Sorting in the Labor Market*

Jan Eeckhout[†]

October 9, 2017

Abstract

This paper surveys the literature on sorting in the labor market. There are inherent differences in worker ability and across firm productivity. A fundamental question is whether the exact composition of skills of workers and productivity of firms affects output, and how it determines the equilibrium allocation of workers within a firm and between firms. There has been a surge of research investigating the causes and consequences of the allocation process of heterogeneous workers to firms. The focus here is on the theory that sheds light on open questions in macroeconomics, labor and industrial organization, with a particular emphasis on the role of firm size. Those models allow us to infer from the observed sorting patterns (who matches with whom) what the underlying technological determinants are, and how they have evolved in recent decades. And they help us understand the technological origins of important labor market trends such as the increase in wage inequality and the change in labor market and firm dynamics.

Keywords. Labor Markets. Sorting. Wages. Inequality. Firm Size Distribution. Assortative Matching. Complementarities. Supermodularity. Unemployment.

*I benefitted from the detailed comments by the editor, Gene Grossman, and from discussion with and feedback from numerous colleagues. I gratefully acknowledge the support from the ERC, Advanced Grant 339186, and from ECO2015-67655-P.

[†]Department of Economics, UPF (Barcelona GSE/ICREA) and University College London, `jan.eeckhout@upf.edu`.

Contents

1 Introduction

Firms spend a lot of time and resources finding the right (wo)man for the job. This is true for high skilled personnel such as managers and professionals, but also for low skilled workers. The average total cost of hiring a worker is estimated at multiple monthly salaries. Given that the average duration of a job is only a few years, this is a large share of the cost of employment. Such a large expense indicates that there are big benefits from picking the right worker and equally big costs from hiring the wrong one. In large part, those benefits emanate from the impact that workers have on one another and how their productivity is affected by the jobs they are in. Work is to a large extent a team effort, and we know from sports teams that a star player is only as good as the team around her. Talent thrives thanks to the talent and productivity of others.

This immediately begs the question how talented workers are assigned when there are such peer effects across different types of workers and how they should be assigned. This is a problem of *sorting*, or assortative matching. Is there (should there be) Positive Assortative Matching (PAM) where the most productive workers are assigned to work in teams with other high productivity workers, and the least productive workers in teams amongst other low productivity workers? Or is there Negative Assortative Matching (NAM) with diversity of workers within teams where high productivity workers are matched with low productivity workers? The answer to these questions depends on the relative effect that more productive workers have. If they affect the high productivity workers *more* than the low productivity workers, then PAM will be optimal, and NAM if they affect them less. This relative effect on other workers' productivity is measured by the extent to which worker inputs are complementary or substitutable in production of output. In mathematical terms, this is equivalently measured by whether the match surplus function is supermodular or submodular.[1]

Understanding the sorting patterns – who matches with whom – allows us to better understand the underlying technology that generates the surplus of such a match. The way workers sort into firms has changed dramatically in recent decades. Wage inequality has gone up substantially, but most of that increase is between firms, not within. Increasingly workers within firms look more similar to one another than workers do between firms. A few decades ago, every firm had both high paid professionals and low paid janitors and cleaners. Now many firms hire the services of a security firm and a cleaning company. As a result, a law firm has mainly high skilled lawyers, and a cleaning company has mainly low skilled cleaning personnel. There is ample evidence in support of this fact – that the increase in wage inequality is mainly driven by an increase in between-firm inequality and virtually no increase in within-firm inequality – for different countries (for references, see section 2.3.3).[2] The evolution of wage inequality and how it varies between and within firms indicates that the sorting pattern is changing, which is the result of a changing technology. While we cannot observe the underlying technology,

[1] Evidence suggests that this is not only the case for top sports professionals. In a study of peer effects amongst randomly assigned check out cashiers, Mas and Moretti (2009) find that high productivity workers positively affect the productivity mostly of less productive workers. As a result, in that case the optimal assignment is NAM where teams are diverse.

[2] No doubt, some of this outsourcing is motivated by the firms' responses to regulation (such as the requirement to provide health and other benefits in the United States). However, the fact that this is so pervasive in most advanced economies with very different regulation indicates that there is something more primitive beyond regulation that is underlying this development.

the change in the sorting pattern and how workers are allocated across firms will give us an insight into the underlying technological change. Sorting is therefore important for our understanding of the determinants of wage inequality.

In this article, I survey the theoretical literature on sorting in the labor market. The building block of this review is the sorting model without matching frictions. More precisely, the goal is to consider economic problems of sorting where a nontrivial matching problem of heterogeneous agents takes place, both when matching one-to-one (one firm to one worker, or worker-to-worker in teams of two) and when matching with large firms (where many workers match to one firm). I will also consider the role of search frictions to analyze sorting in environments with employment and/or mismatch. This review is unique in its focus on the labor market with an emphasis on large firms and mismatch. The broader literatures on assortative matching and search have been reviewed elsewhere (see Smith (2011), Rogerson, Shimer, and Wright (2005), Chade, Eeckhout, and Smith (2017) and Roth and Sotomayor (1990)), and this review is more specialized than and therefore complementary to those reviews.

The terminology for sorting in economics has its origin in the biology literature on assortative *mating*.[3] When sexually reproducing organisms choose to mate with partners that are similar in phenotype – i.e. any discernible physical characteristic – there is positive assortative mating. When the sorting is on inheritable traits, positive assortative mating tends to decrease the trait variation within the population of a species, but it renders the species more specialized. This allows the breed to excel in a narrowly defined attribute. As Darwin pointed out with his dog breeding example, humans can manipulate natural environmental circumstances that lead to specialization of the breed: fast hounds, elegant retrievers and intelligent shepherds. Negative assortative mating instead increases the trait variation. A mutt is neither elegant nor fast, but a generalist and adaptable to different environments. Those general and broad traits endow the species with a wide range of characteristics that allow it to survive in very different environments.[4] Negative sorting is therefore considered a stabilizing force for the species precisely because it provides the whole species with versatility and adaptability to a changing environment.

The theoretical literature on frictionless matching dates back to the early work by Kantorovich (1948) on the mass transportation problem, a problem initially posed by Monge (1781), and to Koopmans and Beckmann (1957) who introduce a pricing system to solve this problem. The work by Becker (1973) on the marriage market has focussed the spot light on the conditions for sorting, who matches with whom. Becker characterizes the relation between the properties of the match value function (supermodularity) and the resulting equilibrium allocation (positive or negative assortative matching).

By now, there has been enormous progress in the literature on assortative matching, both without frictions and in the presence of search. The renewed interest is fueled by the new theoretical challenges as well as by the relevance for applications in the labor, housing, macro, industrial organization, trade,... The main idea here is to offer a unified framework of the basic models. The focus is on aspects of the

[3]It is not clear when and where the adoption of the biological term in economics happened – some claim it was Becker (1973) when he studied the marriage market – but it is clear that economists, willingly or not, mostly talk about matching and not mating.

[4]When the Sputnik engineers were choosing the canine Laika to be sent into orbit, they picked a young adult stray dog from the streets of Moscow whom they anticipated would best be able to survive long periods in adverse environments of cold, heat and hunger.

theoretical sorting literature that are of relevance for applications, such as firm size, risk aversion, and mismatch.

In Section 2, I start with the benchmark model without frictions, and then introduce non-linear preferences. Then I introduce firm size, first exogenously, and then endogenously chosen. Firm size is an important aspect of the macro and industrial organization literatures and this section thus links it also to sorting in the labor market. Finally, we consider the importance of negative assortative matching, in particular in the context of information aggregation and of matching with externalities. In Section 3, I introduce unemployment in the presence of sorting. The focus is on mismatch and large firms. I also analyze the role of risk aversion for assortative matching in the presence of search frictions. In Section 4, I analyze three sources of mismatch: search friction, stochastic sorting and multidimensional types and discuss how this helps identifying sorting. I conclude in Section 5.

2 Sorting, Wages, and Firms

In this section I first lay out the benchmark model that has one-to-one matching of workers to firms, both with linear preference (Transferable Utility, TU) and with non-linear preferences (Imperfectly Transferable Utility, ITU), and derive the equilibrium conditions for sorting and the equilibrium allocation and wages. Then I focus attention on sorting with large firms, which is particularly relevant for applications in macro, labor and industrial organization. Finally, I return to matching problems where Negative Assortative Matching (NAM) is natural, namely when there is information aggregation or when there are externalities across matches.

2.1 The Basic Model

We start with an illustrative example.

Example 1 *Consider a set of workers $\mathcal{X} = \{1, 2, 3\}$ and a set of firms $\mathcal{Y} = \{1, 2, 3\}$. Let the match surplus function be $f(x, y) = x \cdot y$. The match surplus can be written as*

$$f(x,y) = \begin{pmatrix} 9 & 6 & 3 \\ 6 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix} \tag{1}$$

We are looking for an allocation μ that indicates which worker is matched with which firm: $y = \mu(x)$ is a function that maps worker types x into firm types y. Utility is transferable, and an allocation will induce payoffs, wages $w(x)$ and profits $\pi(y)$. Shapley and Shubik (1972) consider the equilibrium concept of stability for this assignment game. An outcome (w, π, μ) is stable if there is no blocking pair, i.e., $w(x) + \pi(y) \geq f(x, y), \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$. This implies a system of 9 inequalities, where along the equilibrium allocation it holds with equality $w(x) + \pi(\mu(x)) = f(x, \mu(x))$.

Shapley and Shubik (1972) show that the set of stable outcomes exists and coincides with the core and with the solution of the dual of the linear programming problem of the assignment game. In this example, the solution satisfies $\mu(1, 2, 3) = (1, 2, 3)$ and $1 \leq w_2 - w_1 \leq 2, 2 \leq w_3 - w_1 \leq 6, 2 \leq w_3 - w_2 \leq 3$. There is a unique allocation that exhibits positive assortative matching. The high type $x = 3$ matches with the high type $y = 3$. Even though the allocation is unique, there are multiple wages that are consistent with the stability of this allocation. Different equilibrium prices do not affect the allocative efficiency, but they do impact the distribution of the rents. One particular wage vector is $(0.5, 2, 4.5)$ which is also equal to the profit vector given symmetry. It is easily verified that this outcome (allocation and payoffs) is stable. For example, if $x = 1$ and $y = 2$ wanted to deviate to form a match, their joint surplus $f(1, 2) = 2$ is no larger than the sum of their current payoffs $w(1) + \pi(2) = 2.5$.

Becker (1973) establishes that there is positive assortative matching if the match surplus function $f(x, y)$ is supermodular, or equivalently, if f exhibits complementarities. A function is supermodular if for any $\overline{x} > \underline{x}$ and $\overline{y} > \underline{y}$, it is the case that

$$f\left(\overline{x}, \overline{y}\right) + f\left(\underline{x}, \underline{y}\right) > f\left(\overline{x}, \underline{y}\right) + f\left(\underline{x}, \overline{y}\right). \tag{2}$$

When $f(x, y)$ is differentiable in both arguments, supermodularity is equivalent to a positive cross-partial derivative $f_{xy} > 0$. In our example, f is supermodular, $f_{xy} = 1$.

Shapley and Shubik (1972) also establish that the stable allocation (w, π, μ) is also optimal in the sense that it maximizes the sum of total payoffs, i.e., there is no other allocation that generates a higher sum of total payoffs. The proof follows immediately from the fact that $\sum_{x\in\mathcal{X}} [w(x) + \pi(\mu(x))] = \sum_{x\in\mathcal{X}} f(x, \mu(x))$ and stability: $w(x) + \pi(y) \geq f(x, y)$, for all (x, y).

This is an example of what is known in the mathematics literature as the optimal transport problem, or transportation theory. The problem was originally posed by Monge (1781), and later solved by Kantorovich (1948) and Koopmans and Beckmann (1957) as a linear programing problem. Interestingly, Kantorovich (1948), then working for the Soviet government as a planner optimizing the production in the plywood industry, solved this problem as a planner's problem. Instead, Koopmans and Beckmann (1957) – unaware of Kantorovich's solution because it was published in Russian – solved it as a decentralized equilibrium with prices. Unknown to them then, the planner's optimal solution and the equilibrium coincide. Following Koopmans and Beckmann (1957), this problem is also called an assignment problem.[5]

Against the backdrop of this example, we analyze the assignment problem in a large economy with a continuum of agents and differentiable distributions of types and match surplus functions. There are no restrictions on the allocation $\mu(x)$, but we will be looking for conditions on the environment that imply that $\mu(x)$ is monotonic and differentiable. The advantage is that the large matrix of inequalities is now substituted by a differential equation thus making the analysis much more tractable for applications.

Let the worker type x and firm type y[6] be distributed according to $\Gamma(x)$ and $\Psi(y)$ (with densities γ and ψ). As before, the match surplus is given by $f(x, y) \geq 0$. Often we will have examples where there is monotonicity in types, $f_x > 0$ and $f_y > 0$, i.e., higher ranked types contribute more to output than lower ranked types, but it is not essential. What is essential for the particular equilibrium allocation is whether or not there are complementarities in the match surplus function f between x and y, that is, whether $f(x, y)$ is supermodular.

[5]For a review of recent developments in the transportation literature (mostly in mathematics), see Villani (2009), and for a treatment of the transportation problem with an angle from the economics literature, see Galichon (2016).

[6]If matching is between teams of co-workers, then we can refer to x as worker 1 and y as worker 2.

An outcome is an assignment of workers to firms $\mu(x) = y$, a wage schedule $w(x)$ and a profit schedule $\pi(y)$. These wage and profit schedules are hedonic price schedules, that relate compensation to worker or firm attributes. Unlike standard price schedule that are linear in quantities, hedonic price schedules are non-linear in attributes. We analyze this problem as a competitive equilibrium where firms y choose the optimal worker x to match with, taking as given a wage schedule $w(x)$. For the assignment problem (two-sided matching with transferable utility), the stable matching, the core and the competitive equilibrium all coincide (Gretsky, Ostroy, and Zame (1992)). The equilibrium allocation also is optimal, i.e., it maximizes the total sum of the surplus (see Koopmans and Beckmann (1957) and Roth and Sotomayor (1990)).

The model is closed with a market clearing condition, basically ensuring that the matching μ is measure-preserving (i.e. the measure of workers matched is equal to the measure of firms). Market clearing then pins down the equilibrium wage schedule. We define positive assortative matching (PAM) as an allocation μ that is a strictly increasing function ($\mu'(x) > 0$) and negative assortative matching (NAM) as $\mu'(x) < 0$. Under PAM, the equilibrium allocation can be written as:[7]

$$\int_{x}^{\overline{x}} \gamma(x)dx = \int_{\mu(x)}^{\overline{y}} \psi(y)dy \iff \Gamma(x) = \Psi(\mu(x)) \iff \mu(x) = \Psi^{-1}(\Gamma(x)). \tag{4}$$

Starting at the top, the highest type $\overline{x}$ matches with the highest type $\overline{y}$ under PAM. Any lower type x has a measure $1 - \Gamma(x)$ above her, and likewise any type y has a measure $1 - \Psi(y)$ that are higher. The type x matches with the type $y = \mu(x)$ if the measure of agents above x is equal to the measure above y, i.e., the matching is measure preserving. This is the continuum version of the discrete case where the first ranked x matches with the first ranked y, the second x with the second y,..., and the 17th ranked x with the 17th ranked y.

Given a wage schedule $w(x)$, each firm y maximizes profits:

$$\pi(y) = \max_{\tilde{x}} f(\tilde{x}, y) - w(\tilde{x}), \tag{5}$$

with first order condition $f_x(x, y) - \frac{\partial w(x)}{\partial x} = 0$. The equilibrium wage schedule must also satisfy market clearing. We thus obtain the equilibrium wages schedule by integrating along the market clearing allocation $y = \mu(x)$:

$$w^{\star}(x) = \int_{0}^{x} f_x(\tilde{x}, \mu(\tilde{x}))d\tilde{x} \ + w_0, \tag{6}$$

where w_0 is the constant of integration which is zero if the measure of firms is smaller than that of workers, equal to $f(0, \mu(0))$ if the measure of firms is larger, or any value $w_0 \in [0, f(0, \mu(0))]$ if the measure of firms is equal.[8]

[7] Under NAM

$$\int_{x}^{\overline{x}} \gamma(x)dx = \int_{\underline{y}}^{\mu(x)} \psi(y)dy \iff 1 - \Gamma(x) = \Psi(\mu(x)) \iff \mu(x) = \Psi^{-1}(1 - \Gamma(x)). \tag{3}$$

[8] Equilibrium profits are given by the residual of output minus wages: $\pi^{\star}(y) = f(\mu^{-1}(y), y) - w^{\star}(\mu^{-1}(y))$. There is also a dual problem that can be solved and that yields the same solution where workers x maximize their payoff by choosing a firm y given a hedonic profit schedule $\pi(y)$.

We are interested in a monotonic equilibrium where $\mu'(x)$ is either positive or negative for all x. The properties of μ can be derived from analyzing the second order condition for a global maximum:

$$f_{xx}(x, y) - w_{xx}(x) < 0. \tag{7}$$

Totally differentiation with respect to x of the first order condition evaluated along the equilibrium allocation $y = \mu(x)$, yields the identity $f_{xx}(x, \mu(x)) + f_{xy}(x, \mu(x))\mu'(x) = w_{xx}(x)$. Using this identity, the second order condition (7) is therefore satisfied (provided $\mu(x)$ is differentiable) when

$$f_{xy}(x, \mu(x))\mu'(x) > 0. \tag{8}$$

From this condition it follows that there is positive assortative matching (PAM, i.e., $\mu'(x) > 0$) whenever f_{xy} is positive, i.e., $f(x, y)$ is supermodular. There is negative assortative matching (NAM, i.e., $\mu'(x) < 0$) whenever f_{xy} is negative, i.e., $f(x, y)$ is submodular. This condition is derived only locally along the equilibrium allocation, where it must always be satisfied since it would violate stability if not.

To see the intuition behind the relation between positive sorting and supermodularity, consider the graphical interpretation of supermodularity in Figure 1. A function $f(x, y)$ is supermodular if the sum of the its value at the extremes (red dots) exceeds that at the intermediates (blue squares): $f(\overline{x}, \overline{y}) + f(\underline{x}, \underline{y}) > f(\overline{x}, \underline{y}) + f(\underline{x}, \overline{y})$. Supermodularity can thus be thought of as a notion of convexity on a multidimensional domain: the function takes on higher values at the extremes than in the interior.

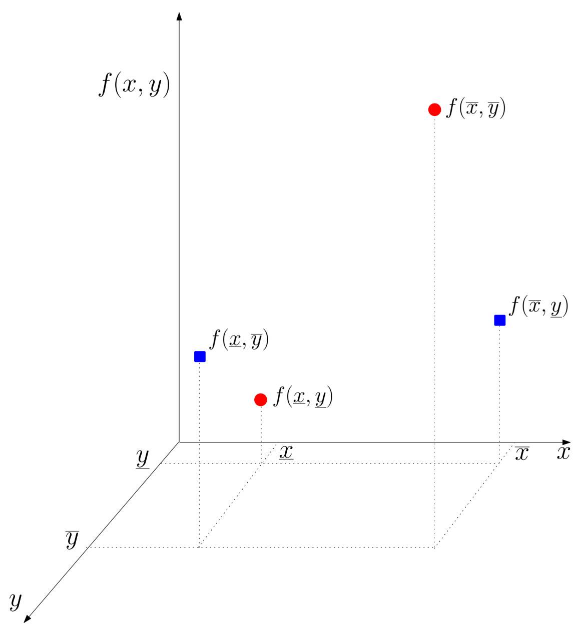

Figure 1: Supermodularity.

An alternative interpretation is in terms of *increasing differences.* The condition for supermodularity can also be written as $f(\overline{x}, \overline{y}) - f(\underline{x}, \overline{y}) > f(\overline{x}, \underline{y}) - f(\underline{x}, \underline{y})$. The increase in match value from moving from $\underline{x}$ to $\overline{x}$ is higher at $\overline{y}$ than at $\underline{y}$. The marginal contribution of a high type $\overline{x}$ is therefore higher when matched with a high type $\overline{y}$.

Outside the equilibrium allocation, supermodularity need not be satisfied for PAM. The second order condition is necessary, but not sufficient, see Eeckhout (2006) for examples. Therefore, if the surplus function for a given distribution of types is supermodular for all $x, y \in \mathcal{X} \times \mathcal{Y}$, this constitutes a sufficient condition for PAM, but it is not necessary. Once we allow for arbitrary distributions of types x and y, the allocation $y = \mu(x)$ can take any value in the domain $\mathcal{X} \times \mathcal{Y}$. Because supermodularity must necessarily be satisfied along the allocation, and the allocation can potentially cover the entire domain, supermodularity must necessarily be satisfied for all x, y. The condition is also sufficient provided it holds strictly. If it holds weakly, then any allocation is an equilibrium and therefore a PAM allocation does not necessarily imply that there is supermodularity.

2.2 Assortative Matching in the Presence of Risk Aversion and Moral Hazard

One of the major limitations for economic applications of the assignment game is that utility is transferable, i.e., the Pareto frontier is linear. While the model with Transferable Utility (TU) is very tractable and insightful, it rules out any economic context with non-linear utilities due to risk aversion for example. Many economic examples match agents that have non-linear preferences. In one of the early applications analyzing the allocation of farmers to landowners in early Renaissance Tuscany, Ackerberg and Botticini (2002) find that the degree of risk aversion of the tenants affects that sorting pattern of farmers to land characteristics. Consequently, Serfes (2005), Chiappori and Reny (2015), Schulhofer-Wohl (2006) and Legros and Newman (2007) have modeled matching in a setting with non-linear preferences.

Legros and Newman (2007) analyze the properties of the assignment when the pairwise frontiers are non-linear. They use as the primitive the Pareto frontier, denoted by $\phi(x, y, u)$. It is the payoff that a firm of type y obtains in a match with worker type x and where the worker x receives a utility u. This general case of Imperfectly Transferable Utility (ITU) has a frontier that is assumed to be concave and differentiable and with $\phi_u < 0$. At the two extremes of this formulation are TU and NTU (Non-Transferable Utility). In the case of TU, $\phi(x, y, u) = f(x, y) - u$. In the case of NTU (Gale and Shapley (1962)), preferences are ordinal and the Pareto frontier can be represented by a single point – at best, because there is really no cardinal utility – where agents take the ordinal utility as given without any transfers.

For discrete types, Legros and Newman (2007) derive a condition for matching that implies assortativeness of the allocation. They call it Generalized Increasing Differences (GID). Here we derive the necessary and sufficient condition for the continuous case:[9]

Result *In the presence of non-linear utilities, the necessary and sufficient condition for positive assortative matching is*

$$\phi_{xy} > \frac{\phi_y}{\phi_u} \phi_{yu}, \tag{9}$$

with the opposite inequality for negative assortative matching.

[9]For the continuous framework, Legros and Newman (2007) derive a sufficient condition for PAM, namely $\phi_{xy} \geq 0$ and $\phi_{yu} < 0$ ($\phi_{xy} > 0$ and $\phi_{yu} \leq 0$ for NAM.

As before with TU where given a wage schedule the firm maximizes profits, now the firm maximizes $\phi(x,y,u)$ where $u(x)$ is the utility that a worker of type x obtains from the match and that corresponds to the hedonic wage schedule above. Then the first-order condition satisfies:

$$\phi_x(x,y,u) + \phi_u(x,y,u)u'(x) = 0 \tag{10}$$

The equilibrium allocation μ that solves this first-order condition is a maximum provided it satisfies the second-order condition:

$$\phi_{xx} + 2\phi_{xu}u' + \phi_{uu}u'^2 + \phi_u u'' < 0. \tag{11}$$

Provided μ is differentiable, differentiating the first order condition along the candidate equilibrium allocation $y = \mu(x)$ implies $\phi_{xx} + \phi_{xy}\mu' + \phi_{xu}u' + \phi_{xu}u' + \phi_{yu}\mu' u' + \phi_{uu}u'^2 + \phi_u u'' = 0$. Therefore, the second-order condition is satisfied provided $\phi_{xy}\mu' + \phi_{yu}\mu' u' > 0$ and using the first order condition to substitute for u' we obtain:

$$\mu' \left[\phi_{xy} - \frac{\phi_x}{\phi_u}\phi_{yu}\right] > 0. \tag{12}$$

There is positive assortative matching, i.e., $\mu' > 0$ provided

$$\phi_{xy} > \frac{\phi_x}{\phi_u}\phi_{yu}. \tag{13}$$

This condition highlights that mere supermodularity between types is no longer the exclusive determinant of positive sorting. How the surplus of the match varies, in particular, how the Pareto frontier varies across different matched pairs through ϕ_{yu} now also affects the sorting pattern. Consider for instance the case where the frontier is steeper the larger the firm type y, i.e., ϕ_{yu} is negative. The same worker type can transfer more utils to the high type firm than to the low type firm in exchange for one of her utils. This would be the case for example if the high type firm's technology is more productive for the same amount of worker effort. That implies there can be negative assortative matching even if there is supermodularity in x and y, provided ϕ_{xy} is not too positive. To see this, recall that $\phi_u < 0$, and therefore the right hand side of the equation is positive. The slope of the Pareto frontier thus captures the extent to which types y and utilities to the worker u can be traded off. If higher y do disproportionately better, less of complementarity is needed to nonetheless obtain positive sorting.

This is evident when considering the case where agents on each side of the market have common, non-linear preferences over consumption. Consider the standard assignment game where output produced by a pair (x,y) is transferable and is denoted by $f(x,y)$. Denote by $U(\cdot)$ the utility obtained by any worker type x over consumption and likewise for any firm's preferences $V(\cdot)$ over profits. This means that there are common preferences for all worker, and common preferences for all firms, but not that workers have the same preferences as firms: $U \neq V$. We assume that both U and V are increasing, concave, differentiable and with bounded derivatives. Then given a wage schedule $w(x)$ and the fact that $u = U(w(x))$, we can write $\phi(x,y,u) = V(f(x,y) - w(x)) = V(f(x,y) - U^{-1}(u))$. It is immediately verified that $\phi_{xy} = V'' \cdot f_x f_y + V' f_{xy}$, $\frac{\phi_x}{\phi_u} = \frac{V' \cdot f_x}{-V' 1/U'} = -U' \cdot f_x$ and $\phi_{yu} = -\frac{V'' \cdot f_y}{U'}$. Now, condition (9) is equivalent to:

$$V'' \cdot f_x f_y + V' f_{xy} > \left(-U' \cdot f_x\right)\left(-\frac{V'' \cdot f_y}{U'}\right) \quad \Longleftrightarrow \quad V' f_{xy} > 0 \tag{14}$$

Therefore, since $V' > 0$, the condition for positive assortative matching reduces to the standard condition of supermodularity in the assignment game with TU. In other words, the result from TU extends to the case where agents have common preferences on each side of the market.

Result *Consider the model with non-linear utilities. Under common preferences for workers and firms, the condition for Positive Assortative Matching is supermodularity:* $f_{xy} > 0$.

This result is interesting because we can analyze a wide class of relevant economic settings that involve risk preferences and contract design. Here, with common preferences on each side of the market, there is no difference in the preferences of different agents, but agents can be exposed to different consumption levels. As we will show below, we can introduce risk in output or in endowments and analyze the role of risk on the matching pattern. For example, this allows us to study the role of risk in the matching of CEO's to firms, where different firms face different stochastic outcomes, and likewise for CEO's. Or in the application in Ackerberg and Botticini (2002), some own plots of land that exhibit properties that are substantially different across different farmers: one may be wetland that generates a steady return from cattle grazing, whereas another plot may harvest fruits and vegetables that are very sensitive to drops in temperature and therefore yield a very volatile return.

Note that this result is consistent with Costinot (2009) and Costinot and Vogel (2015) who show that output produced must be log-supermodular to guarantee PAM. In their model, all agents have common preferences, but output produced by each firm is not perfectly substitutable as it is here. There is a CES aggregator over consumption, and it is the less than perfect substitutability that requires a stronger condition than mere supermodularity.

2.3 Assortative Matching in Macro: Large Firms

In the quintessential model of the macro economy, firms make choices about the size of the labor force or of capital investment. These are often referred to as *intensive* margin decisions, the firm makes a quantity decision. Instead, in the matching models we have discussed so far, the decision is a quality choice: which worker to hire, or whether or not to enter in the labor market. This is often referred to as an *extensive* margin decision. The intensive margin decision has been modeled to explain the firm size distribution. Lucas (1978) models the firm decision to hire a number of people as the management's span of control. The more productive management hires a larger work force thus spanning its control further.

In reality, management at a firm faces a more complex trade-off. Not only does it decide on the number of people to hire, it also decides on which worker types to hire. Now this extensive margin decision is precisely the canonical feature of the matching models. The problem is that in the canonical model of the labor market (Becker (1973)) a firm consists of exactly one job. A notable exception is Kelso and Crawford (1982) who consider a many-to-one matching model of the firm and show sufficient conditions for existence, namely the presence of gross substitutes.

In this section, we review models with heterogeneous firms and workers, and where each firm consists of multiple workers. We start with large firms that have an exogenously given size. Then we endogenize the size as a choice of the firm, with two different specifications that lead to very different outcomes.

2.3.1 Fixed Firm Size

An immediate extension of the one-to-one sorting model with TU that we discussed above is a one-to-N model, as for example in the O-ring theory of Kremer (1993). In his model, there is a measure 1 of firms indexed by y with distribution $\Psi(y)$ and a measure N of workers x with CDF $\Gamma(x)$. A firm of type y produces output with N inputs $x_1, ..., x_N$, potentially different, where $F(x_1, ..., x_N; y) = (\Pi_i x_i)\, y$. As Kremer (1993) shows, in equilibrium, the firm chooses identical inputs $x_i = x$ and the equilibrium allocation $\mu(x)$ satisfies $\Gamma(x) = \Psi(\mu(x))$. Since the measure of workers is N, each firm $y = \mu(x)$ hires exactly N type x workers. In equilibrium, output is given by $f(x, ..., x; \mu(x)) = x^N \mu(x)$.

A slight variant of this model has N pools of different worker types $x_i, i = 1, ..., N$, each distributed according to the CDF $\Gamma_i(x_i)$, and firms y with CDF $\Psi(y)$. Think of each worker pool consisting of a skill or education category, and those workers can only perform jobs in their category. Output can then be written as $f(x_1, ..., x_N; y)$. Now there is a hedonic wage schedule $w_i(x_i)$ and an equilibrium allocation $\mu_i(x_i)$ satisfying market clearing for each skill group i. Under PAM for example, the equilibrium allocation for each skill group satisfies $\Gamma_i(x_i) = \Psi(\mu_i(x_i))$ for all I.

For expositional simplicity, consider the case where $N = 2$. The first order conditions for each x_i satisfy:

$$f_{x_1} = w_1'(x_1) \tag{15}$$
$$f_{x_2} = w_2'(x_2). \tag{16}$$

As before, we will evaluate the second order condition to derive the conditions on the technology that lead to positive or negative sorting. Because the firm now chooses two skill types, we have a multi-dimensional second order condition. To that end, we write the Hessian as

$$\mathbf{H} = \begin{pmatrix} f_{x_1x_1} - w_1''(x_1) & f_{x_1x_2} \\ f_{x_1x_2} & f_{x_2x_2} - w_2''(x_2) \end{pmatrix}. \tag{17}$$

Now we evaluate the Hessian at the equilibrium allocation in order to substitute for the wages w_i''. To do so, we know that, at equilibrium, a change in say x_1 affects the allocation $y = \mu_1(x_1)$ but also the allocation of x_2, since $y = \mu_2(x_2)$ and therefore $x_2 = \mu_2^{-1}[\mu_1(x_1)]$. Hence the two FOCs evaluated at the equilibrium allocation can be written as:

$$f_{x_1}\left(x_1, \mu_2^{-1}[\mu_1(x_1)], \mu_1(x_1)\right) - w_1'(x_1) = 0 \tag{18}$$
$$f_{x_2}\left(\mu_1^{-1}[\mu_2(x_2)], x_2, \mu_2(x_2)\right) - w_2'(x_2) = 0. \tag{19}$$

Taking the total derivative with respect to x_i yields:

$$f_{x_1x_1} + f_{x_1x_2}[\mu_2^{-1}]'\mu_1' + f_{x_1y}\mu_1' - w_1'' = 0 \tag{20}$$
$$f_{x_2x_2} + f_{x_1x_2}[\mu_1^{-1}]'\mu_2' + f_{x_2y}\mu_2' - w_2'' = 0. \tag{21}$$

We can now verify the three conditions for negative definiteness of the Hessian:

$$f_{x_1x_2}[\mu_2^{-1}]'\mu_1' + f_{x_1y}\mu_1' > 0 \quad \text{and} \quad f_{x_1x_2}[\mu_1^{-1}]'\mu_2' + f_{x_2y}\mu_2' > 0 \tag{22}$$
$$\left(f_{x_1x_2}[\mu_2^{-1}]'\mu_1' + f_{x_1y}\mu_1'\right)\left(f_{x_1x_2}[\mu_1^{-1}]'\mu_2' + f_{x_2y}\mu_2'\right) - (f_{x_1x_2})^2 > 0 \tag{23}$$

where the last line follows from setting the determinant $|\mathbf{H}| > 0$ and the first two inequalities from requiring that the principal minors are negative.

To check consistency and gain some intuition, consider the case where $f_{x_1x_2} = 0$, i.e., f is additively separable in x_1 and x_2. Then the three conditions are:

$$f_{x_1y}\mu_1' > 0 \quad \text{and} \quad f_{x_2y}\mu_2' > 0 \quad \text{and} \quad \left(f_{x_1y}\mu_1'\right)\left(f_{x_2y}\mu_2'\right) > 0 \tag{24}$$

which is satisfied by supermodularity of f in x_1, y and in x_2, y. Now consider another case, where distributions of x_1, x_2, y are identical. Hence under PAM, $\mu_1 = x_1$ and $\mu_2 = x_2$ and the derivatives are all 1, also of the inverse. Then the three conditions can be written as

$$f_{x_1x_2} + f_{x_1y} > 0 \quad \text{and} \quad f_{x_1x_2} + f_{x_2y} > 0 \quad \text{and} \quad f_{x_1x_2}\left(f_{x_1y} + f_{x_2y}\right) + f_{x_1y}f_{x_2y} > 0. \tag{25}$$

Now a sufficient condition for PAM is that f is supermodular in all three arguments:

$$f_{x_1x_2} > 0, \quad f_{x_1y} > 0, \quad f_{x_2y} > 0. \tag{26}$$

We now show for this extended Becker (1973) model that if $f(x_1, x_2, y)$ is supermodular (in all arguments), then there is PAM in both skills for any distribution (not just symmetry). Notice that for the general model assuming supermodularity of f implies that the two equations in (22) are automatically satisfied under PAM. The question is what happens to the third equation (23), which can be written as:

$$f_{x_1x_2}\left[f_{x_1x_2}\left([\mu_2^{-1}]'\mu_1'[\mu_1^{-1}]'\mu_2' - 1\right) + f_{x_1y}\mu_1'[\mu_1^{-1}]'\mu_2' + f_{x_2y}\mu_2'[\mu_2^{-1}]'\mu_1'\right] + f_{x_1y}f_{x_2y}\mu_1'\mu_2' > 0.$$

Observe that the derivative of the inverse function is the inverse of the derivative so that $\mu_1'[\mu_1^{-1}]' = 1$ and we obtain:

$$f_{x_1x_2}\left[f_{x_1y}\mu_2' + f_{x_2y}\mu_1'\right] + f_{x_1y}f_{x_2y}\mu_1'\mu_2' > 0.$$

When f is supermodular in all arguments, all cross partials are positive, and as a result, this inequality is satisfied for any distribution under monotone matching.

Notice that supermodularity is also necessary if the condition is to hold for all distributions. To see this, consider for example the first inequality (it holds for all inequalities): $\left[f_{x_1x_2}[\mu_2^{-1}]' + f_{x_1y}\right]\mu_1' > 0$. While f_{x_1y} can be negative as long as $f_{x_1x_2}$ is sufficiently positive, we can always pick distributions such that $[\mu_2^{-1}]' \to 0$, and as a result, the requirement is still that f_{x_1y} is non-negative.

2.3.2 Endogenous Firm Size

In reality, large firms endogenously determine their size. Eeckhout and Kircher (2017) combine the quantity and quality decision of the firm into a joint decision.[10] Firms face a trade off between hiring better and hiring more workers. Let the match output be given by $F(x, y, l, r)$, where x denotes the

[10] The model with fixed firm size is a limit case of this model. Examples of such fixed size models that are a limited case are the hierarchy models like Garicano (2000) (see also Antràs, Garicano, and Rossi-Hansberg (2006)). Applications to trade of sorting with endogenous quantity, augmented with multiple sectors, are in Grossman, Helpman, and Kircher (2017), Grossman and Helpman (2014) and Bombardini, Gallipoli, and Pupato (2014).

worker type, y the firm type, l the labor force, and r the resources the firm has available.[11] Firms can pick different types x, and choose $l(x)$ and $r(x)$ correspondingly. Total output is simply the sum of the output produced by different x types. Eeckhout and Kircher (2017) show that because of the additivity assumption, in equilibrium firms would only want to hire one worker type and the problem reduces to choosing *which* type x and how many of those. Moreover, the focus is on the case of constant returns in l, r, so we can write $F(x, y, l, r) = rf(x, y, \theta, 1)$ where $\theta = \frac{l}{r}$. The span of control θ of the manager can be interpreted as the size of the firm, for example if the manager is the CEO as in Lucas (1978). Normalizing the resources to 1, the problem of a firm y then is:

$$\max_{\tilde{x},\tilde{\theta}} f(\tilde{x}, y, \tilde{\theta}) - \tilde{\theta} w(\tilde{x}), \tag{27}$$

where $w(x)$ is the wage schedule of a worker of type x. The first order conditions satisfy:

$$f_\theta(x, y, \theta(x)) - w(x) = 0 \tag{28}$$

$$f_x(x, y(x), \theta(x)) - \theta(x) w'(x) = 0. \tag{29}$$

As before, this solution is a maximum if the second order condition is satisfied. Given the multiple dimensions of the problem, this requires the Hessian to be negative semidefinite. The Hessian is

$$\mathbf{H} = \begin{pmatrix} f_{\theta\theta} & f_{x\theta} - w'(x) \\ f_{x\theta} - w'(x) & f_{xx} - \theta w''(x) \end{pmatrix}. \tag{30}$$

and negative semidefiniteness requires $|\mathbf{H}| \geq 0$ or

$$f_{\theta\theta}[f_{xx} - \theta w''(x)] - (f_{x\theta} - w'(x))^2 \geq 0, \tag{31}$$

as well as $f_{\theta\theta} < 0$, which is assumed. Differentiating the first order conditions (evaluated at the equilibrium allocation) with respect to x, and substituting, this condition requires:

$$-\mu'(x)[f_{\theta\theta} f_{xy} - f_{y\theta} f_{x\theta} + f_{y\theta} f_x / \theta] \geq 0 \tag{32}$$

Positive sorting means $\mu'(x) > 0$, requiring $[\ \cdot\ ] < 0$. After writing this condition in terms of F, using the fact that from homogeneity of degree one, $-F_{lr} = \theta F_{ll}$ and $F_x(x, y, \theta, 1) = \theta F_{xl} + F_{yr}$:

$$F_{xy} F_{lr} \geq F_{yl} F_{xr}. \tag{33}$$

Sorting between workers and firms is positive under cross-margin-complementarity. It relates the within-complementarities in extensive and intensive margin to the between-complementarities across intensive and extensive margin. The within-complementarity in the extensive margin is the standard complementarity between types as in Becker (1973). Instead, the between-complementarity between job type and the number of workers captures span of control as in Lucas (1978): it is positive if the span of control of a better manager is larger. The between-complementarity of the worker type and the amount of firm resources indicates the role of worker resources: it is positive if the productivity of better workers is

[11] Here the set of workers and firms is disjoint. For a model that considers occupational choice where a pool of agents chooses whether to become a worker or a manager, see Eeckhout and Jovanovic (2011).

increasing more if they have more resources, say a better computer or more support. The equilibrium condition for sorting then compares the marginal increase of better workers versus the marginal impact of more workers.

Condition (33) can be shown to nest a large number of well-known models, such as Becker (1973) with fixed quantities, Lucas (1978) without worker heterogeneity, the model of efficiency units where heterogeneous agents are perfectly substitutes in a fixed proportion,... And it is worth pointing out that in the special case of Cobb-Douglas in quantities, condition (33) boils down to log-supermodularity.

The combination of two-sided heterogeneity and endogenous factor intensity pins down the equilibrium allocation and determines wages. The equilibrium solution is obtained after solving for a system of three differential equations that follow from the first order conditions and market clearing. In the case where the distributions of x and y are identical, under PAM these three differential equations are:

$$F_x = w'(x), \quad \mu'(x) = \frac{1}{\theta(x)}, \quad \text{and} \quad \theta'(x) = \frac{F_{yl} - F_{xr}}{F_{lr}}. \tag{34}$$

Whether the firm size is increasing in firm productivity depends on the relation between the span of control complementarity F_{yl} relative to the complementarity of managerial resources and worker skills F_{xr}. Under PAM, when $F_{yl} > F_{xr}$, the more productive firms are larger.

This simple setup has important implications for our understanding of technological change as we can infer from studying the composition of firms. The working paper version of Eeckhout and Kircher (2017) applies this setup to estimate a CES version of the model using German matched employer-employee data. What the estimation reveals is that there has been substantial skill biased technological change. But with the endogeneity of the firm size, there has also been a change in the technology that governs the span of control within the firm, i.e. there has been quantity biased technological change. Firms as a result have grown larger, and the increase in the wage premium that we observe is muted by the change in the firm size. The implication is that the skill premium would have been substantially higher had it not been for the quantity biased technological change.

2.3.3 Endogenous Firm Size and Full Support

Starting with the Lucas (1978) span-of-control model, the canonical model of firm size has allowed for heterogenous managers who have a different size work force underneath them – their span of control – depending on the manager's skill. But even if the workers are heterogeneous in skill, as in Eeckhout and Kircher (2017), the firm in equilibrium chooses exactly one type of worker.[12] There is sorting of heterogeneous managers with heterogeneous workers, but in equilibrium managers choose exactly one worker type. This is an equilibrium outcome – firms could choose multiple worker types, but it is not optimal to do so – that follows from the assumption that output is additively separable in worker inputs. This simplifying assumption makes the model analytically tractable.

While there is no general characterization result of the most general many-to-one matching model (Kelso and Crawford (1982)), there are some more specific models that have endogenous firm size *and*

[12]In an extension of the basic model, Eeckhout and Kircher (2017) show how the framework can be adjusted to allow for agents from disjoint distribution – as in the variation of Kremer (1993)'s model discussed above, but with *endogenous* size within each of the skill categories.

the equilibrium matching of different worker types in the same firm. Here we lay out the basic model in Pinheiro and Eeckhout (2014).

Let there be a continuum with measure one of firms indexed by y, distributed according to CDF $\Psi(y)$, and a measure M of workers of type x with CDF $\Gamma(x)$. Assume the technology is given by:

$$F(l(x), y) = y\left(\int_0^\infty h(l(x))x dx\right)^\beta, \tag{35}$$

where $l(x)$ is the number of workers of type x the firm hires and $h(\cdot)$ is an increasing and concave function. For example, when $h(l) = l^{\frac{1}{\beta}}$ this is the standard CES formulation of the labor aggregator.

The firm y chooses the function $l(x; y)$ for all x to solve the objective: $\max_{\tilde{l}(x)} F(\tilde{l}(x), y) - w(x)\tilde{l}(x)$. Markets are competitive, so it takes the hedonic wage schedule $w(x)$ as given, which is pinned down from market clearing $\int l(x; y)d\Psi(y) \leq M\Gamma(x)$. The equilibrium solution $l^\star(x, y)$ to this problem satisfies the First Order Condition $yh'(l) \leq w(x)$ (with equality if the solution is interior). Given $l^\star(x; y)$, we can construct the CDF $G(x; y) = \frac{l^\star(x;y)}{\int l^\star(x;y)dx}$, which is normalized to have measure one.

In principle, this technology can give rise to any allocation of worker skills $l^\star(x; y)$. An immediate result is that if $h'(0) = \infty$ (for example when $h(\cdot)$ satisfies the Inada conditions) then there is full support of all skills in all firms. Because the marginal product of the first worker is infinite, with finite wages all firms must hire at least some measure of all worker types.

Somewhat more surprising is that when the labor aggregator is CES (when $h(l) = l^{\frac{1}{\beta}}$), then the distribution of workers G is the same in all firms y and equal to the economy-wide skill distribution Γ. That is, firms can be orders of magnitude more productive (have a higher y) and yet they choose the same distribution of workers. Of course, more productive firms will choose a larger measure of workers. Productivity y thus determines firm size, but under CES it does not affect the skill distribution across firms.

Result. *If the labor input aggregator is CES, then all firms have identical skill distributions in equilibrium.*

Observing the skill distribution across firms then allows us to infer what the underlying production technology is. In particular, there is ample evidence that firms are not identically distributed, which would refute the CES technology. More importantly, there has been consistent evidence across different countries that the firm wage and skill distribution has evolved over time. In particular, most of the change in wage inequality can be attributed to the change in between-firm inequality rather than to the change in the within firm inequality. This has been shown for Germany (Card, Heining, and Kline (2013)), for the US (Song, Price, Guvenen, Bloom, and von Wachter (2015) and Barth, Bryson, Davis, and Freeman (2014)), for Brazil (Benguria (2015)), and for Sweden (Vlachos, Lindqvist, and Hakanson (2015)).

The fact that the model can produce a distribution of skills in each firm with a skill mix (including full support of all skills) is appealing in order to match the facts described. In particular, using the observed evolution of the distribution of skills, it permits us to investigate the properties of the technology at any given point in time. Observe that the technology that aggregates the skill inputs embodies a measure of complementarity. We can simply calculate the cross-partial $\frac{\partial^2 F}{\partial l(x_i)\partial l(x_j)}$ which indicates the strength of the complementarity. This will depend on the term β as well as on the curvature of $h(\cdot)$. Moreover, we can investigate how the technology has changed over time. Intuitively, we would expect that given the increase in between firm inequality, the complementarity between close skill types has gone up relative to the complementarity of distant skill types.

Exploiting the variation across cities, Eeckhout, Pinheiro, and Schmidheiny (2014) uses a similar framework with a city-level aggregate production technology. Now the type y represents city-specific Total Factor Productivity (TFP). More productive cities pay higher wages and attract more workers. Now the countervailing force that stops all workers to live in the most productive city is the cost of living, i.e., housing prices. Large cities pay higher wages. What is unclear is whether this is actually due to the fact that they attract more productive workers or simply because workers are uniformly more productive. In other words, is there any sorting based on skills across different cities. The data shows that the average skill is constant across cities of different sizes, but the variance of skills is larger in bigger cities. Therefore, there is spatial sorting, not on the average skill level, but across the entire distribution.

2.4 Whither Negative Assortative Matching?

In economic settings, it seems that all forces pull towards PAM. Like athletes in football, basketball, or hockey teams, putting the best players together brings out the best of each of these already star players. Firms trying to improve the quality of their professionals often make hiring decisions of groups of equally talented experts. All this indicates that there is a strong tendency towards PAM and with that, there is a tendency towards higher inequality in outcomes, where the high types generate most of the output. For example, on a uniform distribution of x and y and with a technology $f = x \cdot y$, under PAM the match of the highest types produces output of 1 and the lowest types 0. Under NAM, the output of the $(1, 0)$ and $(0, 1)$ matches is 0 and the highest match value produced by the mixed match $\left(\frac{1}{2}, \frac{1}{2}\right)$ is equal to $\frac{1}{4}$. Under NAM, there is a tendency for output to be distributed more equally. Of course the comparison is incomplete because the NAM allocation is suboptimal and it makes little sense to compare different technologies, but this is just meant as an illustration. The point is that under NAM, the distribution of output tends to be more equal than under PAM where, by supermodularity, the resources produced are concentrated with the high types at the detriment of the low types who are forced to match with each other.

But do we actually see examples of NAM in economic applications? A simple intuitive example of submodular technologies is the matching of pilots and co-pilots. The best pilot is very unlikely to need any assistance because she can handle the most complex circumstances in the cockpit. Therefore the added value of a co-pilot is minor. Instead, the worst pilot makes more mistakes and can deal with extreme circumstances less well, so the presence of an able co-pilot is all the more important. The optimal allocation of pilots to co-pilots in this example would therefore be NAM: the best pilot is matched with the worst co-pilot and the worst pilot is matched with the best co-pilot. Another example is WalMart, where the sales people are typically very low skilled whereas management are the most coveted executives.[13] These examples may seem somewhat far-fetched to have general applicability. In

[13] Grossman and Maggi (2000) argue that research teams might be submodular in some cases. A research project relies on the best idea and it is a waste to have a good "second-best" idea. Of course, this neglects the complementarities

biology, NAM is most prevalent precisely because it is a force towards stability. So is it at all that rare to see the negative in sorting in economics?

The answer: not necessarily. Once we dig into the micro foundations of the match surplus, it is not that rare to find NAM. We illustrate this with a canonical example of information aggregation.[14]

2.4.1 Aggregating Information leads to NAM

Consider a firm of fixed size, for now assume two workers. In the tradition of Radner (1962) and Marshak and Radner (1972), we assume that a significant role of the firm is to aggregate decentralized information held by the workers, who have a common objective to act optimally and produce maximal output given their joint information. There is uncertainty about the state of the world in which the firm operates and workers each draw a signal that is informative about this unknown state. Workers are heterogeneous in the information they possess, expressed by the precision of their signal. A worker type x draws a signal with a common mean (equal to the true state) and precision (inverse variance) x. This is the canonical set op of the Radner (1962) quadratic-normal model: the state of the world and the signals are normally distributed, and payoffs (losses) are quadratic in the distance from the realization of the state to the action chosen. The joint action a is chosen to maximize $\pi - (a-s)^2$ where π is a positive constant, and s is the realization of the state (normally distributed with variance τ^{-1}).

Given this setup, now consider a world in which firms compete to hire workers before they produce. In this first stage, there is a competitive matching market with wages $w(x)$. The willingness to pay for a worker of types x will depend on her x but also on the other worker that is in this two agent firm. In the second stage, and given the types of the two workers x_1, x_2, output is produced. We follow Chade and Eeckhout (2017b) in what follows. They show that the ex ante value of a firm that optimally chooses actions given independent signals is $V(x_1, x_2) = \pi - \frac{1}{\tau + x_1 + x_2}$. If signals are correlated, then the value further depends on the correlation coefficient. This payoff is increasing in the precision of any of the workers' signals x_i. Most interestingly though, the match value $V(x_1, x_2)$ is strictly submodular. That is easily seen from calculating the cross-partial derivative $V_{x_1 x_2} = \frac{-2}{(\tau + x_1 + x_2)^3} < 0$. With two workers in each firm, this implies that the optimal and equilibrium allocation is NAM, high types are matched with low types. The reason is that with a high type already, the additional value from a high precision worker is small. Instead, a firm with a low precision type has a much higher value from hiring a high precision worker. Hence they are willing to pay a higher wage. This negative sorting pattern leads to diversification: workers are maximally different within and minimally different between firms.

The same logic goes through once firms have sizes larger than two. There is still maximal diversification of the composition of the firms for the same reason as in the two worker firms. However, we cannot call this Negative Assortative Matching. While PAM readily extends to the case of more than two workers (form example in Kremer (1993)'s O-Ring technology), under diversification with more than two agents it is unclear how to define NAM, as there are more than two different agents involved. There is however a notion in which we can state that the teams are maximally diversified, based on the

between researchers.

[14]This example is not unique. Schulhofer-Wohl (2006) shows the NAM arises naturally in a the context of matching risk averse agents. The least risk averse agent takes on the risk of the most risk averse agent. This is optimal since the utility cost of the same risk is higher for the more risk averse agents. There are thus gains from trade, where the more risk averse agents pays a deterministic price that can induce the least risk averse agent to take on the higher risk.

notion of majorization of vectors. This requires finding the partition of worker precisions where firms are as equal as possible, and this is computationally intensive (the problem is NP-hard).

In sum, in the context of matching in the presence of a canonical information aggregation problem, NAM is the natural outcome when matches are one-to-one. Once the number of agents in a match is larger than two, the extension of NAM is maximal diversification. Unfortunately, the mathematical properties of that problem make the solution a lot more challenging.

2.4.2 Matching with Externalities

Even if we tend to observe positive sorting as an equilibrium outcome, negative sorting may actually be optimal. Of course, in the standard two sided matching model with TU, there is no source of inefficiency since the equilibrium allocation and the planner's solution coincide. But what if there are externalities?

In many output markets, there are few competitors and firms exert some market power. Or, firms may be competing for a patent in a zero sum race. The question we address here is how the presence of market power or externalities in the output market affects the labor market, and in particular the composition of skills within and across firms. Imagine therefore a two-stage matching problem in the spirit of the information aggregation setup above where in the first stage matches are formed in a competitive labor market. Once the workers are hired by the firms, firms compete in an output market with externalities or in an oligopolistic market. This is where the problem differs from the information aggregation model. The payoff of a given firm now is no longer a function of the composition of its own work force, but also a function of the work force composition of its competitors. Due to the externalities or the less than perfect competition in the output market, the central premise of the benchmark TU matching model is violated.

Consider the simplest possible set up (based on Chade and Eeckhout (2017a)) with a continuum of firms and workers, of which there are two types of workers $x \in \{\overline{x}, \underline{x}\}$ with equal measure. All firms are identical, and they can either hire two high types $(\overline{x}, \overline{x})$ denoted by $\overline{X}$, two low types $\underline{X} = (\underline{x}, \underline{x})$ or a mixed worker composition $\hat{X} = (\overline{x}, \underline{x})$. The labor market is competitive with wages $\overline{w} = w(\overline{x})$ and $\underline{w} = w(\underline{x})$. Output V and wages are determined in the output market and depend on the market structure or the exact nature of the externality. Assume that once the workers are hired, firms randomly match with another firm (Chade and Eeckhout (2017a) consider a general setup that also includes deterministic matching and economy-wide spillovers). Therefore, if a firm with worker composition X_i matches with a firm X_j, the payoff will depend on both firm compositions: $V(X_i|X_j)$. Clearly, a firm with high skilled workers is more likely to obtain the patent (or will achieve a higher market share) if its competitor has low skilled workers only or a mixed composition than if it has all high skilled workers as well. Now under random matching, the expected payoff depends on whether the matching pattern μ is PAM (μ_+) or it is NAM (μ_-). The expected payoff in both cases is:

$$\text{PAM:}\quad \mathcal{V}(X_i|\mu_+) = \frac{1}{2}V(X_i|\overline{X}) + \frac{1}{2}V(x_i|\underline{X}) \tag{36}$$

$$\text{NAM:}\quad \mathcal{V}(X_i|\mu_-) = V(X_i|\hat{X}). \tag{37}$$

Because of the externalities, the allocation in the hiring stage will differ in equilibrium from the planner's solution, that is, a planner who takes the externality or the market structure as given. The planner chooses the allocation that maximizes the total sum of match values, μ_+ or μ_-. In fact, it can be shown that for some match values there is even an interior solution to the planner's problem with a fraction $\alpha \in (0,1)$ of matches μ_+ and a fraction $1-\alpha$ of matches μ_-. Under some conditions and after solving the planner's optimization problem, there will be PAM provided $\mathcal{V}(\overline{X}|\mu_+) + \mathcal{V}(\underline{X}|\mu_+) > 2\mathcal{V}(\hat{X}|\mu_-)$.

In equilibrium, firms take wages as given. Then under PAM, an equilibrium must satisfy:

$$\mathcal{V}(\overline{X}|\mu_+) - \overline{w} > \mathcal{V}(\hat{X}|\mu_+) - \underline{w} \tag{38}$$

$$\mathcal{V}(\underline{X}|\mu_-) - \underline{w} > \mathcal{V}(\hat{X}|\mu_-) - \overline{w} \tag{39}$$

and after summing both equations, it must be that $\mathcal{V}(\overline{X}|\mu_+) + \mathcal{V}(\underline{X}|\mu_+) > 2\mathcal{V}(\hat{X}|\mu_+)$. Under NAM, the condition is $\mathcal{V}(\overline{X}|\mu_-) + \mathcal{V}(\underline{X}|\mu_-) < 2\mathcal{V}(\hat{X}|\mu_-)$. It is clear here that in equilibrium, firms do not take into account the effect their choice has on other firms.[15] This can therefore for obvious reasons lead to inefficient allocations. In particular, it may be that the equilibrium outcome is PAM because $\mathcal{V}(\overline{X}|\mu_+) + \mathcal{V}(\underline{X}|\mu_+) > 2\mathcal{V}(\hat{X}|\mu_+)$, yet the planner would choose NAM because $\mathcal{V}(\overline{X}|\mu_+) + \mathcal{V}(\underline{X}|\mu_+) < 2\mathcal{V}(\hat{X}|\mu_-)$. It may seem a minor notational difference (the fact that on the right hand side of these two equations the planner takes the changed allocation μ_- into account whereas the firms do not in equilibrium), but the impact can be large. Note that in this example, the planner chooses a radically different allocation (NAM) than what results as the equilibrium outcome (PAM).

This has important policy implications. If the planner cannot regulate the output market that gives rise to the externalities in the first place, then it may nonetheless be able to affect welfare through intervention in the competitive labor market. In the example above, welfare improves if the firms can be induced to hire mixed work forces (NAM) rather than perfectly sorted work forces (PAM). That would mean that incentives will be provided to ensure that say one pharmaceutical company does not have all the best scientists but rather that all firms have some of the best mixed with some less talented scientists. Some of these policy interventions already exists in the context of sports competitions, especially in the US. There, policies are in place to ensure that there is not one team that manages to sign all the best players. Those policies include a rookie draft where the worst performing team gets the first pick of the new incoming players, as well as salary caps and taxes for overspending. Those policies ensure that the competition is balanced and as a result, the likelihood that a team repeats as the league winner is much lower than in European soccer competition where no such policies are in place. The balanced competition is meant to offset the zero sum nature of competitions with a winner-takes-all aspect. The best teams tend to win repeatedly and more often, and thus generate more revenues that allow them to sign even better players.

3 Unemployment and Sorting

In order to find the right (wo)man for the jobs, firms spend quite a bit of resources. Some of this is monetary, but quite a bit is costly search. With heterogeneity of worker types, search becomes particularly interesting because in addition the time it takes to find a worker, some workers are rejected because they are not the right match. Search frictions naturally give rise to equilibrium unemployment.

[15] Unlike the benchmark TU matching model, the model with externalities can also give rise to multiple equilibria.

Productive workers (and productive jobs) remain idle because it is mutually optimal to delay forming a match because waiting for a better match outweighs the search cost.

The search literature has been extensively reviewed. For an overview on search frictions in the labor market in the absence of sorting, see Rogerson, Shimer, and Wright (2005). For a review of search with unemployment, see Smith (2011) and Chade, Eeckhout, and Smith (2017).[16] Here, I will restrict myself to three aspects. First, I will present a highly stylized model of random search that serves as the basis for the analysis of mismatch and to make the argument about identifying sorting below. Second, I will consider directed search with sorting when firms are large. Third, I will briefly address how asset holdings lead to sorting by risk averse workers who look for ways to smooth consumption, even in the absence of technological complementarities.

3.1 Random Search and Sorting

Consider an extremely stylized version of the random search model with sorting by Shimer and Smith (2000), based on Eeckhout and Kircher (2011). It has only two periods and the search cost is additive (as in Atakan (2006) or Chade (2001)) rather than from discounting. The reason for this stylized model is that we want to obtain an analytical solution for the wages that we will exploit below. As is apparent from the value functions in the infinite horizon version of the model that we discuss below, there is no explicit solution for wages in Shimer and Smith (2000).

In the first stage, there is costless random meetings between workers x and firms y. If a match is formed in the first stage, the wage is determined as the equal split of the surplus over the value of rematching in stage two. In stage two, at a cost c to both the firm and the worker, there is a round of frictionless matching and wages are determined as in the Becker model. At the end of stage two all agents are matched and production of all pairs from both stages is realized. There is no discounting.

Assume symmetry in the match value function $f(x, y) = f(y, x)$, and uniform distributions of types x and y. Denote by $\mu(x)$ the Beckerian, frictionless allocation and by $w^\star = w(x, \mu(x))$ the Beckerian wage. Likewise for profits $\pi^\star$. The continuation values at the end of stage two are then $w^\star(x) - c$ and $\pi^\star(y) - c$. A match between a pair (x, y) is formed in stage one if the surplus is positive:

$$f(x, y) - (w^\star(x) + \pi^\star(y) - 2c) \geq 0. \tag{40}$$

For a given x, there is a marginal type $\underline{y}$ below the Beckerian type and a marginal type $\overline{y}$ above where the surplus is zero. The acceptance region is illustrated in Figure 2, where the solid line is the Beckerian allocation $\mu(x)$.

Wages are equal to the outside option plus half the share of the surplus:

$$w(x, y) = \frac{1}{2}\left[f(x, y) - w^\star(x) - \pi^\star(y) + 2c\right] + w^\star(x) - c = \frac{1}{2}\left[f(x, y) + w^\star(x) - \pi^\star(y)\right]. \tag{41}$$

Below, when discussing mismatch, we will return to the properties of the wages.

[16] Initially the frictionless matching literature (reviewed in Roth and Sotomayor (1990)) and the Diamond-Mortensen-Pissarides search literature developed in parallel with little interaction. Then, there was a burst of simultaneous discoveries of the sorting model with search frictions and without transfers: McNamara and Collins (1990), Morgan (1996), Burdett and Coles (1997), Eeckhout (1999), Bloch and Ryder (1999), Chade (2001), and Smith (2006). This then later led to the sorting model with search frictions and transfers in Shimer and Smith (2000) and Atakan (2006).

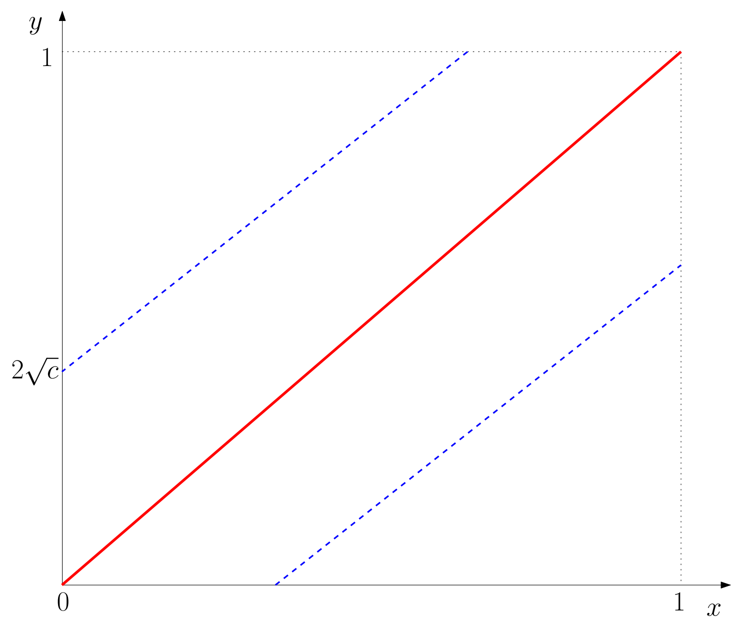

Figure 2: Acceptance sets for $f = xy$.

In an infinite horizon version of this simplified random search model, the same logic applies. Consider a fully symmetric setup with type distributions $\Gamma(x) = \Psi(y)$. Denote by $v(x)$ the value function (equal to $v(y)$):

$$v(x) = \int_{\mathcal{M}(x)} w(x,y) d\Gamma(y) + \int_{y \notin \mathcal{M}(x)} d\Gamma(y)[v(x) - c], \tag{42}$$

where $\mathcal{M}(x)$ is the matching set for a type x. Wherever the match surplus is positive a match is accepted. The marginal type $\underline{y}$ satisfies: $f(x, \underline{y}(x)) - v(x) - v(\underline{y}(x)) = -2c$. The wage can be written as: $w(x,y) = \frac{1}{2}\left[f(x,y) - v(x) - v(y) + 2c\right] + v(x) - c$. Note that the wage expression depends on the value $v(x)$, which means that there is no explicit solution. Still, the wage will have the same properties as in the two period model.

The expression in (42) can be written more intuitively. Denote the lowest wage $\underline{w}(x) = v(x) - c$ and the average wage $\mathbb{E}w(x)$. Now rejection of a match does not lead to the Beckerian outcome with the highest wage, but a new random draw. If accepted, that draw yields the average wage conditional on acceptance. Therefore, the value function is equal to

$$v(x) = \pi \mathbb{E}w(x) + (1 - \pi)\underline{w}(x), \tag{43}$$

where $\pi = \text{Prob}\{\mathcal{M}\}$ is the probability of acceptance of a match, and we can further write $c = \left[\mathbb{E}w(x) - \underline{w}\right]\pi$.

In this setting, there is PAM whenever $f(x,y)$ is supermodular (Atakan (2006)). Things change when instead of an additive cost as in this example, there is discounting and the cost of search is the opportunity cost of not matching. The central result established in Shimer and Smith (2000) is that a stronger requirement is needed than supermodularity. Under fairly general conditions, there is PAM if the match surplus function $f(x,y)$ is log-supermodular, i.e., $\frac{\partial^2}{\partial x \partial y} \log f(x,y) > 0$. The reason for this stronger requirement is that with discounting the cost of search is proportional to the value of being matched. Higher types have higher values and therefore a higher opportunity cost of search. If $f_{xy} \approx 0$, then there are little gains from PAM even without search frictions, and with search frictions it is more costly to postpone a match for the high types. Therefore, the high types will choose to trade faster than low types by accepting a wider range of matches. This in turn violates PAM. As the degree of complementarity increases, the higher opportunity cost of search is compensated by a higher value of match. It turns out that under log-supermodularity this benefit outweighs the cost. Shimer and Smith (2000) also show existence of a sorting equilibrium with random search. And finally, equilibrium is not generally unique. This is due to the endogeneity of the type distribution, the selection effect of which can give rise to multiplicity. This has been shown with an explicit example in the context of search with NTU by Burdett and Coles (1997).

3.2 Unemployment, Sorting and Large Firms

When analyzing matched employer-employee data, it is hard to ignore the existence of two sided heterogeneity of workers and firms, of unemployment and of large firms. Guided by data on individual level observations of firm and worker characteristics in combination with detailed information of the transitions of workers between jobs, it becomes apparent that matching models that feature two-sided heterogeneity and sorting as well as unemployment and search frictions are well suited to analyzed those data. Moreover, there is a distribution of firm sizes, each firm matched with multiple workers.

Existing models combine two of these three features, several have been discussed above: unemployment and large firms (Smith (1999), Acemoglu and Hawkins (2014), Kaas and Kircher (2014)), sorting and unemployment (Shimer and Smith (2000), Atakan (2006), Eeckhout and Kircher (2010)), sorting and large firms (Eeckhout and Kircher (2017)). We now present a setup that incorporates all three aspects: firm size, sorting, and unemployment.

It is a directed search model that combines the sorting with large firms model discussed above and the directed search model in Eeckhout and Kircher (2010). Firms post wages w as well as the number of vacancies v at a cost c. Wages in competition with other firms determine the ratio of workers to firms λ for a given firm-worker type pair (x, y), and given a matching function $m(\lambda)$ they also determine the matching probability. As before, given separability of output between different worker types x, it is optimal for a firm to only hire workers of one type x. Then the firm faces the following problem:

$$\max_{l,w,v} F(x, y, l, r) - lm(\lambda)w - vc \tag{44}$$

$$\text{s.t. } l = vm(\lambda); \quad w\frac{m(\lambda)}{\lambda} = U(x) \quad \text{and} \quad \lambda = \frac{u}{v}. \tag{45}$$

This problem is equivalent to:

$$\max_{x,u,v} F\left(x, y, vm\left(\frac{u}{v}\right), r\right) - uU(x) - vc. \tag{46}$$

The solution to this problem will also satisfy the solution to $\max_{x,u}\left[G(x, y, u, r) - uU(x)\right]$, where $G(x, y, u, r) = \max_v \left[F\left(x, y, vm\left(\frac{u}{v}\right)\right) - vc\right]$. We can immediately apply conditions (33) to determine whether there is PAM: $G_{xy}G_{ur} \geq G_{yu}G_{xr}$, provided G is homogeneous of degree one in (u, r). The equilibrium allocations again satisfies the three differential equations in (34) applied to G.

In this setup, queue lengths vary by worker type x and by firm size (which here coincides with the firm type y), as they do in the data. This is the first model that incorporates the joint role of unemployment, sorting and firm size. It can serve to take it to detailed matched employer-employee data.

3.3 Asset Holdings and Unemployment

In the presence of unemployment and search frictions, sorting actually does not need any technological complementarities. Following Eeckhout and Sepahsalari (2014), here I illustrate how the need for consumption smoothing by risk averse agents naturally leads to sorting. Workers with different asset holdings sort into different productivity jobs, despite them having the same ability or skill.

Consider a simple two period economy. Workers are heterogeneous in asset holdings a and firms are heterogeneous in the productivity of the their vacant job y. A worker is unemployed in period one and searches for a job in period two, when she may either become employed or stay jobless. She faces a joint decision problem: given current assets a, she chooses next period's assets a' as well as the type of job y to apply to. Search is directed, so the higher the wage paid at a job, the more applicants there will be. The worker's problem can be written as:

$$U(a) = \max_{a',\theta} \left\{ u(a-a') + \beta \left[m(\theta)u(Ra'+w') + (1-m(\theta))u(Ra') \right] \right\}, \tag{47}$$

where $u(\cdot)$ is the worker's utility from consumption, β is the discount factor, R is the risk-free return on assets, w is the firm-specific wage, and $m(\theta)$ is the matching probability given (submarket-specific) tightness θ. Observe that workers face labor market risk because in period two, their consumption may be low if they do not find a job.[17]

In this context, Eeckhout and Sepahsalari (2014) show that despite the absence of any technological complementarities (workers are identical), under fairly standard conditions on the preferences there is sorting of worker assets a on job productivity y. Low asset holders choose lower productivity jobs that pay lower wages because those applications are filled at a higher rate. This allows the workers with low assets to self-insure. High asset holders also face job market risk but they are less exposed given their asset holdings. In the infinite horizon version of this model, workers gradually build up assets while employed and they gradually decrease their asset holdings while unemployed. This also means that the same worker applies for worse jobs as the unemployment duration gets longer.

4 Mismatch and Sorting

The ultimate objective of the models reviewed in this paper is that they allows us to gain insights into the underlying production technology. Like preferences, production technologies are hypothetical constructs that we do not observe. Yet, they are extremely important to understand and predict the behavior of economic agents under counterfactual circumstances, for example as a result of a new policy intervention. Like with revealed preference and consumption choices, we observe hiring decisions by

[17] There is a market incompleteness that prohibits workers from perfectly insuring against labor market risk. See also Golosov, Maziero, and Menzio (2012) and Acemoglu and Shimer (1999) for a setup with homogenous asset holdings.

firms that inform us about the properties of the underlying technology. Here we return to the one-to-one matching environment. By equating a firm with a job, it is as if a firm that has multiple jobs has a technology that is additively separable and the firm size is irrelevant.

One of the obstacles with the benchmark sorting model with TU (say Becker (1973)) is that its prediction is extremely stark. Deterministic types generically give rise to an allocation where a given type x matches exactly one y. Graphically, that means that the matching $y = \mu(x)$ is single valued. That is of course not very useful to take to the data as we never observe such matching pattern predicted by the benchmark sorting model. Instead of the red line in Figure 2, we tend to observe less than such perfectly matched pairs, i.e. a cloud of points rather than a line. In what follows we will refer to this as mismatch, even though it is a bit of a misnomer because it may be completely optimal not to match with one type only.

We consider three reasons for mismatch: search frictions, stochastic sorting and multidimensional types. In each case we discuss how the perceived mismatch can help the identification of the complementarities in production.

4.1 Using Search Frictions to Identify Sorting

Despite the casual observation that better firms hire better workers, there is surprisingly little evidence to support this. Are more skilled workers more productive in better firms? Or is it better for more productive workers to be paired with less productive colleagues as suggested by Mas and Moretti (2009) in their study of checkout cashiers? The amount of effort and resources organizations invest in hiring the "right" person for the job indicates that the exact assignment is important for efficiency. Yet, there is little direct evidence. The most widely cited direct evidence concludes that there is actually neither complementarity nor substitutability between workers and jobs. In a seminal paper, Abowd, Kramarz, and Margolis (1999) analyze the correlation between firm and worker fixed effects from wage regressions. The obtained correlation aims to provide evidence about the complementarity or substitutability, and if so its magnitude. The idea is that more productive firms pay higher wages than lower wage firms irrespective of the exact worker they hire, and the firm fixed effect therefore recovers the ranking of the firms. Using this fixed effects regression method, Abowd, Kramarz, and Margolis (1999) find that the correlation between the worker and firm fixed effects is zero, and they conclude that there is no evidence of sorting.

While this appears plausible, it turns out that in a simple model model with search fictions this logic is not consistent. Eeckhout and Kircher (2011) point out that the central question is how wages are determined in equilibrium, and how observed wage data are affected by equilibrium mismatch. We therefore return to the wage equation (41) explicitly derived in the simple two-period matching model.

Consider a worker-job pair with mismatch (in Figure 2, off the red line and between the blue lines). In the presence of a friction (the cost c of getting a new match), there is a tradeoff between separation followed by a new match and staying in the current match. For given frictions, the larger the mismatch, the larger the incentive to face the cost and rematch. The blue line indicates the worker-firm pairs (x, y) that are indifferent between continuing the search or staying in the match. Therefore, along the line the surplus of the match is zero and by equation (41), the wage must be exactly equal to the outside option. At any point strictly inside the blue lines, the surplus is strictly positive and the wage is strictly larger than the outside option.

Because there is a locus of indifference below (at $\underline{y}$) as well as above (at $\overline{y}$), the wage must be equal to the outside option at two points. As a result, the wage is necessarily non-monotonic in y. Even though more productive firms generate higher output ($f_y(x, y) > 0$), their outside option $\pi^\star$ is also increasing in y. For low y, the wage increases in y whereas for high y the wage decreases in y. At the Beckerian allocation $f_y(x, y) = \pi^\star_y$ and the wage is highest. This is because wages reflect the opportunity cost of mismatch. A high y firm demands a larger share to be willing to match with a low x worker in order not to continue searching.

The non-monotonicity of wages in firm type implies that the standard procedure to use firm fixed effects and to correlate it with worker type is ill-suited. That procedure requires the identifying assumption that wages are monotonic in firm type, which is not the case when there is mismatch due to search frictions. Because of the non-monotonic effect of firm types on wages, the wage cannot be decomposed in an additively separable firm and worker fixed effect and the firm fixed effect misses any direct connection to the true type of the firm.

In other work, Bagger and Lentz (2014), Lise, Meghir, and Robin (2016), and Lopes de Melo (2017) have run this fixed effects regression on data that was generated from simulated search models with strong complementarities. Those regressions give small or negative estimates of the correlation between worker and firm fixed effects. This further indicates there is a systematic bias using the fixed effects method when there is mismatch due to search frictions.[18]

The degree of complementarity can be backed out using the search behavior by workers. When the degree of complementarity is high, acceptance sets are small, and large otherwise. There are two components. First, the cost of search can be extracted from the range of wages paid. The difference between the highest and the lowest wage corresponds to the cost of search: the lowest wage is equal to the outside option, and in turn the outside option is equal to the highest wage less the cost of search. Second, given the search cost, the fraction of the firm population that an agent is willing to match with, i.e., the range of the matching set, identifies the strength of the complementarity. To see this, the loss from mismatch can be expressed by the (absolute value of the) cross-partial of the production function. For the case of PAM:

$$L(x, y) = f(x, y) - w^\star(x) - \pi^\star(y) = -\int_y^x \int_{\tilde{y}}^x |f_{xy}(\tilde{x}, \tilde{y})| d\tilde{x} d\tilde{y} \tag{48}$$

where $w^\star(x) = \int_0^x f_x(\tilde{x}, \mu(\tilde{x})) \tilde{d(y)} x$ and likewise for $\pi^\star$. The search decision hinges on whether the loss from mismatch does not exceed the total cost of search, or equivalently, whether the surplus is positive. At the marginal type $\underline{y}$: $L(x, \underline{y}(x)) = -2c$. This functional equation uses as inputs the range of matches $\underline{y}(x)$ as well as the cost of search.[19]

[18] For recent developments on identification see also Bonhomme, Lamadon, and Manresa (2015), Sorkin (2015), and Hagedorn, Law, and Manovskii (2017).

[19] As Teulings and Gautier (2004) point out, there is also *local* identification of the cross-partial. Considering the equilibrium wage equation, $w(x, y) = \frac{1}{2}[f(x, y) + w^\star(x) - \pi^\star(y)]$, the cross-partial of wages is proportional to that of the match value: $w_{xy} = \frac{1}{2} f_{xy}$.

4.2 Stochastic Sorting

Another reason why workers of a given type are not matched with firms of only one type simply has to do with the stochastic nature of productivity. It has long been recognized that labor can be interpreted as an experience good (see Nelson (1970) and Jovanovic (1979) for example). The firm hires a worker based on what is observable (education, work history, interview,...), but it takes some time to realize her true productivity. This may have to do with general human capital characteristics that are difficult to observe such as non-cognitive skills, or with firm-specific human capital. Note that this feature might be particularly salient the more skilled the worker is[20] and the longer it takes to learn the unobserved characteristic.[21]

There is quite a bit of literature analyzing the role of the gradual revelation of information about the worker's productivity, the early work dating back to Jovanovic (1979). Here we focus on the impact of sorting in the presence of learning: ex ante, both workers and firms are heterogeneous and the stochastic nature of their productivity depends on the ex ante heterogeneity.[22] Consider the simplest possible setting based on Chade and Eeckhout (2014) with two periods. Workers and firms match in period one, and in period two their type is realized and output is produced. Denote the ex ante type by x for workers and y for firms, and the ex post type by ω and σ respectively. Let the cumulative distribution of realized ex post types (ω, σ) be $H(\omega, \sigma|x, y)$ with marginal distributions $F(\omega|x)$ and $G(\sigma|y)$. Think for example of x as being the education of the worker (degree, school,...) and ω as her ability to code in C++. Matching occurs in stage one when only the ex ante types are known, and transfers can be made or can be committed to be made in the second stage. A crucial assumption here is that are no separations. Contingent on the realization of the outcome, matched workers and firms may have incentives to separate and rematch.[23]

When preferences are linear and there is TU, the match surplus function is now simply the expected payoff $V(x, y) = \int_\omega \int_\sigma q(\omega, \sigma)h(\omega, \sigma|x, y)d\sigma d\omega$. From the sorting result in the benchmark model (Becker (1973)) we know that sorting pattern exclusively depends on the match payoff function (here denoted by $q(\omega, \sigma)$): PAM if q is supermodular; NAM if q is submodular. Now the sorting pattern depends not only on the match surplus function q, but also on the stochastic properties of the distribution of types $H(\omega, \sigma|x, y)$. In particular, Chade and Eeckhout (2014) show the following result:

Result *The optimal sorting pattern is PAM if H is supermodular (submodular) in (x, y) and q is supermodular (submodular) in (ω, σ).*

A similar set of conditions apply for the case of NAM (basically there is NAM if H is supermodular and q is submodular or vice versa).

[20]Contrary to most academic economists' belief, uncovering the true ability of a graduate student or of an assistant professor is marred with a remarkable amount of uncertainty, even after we have information on grades or even on the job market outcome. Athey, Katz, Krueger, Levitt, and Poterba (2007) find that the ranking of economics PhD students upon entry is uncorrelated with their ranking in job placement. And Conley and Önder (2014) find that even if the distribution of research output as assistant professors of graduates from the top institutions first order stochastically dominates that of lower ranked institutions, the difference is small and the variance is much more sizable.

[21]For the equilibrium impact of the increasing importance of on-the-job human capital accumulation, see Cairo (2013).

[22]This is also the case in the learning models of Anderson and Smith (2010) and Eeckhout and Weng (2010).

[23]For an infinite horizon model with sorting and shocks to productivity as well as search frictions, see Eeckhout and Weng (2017).

A few things are of interest here. First, the main takeaway is that PAM depends on the properties of *both* q and H, and there could well be PAM if q is submodular provided H has the right properties. Second, when ω and σ are drawn independently, $H = F \cdot G$. If F and G also satisfy first order stochastic dominance, then PAM arises when q is supermodular, since first order stochastic dominance of F, G implies supermodularity of H. Third, sorting with deterministic types (Becker (1973)) is a special case. It simply suffices to choose F and G with a mass point and such that there is first order stochastic dominance. Then the result readily extends. Finally, results need not be restricted to supermodular distributions H. Roughly speaking, supermodularity captures some notion of first order stochastic dominance, but we can also consider second order stochastic dominance to capture notions of risk (or any stochastic order for that matter). The important insight is that the sorting pattern depends *jointly* on the properties of the match surplus *and* the stochastic order of the type distributions.

When preferences are not linear and there is imperfect transferable utility (ITU), the setup can be extended to allow for risk sharing and incomplete contracts. While the derivation of the equilibrium allocation is considerably more complex, the setting with ITU is appealing for applied work. Not only are non linear preferences more realistic, in this setting they ensure that the ex post payoffs are uniquely pinned down.[24] For example, Chade and Eeckhout (2014) embed the standard principal-agent model of Holmström and Milgrom (1987) in a matching framework with stochastic types in order to study CEO performance.[25] Optimal contracting resolves the incentive problem given moral hazard and stochastic output and manager types, and ex ante, firms and workers match based on their observable types. Using data on worker compensation and firm valuation, estimating this model shows that the role of the stochastic component is large. CEO performance embeds a large stochastic component of their ability. If firms could ascertain the realized outcome before contracting, there would be efficiency gains because CEOs would be reassigned across different productivity firms.

4.3 Multidemensional types

Finally, another reason why matches are not as predicted by the Beckerian model with exactly one matched firm type for each worker type derives from the multidimensionality of types. For an exhaustive theoretical treatment of matching multidimensional types in the TU setting, see Lindenlaub (2017).[26] While nothing changes in terms of the equilibrium concept, the notion of sorting takes on a whole new life. To some extent, the two-dimensional model has some of the features of the sorting model with large firms in Eeckhout and Kircher (2017). Agents maximize with respect to a bundle of characteristics (each dimension instead of type and firm size). This now implies that sorting changes smoothly with types.

Quantifying the effect of multidimensional types, Lindenlaub (2017) estimates the model and finds that the increase in wage inequality over the last few decades can be explained in part by the change in the degree of complementarity of cognitive skills. This technological change, even if the distribution of skills remains unchanged, has a profound impact on the job allocation and hence on wages. Introducing

[24] Given that ex ante matched partners care about the expected payoff, with TU there is a continuum of splits of the ex post surplus that generate the same ex ante match value.

[25] Very much like the CEO matching models of Gabaix and Landier (2008) and Terviö (2008), except that now types are stochastic.

[26] For a simple example in the trade literature, see Ohnsorge and Trefler (2007).

search frictions, Lindenlaub and Postel-Vinay (2016) show that using a model with one dimensional types leads to biased estimates of complementarities (such as those in Hagedorn, Law, and Manovskii (2017)) if the true underlying distribution is multidimensional.[27]

With multidimensional types, apparent mismatch can occur if the econometrician observes only a limited number of those characteristics that determine a match. Now if we observe mismatch on observable characteristics, we can posit that the source of mismatch is due to unobservable characteristics. Matching is assumed to be frictionless and mismatch is not there in the eye of the matched agent, but only in the eye of the econometrician. This is the setup proposed by Choo and Siow (2006). The question is whether we can identify the technology (or preferences in the case of marriage). In general, the answer is negative. However progress can be made under the identifying assumption of separability. Separability imposes that the distribution of unobservables within a class is independent of the matched type. Thus, there are only complementarities in the observed component but not in the unobserved component. Though the identification is exact, this now permits progress since variation across time or locations for example can be exploited. The assumption of separability is of course strong. To see this, consider the case where education is the only observable. Then more educated match with more educated, but within an educational class, all educated randomly match on the unobservable characteristics since no supermodularity is allowed on those unobservable characteristics. Galichon and Salanié (2015) are able to somewhat generalize the separability assumption – they can allow for complementarity of an observable component on one side with the unobservable component on the other side of the market – but there cannot be complementarities between the unobservable components on both sides.

5 Concluding Remarks

Workers have heterogeneous skills and a worker's productivity depends on the job she performs, the quality and sophistication of the capital equipment she works with, as well as on the skills of the workers she collaborates with. Sorting, the allocation of workers to firms (and to each other) in the presence of such peer group effects, has implications for the output produced.

In this review, I have focused attention on sorting in settings that are relevant for economics applications in macro, labor and industrial organization. In particular, I have reviewed models that feature large firms, risk aversion, asset holdings, multidimensional characteristics, unemployment and market power. Understanding technology is important for numerous policy questions. The sorting patterns observed – in particular who matches with whom and how workers are distributed within and between firms – allow us to infer indirectly the properties of the technology of production that uses firm and worker characteristics as inputs. This allows us to better understand recent trends in technological change, wage inequality and workers and job dynamics.

[27]See Lise and Postel-Vinay (2016) for more evidence on the importance of multidimensional characteristics with job search.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

Acemoglu, D., and W. B. Hawkins (2014): "Search with multi-worker firms," *Theoretical Economics*, 9(3), 583–628.

Acemoglu, D., and R. Shimer (1999): "Efficient Unemployment Insurance," *Journal of Political Economy*, 107(5), 893–928.

Ackerberg, D. A., and M. Botticini (2002): "Endogenous Matching and the Empirical Determinants of Contract Form," *Journal of Political Economy*, 110(3), 564–591.

Anderson, A., and L. Smith (2010): "Assortative Matching with Evolving Human Capital," *Review of Economic Studies*, 77(1), 3–29.

Antràs, P., L. Garicano, and E. Rossi-Hansberg (2006): "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121(1).

Atakan, A. (2006): "Assortative Matching with Explicit Search Costs," *Econometrica*, 74, 667–680.

Athey, S., L. F. Katz, A. B. Krueger, S. Levitt, and J. Poterba (2007): "What Does Performance in Graduate School Predict? Graduate Economics Education and Student Outcomes," *The American Economic Review*, 97(2), 512–518.

Bagger, J., and R. Lentz (2014): "An empirical model of wage dispersion with sorting," Discussion paper, National Bureau of Economic Research.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2014): "It's where you work: Increases in earnings dispersion across establishments and individuals in the US," Discussion paper, National Bureau of Economic Research.

Becker, G. (1973): "A Theory of Marriage I," *Journal of Political Economy*, 81, 813–846.

Benguria, F. (2015): "Inequality Between and Within Firms: Evidence from Brazil," *Available at SSRN 2694693*.

Bloch, F., and H. Ryder (1999): "Two-Sided Search, Marriages, and Matchmakers," *International Economic Review*, 41, 93–115.

Bombardini, M., G. Gallipoli, and G. Pupato (2014): "Unobservable skill dispersion and comparative advantage," *Journal of International Economics*, 92(2), 317–329.

Bonhomme, S., T. Lamadon, and E. Manresa (2015): "A distributional framework for matched employer employee data," *Unpublished manuscript, University of Chicago.*

Burdett, K., and M. Coles (1997): "Marriage and Class," *Quarterly Journal of Economics*, 112, 141–168.

Cairo, I. (2013): "The slowdown in business employment dynamics: The role of changing skill demands," *Unpublished manuscript, Universitat Pompeu Fabra.*

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128(3), 967–1015.

Chade, H. (2001): "Two-Sided Search and Perfect Segregation with Fixed Search Costs," *Mathematical Social Sciences*, 42, 31–51.

Chade, H., and J. Eeckhout (2014): "Stochastic Sorting," UCL mimeo.

——— (2017a): "Competing Teams," University College London Working Paper.

——— (2017b): "Matching Information," *Theoretical Economics*, forthcoming.

Chade, H., J. Eeckhout, and L. Smith (2017): "Sorting through search and matching models in economics," *Journal of Economic Literature*, 55(2), 493–544.

Chiappori, P.-A., and P. Reny (2015): "Matching to Share Risk," forthcoming, Theoretical Economics.

Choo, E., and A. Siow (2006): "Who marries whom and why," *Journal of political Economy*, 114(1), 175–201.

Conley, J. P., and A. S. Önder (2014): "The Research Productivity of New PhDs in Economics: The Surprisingly High Non-Success of the Successful," *The Journal of Economic Perspectives*, 28(3), 205–215.

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

Costinot, A., and J. Vogel (2015): "Beyond ricardo: Assignment models in international trade," *economics*, 7(1), 31–62.

Eeckhout, J. (1999): "Bilateral Search and Vertical Heterogeneity," *International Economic Review*, 40, 869–887.

——— (2006): "Local supermodularity and unique assortative matching," Discussion paper, Universitat Pompeu Fabra.

Eeckhout, J., and B. Jovanovic (2011): "Occupational choice and development," *Journal of Economic Theory*, 147(2), 657–683.

Eeckhout, J., and P. Kircher (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78, 539–574.

Eeckhout, J., and P. Kircher (2011): "Identifying Sorting–In Theory," *Review of Economic Studies*, 78(3), 872–906.

——— (2017): "Assortative Matching with Large Firms," *Econometrica*, forthcoming.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014): "Spatial Sorting," *Journal of Political Economy*, 122(3), 554–620.

Eeckhout, J., and A. Sepahsalari (2014): "Unemployment risk and the distribution of assets," Discussion paper, Technical report, Manuscript.

Eeckhout, J., and X. Weng (2010): "Assortative Learning," UCL Working Paper.

——— (2017): "The Technological Origins of the Decline in Labor Market Dynamism," Discussion paper, UCL mimeo.

Gabaix, X., and A. Landier (2008): "Why Has CEO Pay Increased So Much?," *Quarterly Journal of Economics*, 123, 49–100.

Gale, D., and L. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, 9–15.

Galichon, A. (2016): *Optimal Transport Methods in Economics*. Princeton University Press.

Galichon, A., and B. Salanié (2015): "Cupid's invisible hand: Social surplus and identification in matching models," mimeo.

Garicano, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of political economy*, 108(5), 874–904.

Golosov, M., P. Maziero, and G. Menzio (2012): "Taxation and Redistribution of Residual Income Inequality," University of Pennsylvania mimeo.

Gretsky, N. E., J. M. Ostroy, and W. R. Zame (1992): "The Nonatomic Assignment Model," *Economic Theory*, 2(1), 103–127.

Grossman, G. M., and E. Helpman (2014): "Growth, trade, and inequality," Discussion paper, National Bureau of Economic Research.

Grossman, G. M., E. Helpman, and P. Kircher (2017): "Matching, Sorting, and the Distributional Effects of International Trade," *Journal of Political Economy*, 125(1), 224–264.

Grossman, G. M., and G. Maggi (2000): "Diversity and Trade," *American Economic Review*, 90(5), 1255–1275.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85(1), 29–65.

Helpman, E., O. Itskhoki, and S. Redding (2010): "Inequality and unemployment in a global economy," *Econometrica*, 78(4), 1239–1283.

Holmström, B., and P. Milgrom (1987): "Aggregation and Linearity in the Provision of Incentives," *Econometrica*, 55, 303–328.

Jovanovic, B. (1979): "Job matching and the theory of turnover," *The Journal of Political Economy*, pp. 972–990.

Kaas, L., and P. Kircher (2014): "Learning and Recombination," New York University.

Kantorovich, L. V. (1948): "On a Problem of Monge," *Journal of Mathematical Sciences*, 133, 1383–1383.

Kelso, Alexander S, J., and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50(6), 1483–1504.

Koopmans, T., and M. Beckmann (1957): "Assignment Problems and the Location of Economic Activities," *Econometrica*, 25(1), 53–76.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 108(3), 551–75.

Legros, P., and A. Newman (2007): "Beauty is a Beast, Frog is a Prince: Assortative Matching with Nontransferabilities," *Econometrica*, 75, 1073–1102.

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

Lindenlaub, I., and F. Postel-Vinay (2016): "Multidimensional Sorting Under Random Search," *Manuscript, University College London*.

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

Lise, J., and F. Postel-Vinay (2016): "Multidimensional skills, sorting, and human capital accumulation," University College London mimeo.

Lopes de Melo, R. (2017): "Firm wage differentials and labor market sorting: Reconciling theory and evidence," *Journal of Political Economy*, forthcoming.

Lucas, R. E. (1978): "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 9(2), 508–523.

Marshak, J., and R. Radner (1972): *Economic Theory of Teams*. Yale University Press, New Haven.

Mas, A., and E. Moretti (2009): "Peers at work," *The American economic review*, 99(1), 112–145.

McNamara, J., and E. Collins (1990): "The Job-Search Problem as an Emplyer-Candidate Game," *Journal of Applied Probability*, 28, 815–827.

Monge, G. (1781): *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale.

Morgan, P. (1996): "Two-Sided Search and Matching," SUNY-Buffalo Working Paper.

Nelson, P. (1970): "Information and consumer behavior," *Journal of political economy*, 78(2), 311–329.

Ohnsorge, F., and D. Trefler (2007): "Sorting it out: International trade with heterogeneous workers," *Journal of political Economy*, 115(5), 868–892.

Pinheiro, R., and J. Eeckhout (2014): "Diverse Organizations and the Competition for Talent," *International Economic Review*, 55(3), 625–664.

Radner, R. (1962): "Team Decision Problems," *The Annals of Mathematical Statistics*, 33, 857–881.

Rogerson, R., R. Shimer, and R. Wright (2005): "Search-Theoretic Models Of The Labor Market: A Surveyv," *Journal of Economic Literature*, 43, 959–988.

Roth, A. E., and M. Sotomayor (1990): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, Cambridge.

Schulhofer-Wohl, S. (2006): "Negative assortative matching of risk-averse agents with transferable expected utility," *Economics Letters*, 92(3), 383–388.

Serfes, K. (2005): "Risk sharing vs. incentives: Contract design under two-sided heterogeneity," *Economics Letters*, 88(3), 343–349.

Shapley, L., and M. Shubik (1972): "The Assignment Game I: The Core," *International Journal of Game Theory*, pp. 111–130.

Shimer, R., and L. Smith (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

Smith, E. (1999): "Search, Concave Production, and Optimal Firm Size," *Review of Economic Dynamics*, 2(2), 456–471.

Smith, L. (2006): "Assortative Matching and Search," *Journal of Political Economy*, 114(6), 1124–1146.

——— (2011): "Frictional Matching Models," *Annual Review of Economics*, 3, 319–338.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2015): "Firming up inequality," Discussion paper, National Bureau of Economic Research.

Sorkin, I. (2015): "Ranking firms using revealed preference," Manuscript, University of Michigan October.

Terviö, M. (2008): "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 98(3), 642–668.

Teulings, C. N., and P. A. Gautier (2004): "The right man for the job," *The Review of Economic Studies*, 71(2), 553–580.

Villani, C. (2009): *Optimal Transport: Old and New.* Springer, Berlin.

Vlachos, J., E. Lindqvist, and C. Hakanson (2015): "Firms and skills: the evolution of worker sorting!," Discussion paper, Stockholm.