63) holds.

For NAC/NAM, note that $f_y(x_2, y_m)/f_y(x_1, y_m) < (1+\kappa(y_m))^{\overline{\rho}}$ by (2). We have

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

where the first inequality follows from $(1+\kappa)^{\overline{\rho}} < 1+\overline{\rho}\kappa \leq 1 + \frac{1-\sigma}{2}\kappa$, and the equality follows from equations (61) and (62). Hence, (60) holds with $>$. We thus have proved that assortative sorting holds locally at all multiplicity points under the sufficient condition in Proposition 4. □

B.6 Proof of Lemma 9

Since $\phi(\mu, \lambda) = \mu m(\lambda)/\lambda$, we have $\phi_\mu(\mu, \lambda) = m(\lambda)/\lambda$, which in turn implies that $\partial \phi_\mu(\zeta\lambda, \lambda)/\partial \zeta = 0$. Therefore, by equation (25), $a^m(\zeta, \lambda) = a^c(\zeta, \lambda)$ for any ζ and λ. Next, consider $a^c(\zeta, \lambda)$ as given by equation (24). Since $\phi_{\mu\mu}(\mu, \lambda) = 0$ and $\phi_{\mu\lambda}(\mu, \lambda) = (m(\lambda)/\lambda)' = (\lambda m'(\lambda) - m(\lambda))/\lambda^2$, we obtain equation (34) for $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$. A discussion of why the right-hand side in equation (34) equals the elasticity of substitution of the total number of matches in a submarket can be found in Eeckhout and Kircher (2010a). □

B.7 General Bilateral Technologies

Under bilateral technologies, the firm's problem is solved by attracting either low- or high-type workers, but not both.[43] Therefore, we do not need need to consider the scenario where $Z(y)$ is unique and interior; we only need to consider multiplicity points, where the optimal ζ is 0 or 1, but not any number in between. Denote such a multiplicity point by y^{EK}.[44] Using $\zeta_1 = 1$, the indifference condition (58) of a firm with type y^{EK} then becomes

$$(m(\lambda_0) - \lambda_0 m'(\lambda_0))\, f(x_1, y^{EK}) = (m(\lambda_1) - \lambda_1 m'(\lambda_1))\, f(x_2, y^{EK}). \tag{64}$$

Further, the local condition for PAC/PAM (60) reduces to

$$m(\lambda_1)\frac{f_y(x_2, y^{EK})}{f_y(x_1, y^{EK})} > m(\lambda_0), \tag{65}$$

i.e. firms with types slightly above y^{EK} strictly prefer high-type workers, while firms with types slightly below y^{EK} strictly prefer low-type workers. If this holds for all multiplicity points, then there exists at most one such point and PAC/PAM holds globally (see Lemma 14). The following proposition establishes the necessary and sufficient conditions for robust sorting. These conditions were first derived by Eeckhout

[43] This result follows from firms' SOC: with bilateral meetings, $\phi_{\mu\mu} = 0$; the right-hand side of (16) is therefore zero and never satisfied for an interior ζ.

[44] The superscript refers to Eeckhout and Kircher (2010a), who first analyzed the bilateral case.

and Kircher (2010a) in a framework with a continuum of worker types; we establish that the same conditions arise with two types.

Proposition 8 (Eeckhout and Kircher, 2010a)**.** *Suppose meetings are bilateral, i.e.* $\phi(\mu, \lambda) = m(\lambda)\mu/\lambda$. *The equilibrium then exhibits robust PAC/PAM if (resp. only if)* $\underline{\rho}$ *is strictly (resp. weakly) larger than* $\overline{a}^c = \overline{a}^m = \overline{a}^{EK}$. *In contrast, the equilibrium exhibits robust NAC/NAM if (resp. only if)* $\overline{\rho}$ *is strictly (resp. weakly) smaller than* $\underline{a}^c = \underline{a}^m = \underline{a}^{EK}$.

For our benchmark technology with $\sigma = 0$, we have $a^{EK}(\lambda) = 1/2$ for any λ. Proposition 8 therefore reveals that Proposition 3 gives the right sorting conditions for this special case (by continuity), even though Proposition 3 was derived under the assumption that the meeting technology is *not* bilateral ($\sigma > 0$) to avoid the technical issue of division by $\phi_{\mu\mu}(\mu, \lambda) = 0$.[45]

Proof. We focus on PAC/PAM; the logic for NAC/NAM is similar. Consider first the sufficient condition. We need to show that equation (64) implies (65) when $\underline{\rho} > \overline{a}^{EK}$. To see this, take logs to rewrite equation (64) as

$$\log f\left(x_2, y^{EK}\right) - \log f\left(x_1, y^{EK}\right) = \log\left(m(\lambda_0) - \lambda_0 m'(\lambda_0)\right) - \log\left(m(\lambda_1) - \lambda_1 m'(\lambda_1)\right),$$

which, by the fundamental theorem of calculus, can be rewritten as

$$\int_{x_1}^{x_2} \frac{f_x(x, y^{EK})}{f(x, y^{EK})} dx = \int_{\lambda_1}^{\lambda_0} \frac{-\lambda m''(\lambda)}{m(\lambda) - \lambda m'(\lambda)} d\lambda. \tag{66}$$

By the definition of $\underline{\rho}$, we have $\frac{f_{xy}(x,y^{EK})}{f_y(x,y^{EK})} - \underline{\rho}\frac{f_x(x,y^{EK})}{f(x,y^{EK})} \geq 0$. Similarly, by the definition of $\overline{a}^{EK}$, we have $0 \geq \frac{m'(\lambda)}{m(\lambda)} - \overline{a}^{EK}\frac{-\lambda m''(\lambda)}{m(\lambda)-\lambda m'(\lambda)}$, where the right-hand side is strictly greater than $\frac{m'(\lambda)}{m(\lambda)} - \underline{\rho}\frac{-\lambda m''(\lambda)}{m(\lambda)-\lambda m'(\lambda)}$, because we assume $\underline{\rho} > \overline{a}^{EK}$. We thus have

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

[45] There is a subtlety: in Proposition 3, when $\sigma > 0$, the necessary and sufficient conditions exactly coincide, because the supremum and the infimum are never attained as maximum and minimum, respectively. However, when $\sigma = 0$, we have $a^c(\zeta, \lambda) = a^c(\zeta, \lambda) = a^{EK}(\lambda) = 1/2$ for any ζ, λ, making firms indifferent between low- and high-type workers when f is CES with $\rho = 1/2$, as can been seen from (38). This indifference makes it necessary to distinguish between necessary and sufficient conditions for bilateral technologies.

because the integrand on the left-hand side is weakly positive, and integrand on the right-hand side is strictly negative. Combing the above inequality with (66) yields

$$\int_{x_1}^{x_2} \frac{f_{xy}(x,y)}{f_y(x,y)} dx > \int_{\lambda_1}^{\lambda_0} \frac{m'(\lambda)}{m(\lambda)} d\lambda, \tag{67}$$

which is the same as (65), by the same logic as what gave us (66) from (64).

Next, consider the necessary condition.[46] Towards a contradiction, suppose that $\underline{\rho} < \overline{a}^{EK}$. Then there exist an x^*, y^* and λ^* such that $\rho(x^*, y^*) < a^{EK}(\lambda^*)$, and by continuity there exists some $\epsilon > 0$ such that $\rho(x, y^*) < a^{EK}(\lambda)$ holds for all x and λ with $|x - x^*|, |\lambda - \lambda^*| < \epsilon$. To derive the contradiction, we then construct an endowment of agents such that NAC/NAM holds globally (i.e. PAC/PAM fails).

Step 1: Set $y^{EK} = y^*$; by continuity, we can find some $\Delta x, \Delta\lambda < \epsilon$ such that $x_1 = x^* - \Delta x$, $x_2 = x^* + \Delta x$, $\lambda_0 = \lambda^* + \Delta\lambda$, $\lambda_1 = \lambda^* - \Delta\lambda$ and equation (64) holds.

Step 2: Set the market utilities $U_1 = m'(\lambda_0) f(x_1, y^{EK})$ and $U_2 = m'(\lambda_1) f(x_2, y^{EK})$.

Step 3: Define $\pi_i(y)$, $i = 1, 2$, as the maximum expected profit by attracting workers of type x_i only. That is, $\pi_i(y) = \max_{\lambda \geq 0} m(\lambda) f(x_i, y) - \lambda U_i$, where U_i is given by Step 2. By construction, we have then $\pi_1(y^{EK}) = \pi_2(y^{EK})$, which then implies that $\pi_1'(y^{EK}) = m(\lambda_0) f(x_1, y^{EK}) > m(\lambda_1) f(x_2, y^{EK}) = \pi_2'(y^{EK})$ by the same logic that equation (64) implies (65), except that now we reverse all the inequalities and only require that $\rho(x, y^{Ek}) < a^{EK}(\lambda)$ for $x_1 \leq x \leq x_2$ and $\lambda_1 \leq \lambda \leq \lambda_0$. Hence, by continuity, we can find a Δy small enough such that $\pi_1(y) < \pi_2(y)$ for $y \in \left[y^{EK} - \Delta y, y^{EK}\right)$ and $\pi_1(y) > \pi_2(y)$ for $y \in \left(y^{EK}, y^{EK} + \Delta y\right]$.

Step 4: Choose any firm type distribution $J(y)$ on $[y^{EK} - \Delta y, y^{EK} + \Delta y]$. When $y > y^{EK}$, the demand of labor is given by the FOC $m'(\lambda) f(x_2, y) = U_2$, and when $y < y^{EK}$, the demand of labor is given by the FOC $m'(\lambda) f(x_1, y) = U_1$.

Step 5: Set the measure of workers (ℓ_1, ℓ_2) equal to firms' demand of labor, which then ensures that U_1 and U_2 are indeed the equilibrium market utilities of workers. □

B.8 Proof of Lemma 10

The desired expression for a^c follows readily from equations (24) and (35). To derive the expression for a^m, note that $\phi(\mu, \lambda) = m(\mu)$ implies that $\phi_\mu(\mu, \lambda) = m'(\mu)$ and

[46] A proof of necessity is required because Proposition 3 restricts attention to the case in which $\phi(\mu, \lambda)$ is strictly concave in μ, implying that the meeting technology is not bilateral.

$h(\zeta, \lambda) = m(\zeta\lambda)/m(\lambda)$. Therefore, the last factor in (25) can be rewritten as

$$1 - \frac{\partial \phi_\mu/\partial \zeta}{\partial \phi_\mu/\partial \lambda} \frac{\partial h/\partial \lambda}{\partial h/\partial \zeta} = 1 - \frac{\lambda m''(\zeta\lambda)}{\zeta m''(\zeta\lambda)} \frac{\frac{\zeta m'(\zeta\lambda)m(\lambda) - m(\zeta\lambda)m'(\lambda)}{m(\lambda)^2}}{\lambda m'(\zeta\lambda)/m(\lambda)} = \frac{m(\zeta\lambda)m'(\lambda)}{\zeta m'(\zeta\lambda)m(\lambda)} = \frac{\varepsilon_0(\lambda)}{\varepsilon_0(\zeta\lambda)}.$$

Note that $a^c(1, \lambda) = a^m(1, \lambda) = 1$ which implies that $\overline{a}^c, \overline{a}^m \geq 1$.

Next, consider $\underline{a}^c$ and $\underline{a}^m$. Since $m(\mu)$ is strictly concave and strictly increasing, $\varepsilon_0(\mu)$ is strictly positive, and $\varepsilon_1(\mu)$ is strictly negative when $\mu > 0$. Hence, $a^c(\zeta, \lambda)$ and $a^m(\zeta, \lambda)$ are always nonnegative. By L'Hopital's rule and $m'(0) > 0$, we have $\lim_{\mu \to 0} \varepsilon_1(\mu) = \lim_{\mu \to 0} \mu m''(\mu)/m'(\mu) = 0$, where, as we argued in footnote 38, $\lim_{\mu \to 0} \mu m''(\mu) = 0$ for invariant technologies. Similarly, $\lim_{\mu \to 0} \varepsilon_0(\mu) = \lim_{\mu \to 0} \mu m'(\mu)/m(\mu) = \lim_{\mu \to 0} 1 + \mu m''(\mu)/m'(\mu) = 1$. Thus, $\lim_{\zeta \to 0} a^c(\zeta, \lambda) = \lim_{\zeta \to 0} \varepsilon_1(\lambda\zeta)/\varepsilon_1(\lambda) = 0$, and $\lim_{\zeta \to 0} a^m(\zeta, \lambda) = \lim_{\zeta \to 0} \varepsilon_1(\lambda\zeta)/\varepsilon_1(\lambda) \cdot \varepsilon_0(\lambda)/\varepsilon_0(\lambda\zeta) = 0$. Hence, $\underline{a}^c = \underline{a}^m = 0$. □

B.9 Proof of Proposition 5

The necessary conditions directly follows from Proposition 3, as its proof is valid for any non-bilateral technology (i.e. ϕ is strictly concave in μ). Hence, we only need to prove sufficiency. As mentioned in the main text, when the technology is invariant, the firms' problem is strictly concave so that the solution is always unique. For PAC/PAM, we therefore only need to verify (29), which now reduces to

$$\frac{f^1 \Delta f_y}{f_y^1 \Delta f} \geq a^i. \tag{68}$$

where we have used the fact that $\phi_\lambda(\mu, \lambda) = 0$ and hence $\phi_{\mu\lambda} = \phi_{\lambda\lambda} = 0$.

Note that $\underline{\rho} \geq \overline{a}^i$ by assumption. Further, $\overline{a}^i \geq 1$, by Lemma 10. Therefore,

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

where the first inequality is because of (2), and the second inequality follows from part i) of Lemma 12.

Next, consider NAC/NAM, where we have assumed that $\overline{\rho} \leq 0 = \underline{a}^c = \underline{a}^m$ (see

Lemma 10). Again by (2), we have

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

□

B.10 Invariant Technologies with N Worker Types

Our analysis of invariant technologies can easily be extended to the case in which there are $N > 2$ worker types, i.e. $x_1 < x_2 < \cdots < x_N$. To do so, define μ_i as the queue length of workers with type x_i *or higher*, for $i = 1, 2, \ldots, N$. That is, the queue length of workers of type x_i is $\mu_i - \mu_{i+1}$, with the convention that $\mu_{N+1} = 0$, and the total queue length is μ_1.

Consider then a firm of type y that faces a queue $(\mu_1, \mu_2, \ldots, \mu_N)$. With probability $m(\mu_1)$, the firm meets at least one worker, which generates a surplus of at least $f(x_1, y)$; with probability $m(\mu_2)$ the firm meets at least one worker with a type higher than or equal to x_2, which generates an additional surplus of at least $f(x_2, y) - f(x_1, y)$, and so on. Using the convention $f(x_0, y) = 0$, expected surplus therefore equals

$$S(\mu_1, \ldots, \mu_N, y) = \sum_{i=1}^{N} m(\mu_i)[f(x_i, y) - f(x_{i-1}, y)], \tag{69}$$

which generalizes equation (8) and is strictly concave in $(\mu_1, \ldots, \mu_N)$.

From equation (69), we can proceed as before. Adding more workers of type x_i increases $\mu_1, \ldots, \mu_i$ simultaneously, which implies that the marginal contribution to surplus of such workers is $\sum_{k=1}^{i} m'(\mu_k)[f(x_k, y) - f(x_{k-1}, y)]$. Since the firm's problem is strictly concave in $(\mu_1, \ldots, \mu_N)$, the optimal queue is unique and is denoted by $(\mu_1(y), \ldots, \mu_N(y))$. Assuming an interior solution to simplify exposition, $\mu_i(y)$ is determined by the FOC $m'(\mu_i(y))[f(x_i, y) - f(x_{i-1}, y)] = U_i - U_{i-1}$, where U_i is the market utility of workers of type x_i, with the convention $U_0 = 0$.

Defining $\zeta_i(y) = \mu_i(y)/\mu_1(y)$ and $\lambda(y) = \mu_1(y)$, differentiation of this FOC along the equilibrium path yields

$$-\frac{m''(\zeta_i(y)\lambda(y))}{m'(\zeta_i(y)\lambda(y))} \left(\zeta_i'(y)\lambda(y) + \zeta_i(y)\lambda'(y)\right) = \frac{f_y(x_i, y) - f_y(x_{i-1}, y)}{f(x_i, y) - f(x_{i-1}, y)},$$

which replicates (26) for the case of invariant technologies. Setting $i = 1$ and using the fact that $\zeta_1(y) = 1$, the above equation implies

$$-\frac{m''(\lambda(y))}{m'(\lambda(y))}\lambda'(y) = \frac{f_y(x_1, y)}{f(x_1, y)},$$

which replicates (27). Combing the above two equations yields

$$\frac{\frac{m''(\zeta_i(y)\lambda(y))}{m'(\zeta_i(y)\lambda(y))}\left(\zeta_i'(y)\lambda(y) + \zeta_i(y)\lambda'(y)\right)}{\frac{m''(\lambda(y))}{m'(\lambda(y))}\lambda'(y)} = \frac{f_y(x_i, y) - f_y(x_{i-1}, y)}{f(x_i, y) - f(x_{i-1}, y)}\frac{f(x_1, y)}{f_y(x_1, y)},$$

which replicates (28). Then, by Lemma 7, the condition for PAC/PAM is

$$\frac{f^1}{f^1_y}\frac{f_y(x_i, y) - f_y(x_{i-1}, y)}{f(x_i, y) - f(x_{i-1}, y)} \geq a^i\left(\zeta_i(y), \lambda(y)\right). \tag{70}$$

where $i = c$ for PAC and $i = m$ for PAM. From Proposition 5 (the case of $N = 2$), we have that $\underline{\rho} \geq \overline{a}^i$ is necessary, now we show that this is also sufficient for any N. Since $f(x, y)$ is log-supermodular, we have $f(x_1, y)/f_y(x_1, y) \geq f(x_{i-1}, y)/f_y(x_{i-1}, y)$. Therefore, we only need to show that

$$\frac{f(x_{i-1}, y)}{f_y(x_{i-1}, y)}\frac{f_y(x_i, y) - f_y(x_{i-1}, y)}{f(x_i, y) - f(x_{i-1}, y)} \geq a^i\left(\zeta_i(y), \lambda(y)\right),$$

which is a replication of equation (68). Hence, by the same argument as in the proof of Proposition 5, Proposition 5 and Corollary 2 continue to hold for general N.

B.11 Proof of Proposition 6

First, we consider the unconditional probability that an applicant generates a positive signal $\widetilde{x}_2$. The probability of this event equals $\mathbb{P}(\widetilde{x}_2) = \frac{\mu}{\lambda} + \frac{\lambda-\mu}{\lambda}(1-\tau)$, and the queue length of such applicants is $\widetilde{\lambda} = \lambda\mathbb{P}(\widetilde{x}_2) = \mu + (\lambda - \mu)(1 - \tau)$. Given a positive signal $(\widetilde{x}_2)$, the probability that an applicant is of high type (x_2) is $\mathbb{P}(x_2 \,|\, \widetilde{x}_2) = \mathbb{P}(x_2)\mathbb{P}(\widetilde{x}_2 \,|\, x_2)/\mathbb{P}(\widetilde{x}_2) = \mu/\widetilde{\lambda}$, where the first equality is simply Bayes' rule.

Next, we consider the probability that the firm interviews at least one high-type worker, $\phi(\mu, \lambda)$. For this, we can ignore the existence of applicants with negative signals; they are low-type workers for sure and do not affect the meeting process between firms and workers with positive signals. By equation (7), the probability that a firm interviews someone from the queue μ of high-type applicants, given a queue $\widetilde{\lambda}$ of applicants with positive signals, is $\phi(\mu, \lambda) = \mu/(1+\sigma\mu+(1-\sigma)\widetilde{\lambda})$, which yields the desired result after substitution of $\widetilde{\lambda}$. $\square$

B.12 Proof of Lemma 11

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

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

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

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

B.13 Proof of Proposition 7

Our proof consists of three parts. First, we provide the details regarding a firm's optimal choice of σ that we omitted from the main text. Subsequently, we move to the analysis of NAC/NAM. The final part concerns PAC/PAM.

B.13.1 Individual Firm's Problem

Consider a firm of type y which thinks about choosing $\sigma = 1$. As we illustrate in Figure 2, there are four possibilities regarding the firm's optimal applicant pool:

(i) *No applicants.* If $f(x_1, y) \leq U_1$ and $f(x_2, y) \leq U_2$, then the firm will not attract any applicants, such that $\overline{\pi}(y) = 0$.

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

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

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

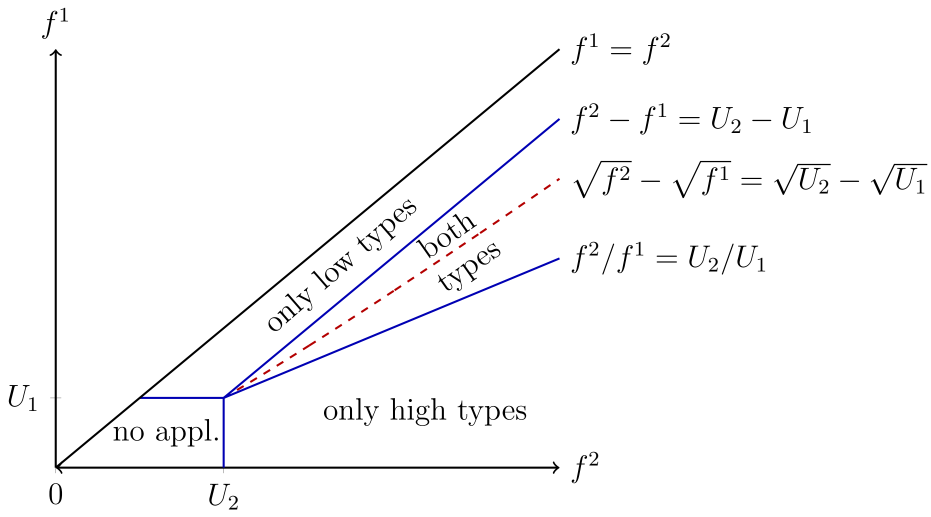

Figure 2: Optimal applicant pool for a firm, conditional on $\sigma = 1$.

U_2/U_1, then the firm strictly prefers a mix of both types of workers in their application pool. By the FOCs, the optimal queue is given by $\mu = \sqrt{\Delta f/\Delta U} - 1$ and $\lambda = \sqrt{f^1/U_1} - 1$. In this case, $\overline{\pi}(y)$ is given by (40).

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

$$\overline{\pi}(y) - c > \max\{\pi_1(y), \pi_2(y)\} \quad \Longleftrightarrow \quad \Delta f > \Delta U \quad \text{and} \quad \frac{f(x_2, y)}{f(x_1, y)} < \frac{U_2}{U_1}. \tag{71}$$

As the red dashed line in Figure 2 shows, the region described by (71) is divided into two parts by the curve $\pi_1(y) = \pi_2(y)$, or equivalently

$$\sqrt{f^2} - \sqrt{f^1} = \sqrt{U_2} - \sqrt{U_1}. \tag{72}$$

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

Lemma 15. *If a firm is indifferent between attracting low- and high-type workers conditional on $\sigma = 0$, i.e. $\pi_1(y) = \pi_2(y)$ or equivalently (72) holds, then this firm attracts both types of workers conditional on $\sigma = 1$, i.e. (71) also holds. In the region* *characterized by (71), the difference in profits between $\sigma = 1$ and $\sigma = 0$ equals*

$$\Delta\pi(y) = \begin{cases} \left(\sqrt{\Delta f} - \sqrt{\Delta U}\right)^2 & \text{if } \pi_1(y) \geq \pi_2(y), \quad (73a) \\ 2\left(\sqrt{f^2 U_2} - \sqrt{f^1 U_1} - \sqrt{\Delta f \Delta U}\right) & \text{if } \pi_1(y) \leq \pi_2(y). \quad (73b) \end{cases}$$

Proof. Equation (72) can be rewritten as $\sqrt{f^2/f^1} - 1 = \sqrt{U_1/f^1}(\sqrt{U_2/U_1} - 1)$. Since $U_1/f^1 < 1$, it follows that $\sqrt{U_2/U_1} - 1 > \sqrt{f^2/f^1} - 1$, and thus $U_2/U_1 > f^2/f^1$. Similarly, (72) can also be rewritten as $(f^2 - f^1)/(\sqrt{f^2} + \sqrt{f^1}) = (U_2 - U_1)/(\sqrt{U_2} + \sqrt{U_1})$. Because $f^1 > U_1$ and $f^2 > U_2$, we have $\Delta f > \Delta U$. Hence, (71) holds. Equation (73) then follows from substituting the relevant version of (38) into $\Delta\pi(y) = \overline{\pi}(y) - \max\{\pi_1(y), \pi_2(y)\}$. □

The characterization of $\Delta\pi(y)$ completes the analysis of the firm's choice problem given by (37): the firm's optimal σ is 1 if $\Delta\pi(y) > c$, 0 if $\Delta\pi(y) < c$, and indeterminate in the knife-edge case $\Delta\pi(y) = c$. If the optimal σ is 1, then the optimal (μ, λ) must be interior, and given by $\mu = \sqrt{\Delta f/\Delta U} - 1$ and $\lambda = \sqrt{f^1/U_1} - 1$. When the optimal σ is 0, then the firm will attract either only low-type or only high-type workers, depending on whether $\sqrt{f^2} - \sqrt{f^1}$ is larger than $\sqrt{U_2} - \sqrt{U_1}$, as discussed after (38).

B.13.2 The Analysis of NAC/NAM

As mentioned in the main text, necessity of submodularity of $f(x, y)$ for NAC/NAM follows from the special case $c = 0$ (see Proposition 4). Next, we show that strict submodularity of $f(x, y)$ is sufficient for NAC/NAM. From the discussion after equation (38), it follows that when $f(x, y)$ is strictly submodular, and thus strictly square-root submodular, there exists a unique y^{EK} which solves (72). Furthermore, $\pi_2(y) > \pi_1(y)$ for firms with $y < y^{EK}$, and vice versa.

Since f is strictly submodular, both $f^2 - f^1$ and f^2/f^1 are strictly decreasing in y. The first part of Lemma 15 states that y^{EK} must belong to the region characterized by (71). There exists at most one $y' < y^{EK}$ such that $f(x_2, y')/f(x_1, y') = U_2/U_1$ (otherwise set $y' = \underline{y}$), and at most one $y'' > y^{EK}$ such that $f(x_2, y'') - f(x_1, y'') = U_2 - U_1$ (otherwise set $y'' = \overline{y}$). The region characterized by (71) is thus $y \in (y', y'')$. The following Lemma establishes that $\Delta\pi(y)$ is single-peaked at $y = y^{EK}$.

Lemma 16. *Suppose that $f(x, y)$ is strictly submodular. In the region characterized by (71), $\Delta\pi(y)$ is strictly increasing in y for $y \leq y^{EK}$ and strictly decreasing in y for*

$y \geq y^{EK}$.

Proof. For submodular f, $\pi_2(y) > \pi_1(y)$ if $y < y^{EK}$, and vice versa. As we remarked before, the region characterized by (71) is (y', y''), which contains y^{EK}. Hence,

$$\Delta\pi'(y) = \begin{cases} \left(1 - \frac{\sqrt{\Delta U}}{\sqrt{\Delta f}}\right) \Delta f_y & \text{if } y > y^{EK}, \quad \text{(74a)} \\ -\left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_2}{f^2}}\right) f_y^2 + \left(\sqrt{\frac{\Delta U}{\Delta f}} - \sqrt{\frac{U_1}{f^1}}\right) f_y^1 & \text{if } y < y^{EK}. \quad \text{(74b)} \end{cases}$$

To establish the sign of (74a), note that $\Delta f_y = f_y^2 - f_y^1 < 0$ when f is strictly submodular; hence, $\Delta\pi'(y) < 0$ for $y > y^{EK}$. To establish the sign of (74b), note that $f^2/f^1 < U_2/U_1$ is equivalent to $\Delta U/\Delta f > U_1/f^1$ or $\Delta U/\Delta f > U_2/f^2$. The coefficient of f_y^2 in (74b) is therefore negative. Since f is submodular, $f_y^2 \leq f_y^1$, and we have

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

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

This result implies that firms with type y^{EK} have the strongest incentive to screen. If all firms choose $\sigma = 1$ in equilibrium, then sufficiency follows from Proposition 4; if all firms choose $\sigma = 0$ in equilibrium, then sufficiency follows from Proposition 8 or Eeckhout and Kircher (2010a). In the remaining case, where the equilibrium features both firms choosing $\sigma = 1$ and firms choosing $\sigma = 0$, we must have $\Delta\pi(y^{EK}) > c$ (otherwise all firms will choose $\sigma = 0$). There exist then two firm types $\underline{y}^s$ and $\overline{y}^s$ with $y' \leq \underline{y}^s < y^{EK} < \overline{y}^s \leq y''$, where firms of type $\underline{y}^s$ and $\overline{y}^s$ are indifferent between choosing $\sigma = 0$ and 1, i.e. $\Delta\pi(\underline{y}^s) = \Delta\pi(\overline{y}^s) = c$. Firms with $y < \underline{y}^s$ will choose $\sigma = 0$ and attract only high-type workers; firms with $y \in (\underline{y}^s, \overline{y}^s)$ will choose $\sigma = 1$ and attract both types of workers; finally, firms with $y > \overline{y}^s$ will choose $\sigma = 0$ and attract only low-type workers. Since all firm types y between $\underline{y}^s$ and $\overline{y}^s$ choose $\sigma = 1$, submodularity implies that NAC/NAM holds within this interval. Combining the above results implies that NAC/NAM holds globally.

Note that we can not weaken the requirement of strict submodularity to mere submodularity for the sufficient condition. To see this, set $f(x, y) = x + y$ and initially set c large enough so that all firms choose $\sigma = 0$. Then for $y \geq y^{EK}$, $\Delta\pi(y)$ is a constant by equation (73a). If we set $c = \Delta\pi(y^{EK})$, all firms with $y \geq y^{EK}$ are indifferent between choosing $\sigma = 0$ with low-type applicants and $\sigma = 1$ with both types of applicants. This indeterminacy violates NAC/NAM.

B.13.3 The Analysis of PAC/PAM

First, with a slight abuse of notation, given x_1 and x_2, we define $\rho(x_1, x_2, y)$ as the solution to

$$\frac{f_y(x_2, y)}{f_y(x_1, y)} = \left(\frac{f(x_2, y)}{f(x_1, y)}\right)^{\rho(x_1, x_2, y)}. \tag{75}$$

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

We now provide a claim which is stronger than the statements in Proposition 7.

Claim. *Consider a log-supermodular function f. Given an endowment of agents and a screening cost c, PAC/PAM holds in equilibrium as long as, for each y,*

$$\rho(x_1, x_2, y) \geq \Omega(\kappa(y)). \tag{76}$$

In contrast, given x_1, x_2 and $J(y)$, if for some $y^* \in (\underline{y}, \overline{y})$, we have

$$\rho(x_1, x_2, y^*) < \Omega(\kappa(y^*)), \tag{77}$$

then we can find (ℓ_1, ℓ_2) and c such that PAC/PAM fails in equilibrium.

Since $\Omega(\cdot)$ is strictly decreasing and with log-supermodular f, $\kappa(y)$ is increasing in y), the right-hand side of (76) reaches its maximum at $y = \underline{y}$. Also since $\rho(x_1, x_2, y) \geq \underline{\rho}$, the sufficient condition (42) in Proposition 7 then implies (76). On the other hand, given any log-supermodular function, whenever $x_1, x_2 \to x$, then $\kappa(y) \to 0$ and $\Omega(\kappa(y)) \to \infty$, and (77) holds for all $y^* \in [\underline{y}, \overline{y}]$, which, by the above claim, implies that we can find (ℓ_1, ℓ_2) and c such that PAC/PAM fails in equilibrium.

Note that for a CES production function, (76) reduces to $\rho \geq \Omega(\kappa(\underline{y}))$ and and (77) reduces to $\rho < \Omega(\kappa(\underline{y}))$. Thus, although the sufficient condition (42) is slightly weaker than (76), it is still sharp in the special case of CES production functions.

Before we move to the detailed proof, we first give a brief sketch. Under the sufficient condition (76), $\Delta\pi(y)$ is single-peaked at $y = y^{EK}$, so PAC/PAM follows from the same logic that was used for the case of NAC/NAM. In contrast, if (77) holds, then we can find (ℓ_1, ℓ_2) and a large c such that all firms choose $\sigma = 0$ in equilibrium, and $\Delta\pi(y)$ reaches its maximum at some point $\widetilde{y} > y^{EK}$ (note that the maximum is between 0 and c here). Now decrease c gradually till firms near $\widetilde{y}$ find it optimal to choose $\sigma = 1$ and screen ex-post while firms with types slightly above y^{EK} will continue choosing $\sigma = 0$ and accordingly attract high-type applicants only. PAC/PAM then fails in this case. Below, we prove this claim formally.

Similar to the analysis of NAC/NAM, since $f(x, y)$ is log-supermodular, and therefore strictly square-root supermodular, there exists a unique y^{EK} which solves (72). The first part of Lemma 15 states that y^{EK} must belong to the region characterized by (71). Furthermore, $f^2 - f^1$ is strictly increasing so that there exists at most one $y' < y^{EK}$ such that $f^2 - f^1 = U_2 - U_1$ (otherwise set $y' = \underline{y}$). Since we only assume weak log-supermodularity, f^2/f^1 is weakly increasing. Set $y'' = \min\{y \mid f^2/f^1 \geq U_2/U_1\}$ (if this set is empty, then set $y'' = \overline{y}$). The region characterized by (71) is then $y \in (y', y'')$. The following Lemma establishes that under the sufficient condition (76), $\Delta\pi(y)$ is single-peaked at $y = y^{EK}$.

Lemma 17. *Suppose that $f(x, y)$ is log-supermodular. In the region characterized by* (71)*, $\Delta\pi(y)$ is strictly increasing in y for $y \leq y^{EK}$, and if condition* (76) *holds for each $y \in (\underline{y}, \overline{y})$, then it is strictly decreasing in y for $y \geq y^{EK}$.*

Proof. If $y \in (y', y^{EK}]$, then $\Delta\pi(y)$ is given by (73a) and its derivative is given by (74a), so it is strictly increasing in y since $\Delta f_y > 0$. If $y \in [y^{EK}, y'')$, then $\Delta\pi(y)$ is given by (73b) and its derivative is now given by (74b) and can be rewritten as

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

where, to simplify notation, we shorten $\rho(x_1, x_2, y)$ as $\rho(y)$, and we used the identities $f^2/f^1 = 1 + \kappa(y)$ and $f_y^2/f_y^1 = (1+\kappa(y))^{\rho(y)}$.

Furthermore, define

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

which implies $\sqrt{U_2/\Delta U} = \sqrt{(\kappa(y)+\delta(y)^2)/\kappa(y)}$, and $\Delta\pi'(y)$ can be rewritten as

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

Thus, $\Delta\pi'(y)$ has the same sign as the numerator of the last factor in the last line. Single out the numerator and define

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

which is a quadratic function of δ with a strictly positive second-order coefficient since we assume $\rho \geq 1$ (log-supermodularity). Note that $\mathcal{S}(1,\kappa,\rho)=0$ and $\frac{\partial\mathcal{S}(\delta,\kappa,\rho)}{\partial\delta}\big|_{\delta=1} = 2(1+\kappa)^\rho((1+\kappa)^{\rho-1}-1)\geq 0$. Therefore, if $\mathcal{S}(0,\kappa,\rho)\leq 0$, then $\mathcal{S}(\delta,\kappa,\rho)<0$ for all $\delta\in(0,1)$. Note that $\mathcal{S}(0,\kappa,\rho)=\kappa\,(1+\kappa)^{2\rho-1}-((1+\kappa)^\rho-1)^2$, Thus $\mathcal{S}(0,\kappa,\rho)\leq 0$ if and only if $\sqrt{\frac{\kappa}{1+\kappa}}(1+\kappa)^\rho \leq (1+\kappa)^\rho-1$, or equivalently $\rho\geq\Omega(\kappa)$.

If for each $y\in(\underline{y},\overline{y})$, we have $\rho(y)\geq\Omega(\kappa(y))$, then by the above argument, $\mathcal{S}(\delta(y),\kappa(y),\rho(y))<0$ and hence $\Delta\pi'(y)<0$ for $y\in[y^{EK},y'')$. □

Similar to the case of NAC/NAM, we only need to consider the case where the equilibrium features both firms choosing $\sigma=1$ and firms choosing $\sigma=0$. Then there exist two firm types $\underline{y}^s$ and $\overline{y}^s$ that are indifferent between choosing $\sigma=0$ and 1, where $y'\leq\underline{y}^s<y^{EK}<\overline{y}^s\leq y''$. Firms with $y<\underline{y}^s$ will choose $\sigma=0$ and attract only low-type workers; firms with $y\in(\underline{y}^s,\overline{y}^s)$ will choose $\sigma=1$ and attract both types of workers; finally, firms with $y>\overline{y}^s$ will choose $\sigma=0$ and attract only high-type workers. Since all firms of y between $\underline{y}^s$ and $\overline{y}^s$ choose $\sigma=1$, log-supermodularity implies that PAC/PAM holds within this interval. Combining the above results then implies that PAC/PAM holds globally.

For the second part of the claim, we first prove the following. Given a log-supermodular function $f(x,y)$ and an endowment of agents, a necessary condition for PAC/PAM to hold for all c is that $\Delta\pi'_+(y^{EK})\leq 0$ when c is sufficiently large (for example, $c\geq f(x_2,\overline{y})$) so that all firms choose $\sigma=0$, where $\Delta\pi'_+(y^{EK})$ is the right derivative of $\Delta\pi(y)$ at point y^{EK}.

Suppose otherwise that $\Delta\pi'_+(y^{EK})$ is strictly positive; the maximum value of $\Delta\pi(y)$ must then be reached at some point $\widetilde{y} > y^{EK}$, since $\Delta\pi(y)$ is always strictly increasing when $y \in (y', y^{EK})$ (see Lemma 17). Now define $\widetilde{c} = \Delta\pi(\widetilde{y})$ and gradually decrease it from $f(x_2, \overline{y})$ to values around $\widetilde{c}$. What is the impact of this change on the sorting pattern? As long as $c \geq \widetilde{c}$, no firm is willing to invest in screening, so the equilibrium allocation remains the same. When c is slightly below $\widetilde{c}$, then firms with types sufficiently close to $\widetilde{y}$ will choose $\sigma = 1$. Note that the equilibrium market utilities U_1 and U_2 will change slightly, so that y^{EK} also changes only slightly. As before, firms with types slightly above y^{EK} will therefore choose $\sigma = 0$ and hire high-type workers only, while firms with types sufficiently close to $\widetilde{y}$ will attract both types of workers. Hence, PAC/PAM fails to hold when c is slightly below $\widetilde{c}$.

Below, we complete the proof by showing that for any log-supermodular function $f(x, y)$ and $(x_1,\ x_2,\ J(y))$, if (77) holds for some $y^* \in (\underline{y}, \overline{y})$, then we can choose (ℓ_1, ℓ_2) such that $\Delta\pi'_+(y^{EK}) > 0$ when c is sufficiently large that all firms choose $\sigma = 0$. The idea of construction is similar to the counterexample for the necessity part of Proposition 8 (the bilateral case).

Step 1: Since $\rho(y^*) < \Omega(\kappa(y^*))$, we have $\mathcal{S}\left(0, \kappa(y^*), \rho(y^*)\right) > 0$, where $\mathcal{S}$ is defined in equation (79). Thus, by continuity, we can find a δ^* small enough such that $\mathcal{S}(\delta^*, \kappa(y^*), \rho(y^*)) > 0$. Next, we construct (U_1, U_2) from the following two equations,

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

These equations are reminiscent of (72) and (78), respectively. The main difference is that there we considered the market utilities as known and solved for y^{EK} and $\delta(y)$; here we treat y^* and δ^* as known and solve for market utilities instead. Denote the unique solution by (U_1^*, U_2^*).

Step 2: Given (U_1^*, U_2^*), y^* is then the firm type that corresponds to y^{EK} defined before. Since f is log-supermodular and hence strictly square-root supermodular, firms with types $y > y^*$ will attract only high-type applicants, and firms with types $y < y^*$ will attract only low-type applicants. The firms' problem is $\max_\lambda m(\lambda)f(x_1, y) - \lambda U_1^*$ for $y \leq y^*$, and $\max_\lambda m(\lambda)f(x_2, y) - \lambda U_2^*$ for $y \geq y^*$.

Denote the solution by $\lambda(y)$ for all y.

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