Search and Matching Models in Microeconomics

Hector Chade* Jan Eeckhout† Lones Smith‡

February 2, 2015

*Arizona State University, Department of Economics.
†University College London, Department of Economics, and Barcelona GSE-UPF.
‡University of Wisconsin - Madison, Department of Economics.

Contents

1 Introduction

Economics is built on the Walrasian supply and demand cornerstone, with trade anonymously guided by a fictitious impartial auctioneer. This survey article explores the literature that has largely emerged in the last quarter century on decentralized matching models with and without frictions. Matching models enrich the Walrasian paradigm, capturing person-specific goods and relationships. The frictional matching literature replaces the auctioneer's gavel by a mixture of dynamic choice and chance. It thus impedes the invisible hand with costs or imperfect information.

Of the two thrusts in the matching literature — matching for production or relationships, and for trade — we will focus on the former. In this research thread of the assignment and matching literature, a dominant theme is assortative matching — loosely, "the best matched with the best". For firms devote significant resources to hiring the right employee; the government spends large sums on unemployment benefits to provide incentives for workers to search for the right jobs; people invest great time resources searching for the right mates — recently even pursuing online dating markets to find more and better partners; and house buyers generally hire agents to help find their ideal property matching their tastes.

Even in the Walrasian setting with centralized trade, frictionless matching of heterogeneous agents makes explicit the sorting patterns between agents. The theoretical literature on frictionless matching has largely pursued two main lines of thought. In one, match payoffs are nontransferable, and equilibrium (stability) requires checking pairwise double coincidence of wants. This work began with the path-breaking math article Gale and Shapley (1962) that developed an intuitive algorithm for deducing stable matchings; this is the cornerstone of the large centralized matching literature. Meanwhile, a parallel model allowing transferable payoffs emerged, closer in spirit to market economics, in which a welfare theorem held. This social planner's problem for the matching literature dates back to the early work by Monge and Kantorovic on the *mass transportation problem*, and to Koopmans and Beckmann (1957) who introduced a pricing system to solve the problem. This literature saw its fruition in Shapley and Shubik (1972). Whereas Gale and Shapley allowed heterogeneous preferences, the seminal marriage model paper by Becker (1973) assumed common ordinal preferences over partners. He found that matching was assortative when the match payoff function was supermodular. This literature was naturally drawn to this pivotal sorting question in a variety of economic contexts like marriage markets, labor markets, housing markets, industrial organization, and international trade.

Concurrent with the matching literature, the economics of search theory was developing. Motivated by the failure of the law of one price, Stigler (1961) had formulated the first search optimization in economics. This has since proven useful for understanding wage formation and unemployment in the labor market. Search in that sense offers a tool to formalize decentralized trade. In many settings, the Walrasian assumptions on price setting are too strong. For example, agents do not see all prices of houses transacted, and even if they did, in the presence of information frictions or because of differences in private valuations, they would need costly inspection.

Until the 1990s, the matching and search theory literatures largely proceeded in isolation.[1] But the development of frictional matching models begun in the early 1990s has sparked renewed interest in both the frictionless matching and search paradigms. This has been driven by the importance of heterogeneity and sorting in many economic environments where search frictions are significant. Since then, the two literatures have been bed-fellows. For search frictions create equilibrium feedback between types who would otherwise remain unmatched. For instance, when low productivity jobs are filled by high ability workers, this affects the labor market prospects of the low ability workers. This has been a major technical theme that has emerged.

Pursuant to this merger of the search and matching literatures, a reduced form approach to search theory developed. Rather than explicitly model the dynamic matching process, an approach known as directed search developed. One approach here captured market clearing failures in two sided matching models in the spirit of Gale and Shapley (1962), by explicitly modeling the queues that form. Buyers arrive at sellers, and the queue length acts like a price, as it does at an amusement park. First partners in trade meet, then they negotiate the price. Another approach, directed search, instead explicitly models the stockouts that emerge — students apply for slots at colleges, and are generally rejected. In essence, they are told that no slot is available. Directed search often also exploits the role of prices: sellers post prices first, upon which buyers make their purchase decisions, taking into account that there are matching frictions.

We offer a self-contained review of this literature. We introduce the benchmark matching models without frictions. Motivated by some unrealistic implications, we then explore the search models that have emerged that best address these failings. We finally assemble these pieces, fleshing out matching models beset by search and information frictions. We try to present each unit in as teachable a fashion as possible, by focusing on the main analytic idea sometimes by way of example, and then fleshing out some salient economic applications. For in recent years, several features of the matching model have been used in applied settings. Examples without frictions include hierarchies, international trade, finance, CEO compensation, FDI and development[2]. Matching models with frictions have served to analyze numerous aspects of unemployment in the presence of sorting, such as mismatch, the transmission risk, and the impact of macro economic fluctuations.[3] Throughout this review we refer further to some of these papers, and highlight open frontier research agendas as a roadmap for future work.

[1] Sattinger (1993) nicely surveyed the matching models that were standard in labor market applications until the 1990s. Search and information frictions play only a minor role in that survey. The large literature that we cover in this paper illustrates how much it has progressed in the last twenty five years.

[2] See amongst many others, Garicano (2000), Sørensen (2007), Antràs, Garicano, and Rossi-Hansberg (2006), Grossman, Helpman, and Kircher (2013), Grossman and Maggi (2000), Tervio (2008), Gabaix and Landier (2008), Guadalupe, Rappoport, Salanie, and Thomas (2014) and Ackerberg and Botticini (2002).

[3] See for example Lise, Meghir, and Robin (2013), Lamadon (2014), Robin and Lise (2013)

This survey tells the story of how two key economic literatures, one in optimization and another in equilibrium, merged to create a cohesive equilibrium story of frictional markets.

2 Frictionless Matching and Assignment

To study how search and/or information frictions shape matching outcomes in economic environments, the first building block we need is a model that can serve as a benchmark when frictions are not present. In this section, we analyze the 'frictionless' matching models that have proved useful in many interesting economic applications.[4]

Many important problems can be analyzed in a unified way as a problem of pairwise matching or assignment of two sides/sets consisting of heterogeneous elements (individuals or goods). This can be accomplished by a benevolent planner or can take place in a decentralized setting where there is competition for agents or objects. Examples abound: sorting men and women into marriages, assigning workers to firms, or locations to plants, buyers to sellers, countries to goods, etc. In all these cases, one of the distinctive features is that agents or objects on each side have different characteristics and are indivisible.

An important modeling choice in this framework is how payoffs are shared within a match. Two polar choices are transferable utility (TU), where agents can freely transfer payoffs between them at each moment, and strict non-transferable utility (strict NTU), where the division of the match surplus is exogenously given and preferences over mates can be fully expressed in ordinal terms. A blend of both cases is NTU, where payoffs are neither fully transferable nor exogenously given. In the rest of the section we provide a detailed analysis of the TU and NTU paradigms, as well as several economic applications. Given our interest in sorting, we only discuss the strict NTU case results that shed light on sorting patters, leaving aside many interesting issues in this framework that are extensively covered in the book by Roth and Sotomayor (1990).

2.1 The Theory of Frictionless Matching with Transferable Utility

The insights below encapsulate the message of a trio of seminal papers: Koopmans and Beckmann (1957), Shapley and Shubik (1972), and Becker (1973).[5] The first analyzed the matching problem between plants and locations and derived the properties of the optimal assignment and competitive equilibrium as solutions to a linear programming problem and its dual. The second one used as a metaphor the assignment of buyers and sellers in a market for heterogeneous houses, and

[4]We abuse the terminology slightly by calling it frictionless, since matching with non-transferable utility can emerge due to some friction that prevents utility from being fully transferable. What we mean is that there are no search frictions (agents can observe all potential partners) and no incomplete information about partners (agents can observe their characteristics).

[5]In the mathematics literature, the Monge-Kantorovich optimal transport problem subsumes many frictionless matching problems. For an authoritative treatise of this problem, see Villani (2009).

provided solid game theoretic foundations to the problem, deriving the optimal assignment, core allocations, and competitive equilibrium in a unified way.[6] None of these papers focus on sorting patterns. It was Becker (1973) who, in a marriage context, provided the fundamental insight about complementarities of partners' characteristics in the match payoff function and the resulting positive or negative assortative matching (PAM or NAM) in the optimal/equilibrium assignment of men and women. His analysis remains a cornerstone of matching theory, and has also become important in empirical work on the subject, since it provides the theory with empirical content.

A. The Basic Model. We will derive the main insights using a simple instance of the matching model with TU (i.e, an assignment game, using Shapley and Shubik (1972) terminology), leaving extensions for later. For definiteness, we cast the problem in terms of a marriage market, but it will be obvious that other applications follow by a simple reinterpretation of the two sides of the market. There are N women and N men. Each man i has a characteristic $x_i \in [0,1]$ and each woman j has a characteristic $y_j \in [0,1]$; for simplicity, we assume that $x_1 < x_2 < \cdots < x_N$ and $y_1 < y_2 < \cdots < y_N$, so that we can identify each agent with his or her type. If man x_i marries woman y_j, then they produce a positive output $f(x_i, y_j)$. We make the innocuous assumption that single agents produce zero output. Crucially, agents' preferences are linear in money (TU), and thus partners can freely divide the match output produced using transfers.[7]

In this setting, we can ask the following questions: What is the optimal matching of men and women? Under what conditions does this assignment exhibit PAM or NAM? Is this allocation in the core of the assignment game? Can it be decentralized as a Walrasian equilibrium? We will provide detailed answers to each of these questions below.

B. The Optimal Assignment Problem. Start with the planner's problem. Since utility is transferable, efficiency demands that an optimal matching maximize the sum of all match outputs.[8] Formally, the optimal matching is the solution to the following maximization problem:

$$\max_{\pi} \sum_{i=1}^{N} f(x_i, y_{\pi(i)}), \tag{1}$$

where the maximization is over all possible permutations $\pi : \{1, 2, ..., N\} \to \{1, 2, ..., N\}$. A well-known result in rearrangement inequalities (e.g., see Vince (1990) and the Becker-Brock result in the appendix of Becker (1973)) shows that the identity permutation $\pi(i) = i$ for all i solves

[6] An excellent source for these results in both the finite and continuous case is Gretsky, Ostroy, and Zame (1999).

[7] Actually, the model allows for transfers to other pairs, but one can easily show that they are not used in the core or competitive equilibrium allocations.

[8] If this were not the case, there would be a rematching of some of the agents that would increase the total size of the pie to be distributed, contradicting optimality.

problem (1) if f is supermodular on $[0,1]^2$.[9][10] This condition is not only *sufficient* but also *necessary* if we want the result to hold for all distributions of types for men and women. In short, PAM optimal if and only if f is supermodular, that is, when men's and women's types are complements in the match output function. In this case, the planner will pair the woman with the best characteristic with the man with the best characteristic, the second best woman with the second best man, and so on.

It is easy to see why supermodularity is sufficient for PAM. Under any other assignment, there would be two women, i and i', $i' > i$, matched with men j and j', respectively, with $j > j'$. The total output of these couples is $f(x_i, y_j) + f(x_{i'}, y_{j'})$, which is lower than $f(x_i, y_{j'}) + f(x_{i'}, y_j)$ by supermodularity. Hence, the planner can increase total output by rematching them in a la PAM.

A similar argument reveals that the reverse permutation $\pi(i) = N - i + 1$ solves the problem if and only if f is submodular in (x, y). Thus, NAM is optimal when types are substitutes in production. In this case, the best woman is paired with the man with the lowest type, the second best woman with the man with the second lowest man, and so on.

A noteworthy feature of PAM and NAM is that they emerge under a property imposed *only* on the match output function f, independently of the distribution of men's and women's types.

An alternative formulation of the optimal assignment problem that will be useful below is the following linear programming formulation:

$$\max_{\alpha} \sum_{i=1}^{N} \sum_{j=1}^{N} f(x_i, y_j)\alpha_{ij} \tag{2}$$

subject to

$$\sum_{j=1}^{N} \alpha_{ij} \leq 1 \ \ \forall i, \quad \sum_{i=1}^{N} \alpha_{ij} \leq 1 \ \ \forall j, \quad \alpha_{ij} \geq 0 \ \ \forall i, j. \tag{3}$$

Notice that α_{ij} is not restricted to be 0 or 1, so the problem permits fractional assignment of men and women. Koopmans and Beckmann (1957) and Shapley and Shubik (1972), however, showed that there is always an optimal solution such that $\alpha_{ij} \in \{0, 1\}$. If f is supermodular, then $\alpha_{ij} = 1$ when $i = j$, and PAM ensues; if not, then as before one can find a profitable rematching. A similar analysis holds for NAM.

[9] A real-valued function z on a lattice $X \subseteq \mathbb{R}^n$ (e.g., $[0,1]^2$) is supermodular if $z(x' \vee x'') + z(x' \wedge x'') \geq z(x') + z(x'')$ for all x' and x'' in X, where $x' \vee x'' = \max\{x', x''\}$ and $x' \wedge x'' = \min\{x', x''\}$. If f is continuously differentiable of order two, then this is equivalent to $\partial^2 z(x)/\partial x_i \partial x_j \geq 0$ for all $i \neq j$. The function is submodular if $z(x' \vee x'') + z(x' \wedge x'') \leq z(x') + z(x'')$ for all x' and x'' in X, and this is equivalent to $\partial^2 z(x)/\partial x_i \partial x_j \leq 0$ for all $i \neq j$ if f is twice continuously differentiable. These concepts are strict if the inequalities are strict.

[10] This is a nonlinear generalization of an inequality for products of vectors in Hardy, Littlewood, and Polya (1952). In the discrete case considered, it states (see Vince (1990)) that if $z_1, ..., z_n$ are real-valued functions on an interval I, then $\sum_i z_i(b_{n-i+1}) \leq \sum_i z_i(b_\pi(i)) \leq \sum_i z_i(b_i)$ for all sequences $b_1 \leq b_2 \leq ... \leq b_n$ and all π if and only if $z_{i+1} - z_i$ is increasing on I for $1 \leq i < n$.

C. Core, Stability, and Walrasian Equilibrium. Instead of the planner's problem, we could envision men and women competing for partners in the *assignment game*, where they can bid for each other and sign contracts specifying how to divide the match output. To motivate the analysis, assume f is strictly supermodular, and thus the optimal assignment is PAM. Let $i > i'$ and $j > j'$; by strict supermodularity, $f(x_i, y_j) + f(x_{i'}, y_{j'}) > f(x_i, y_{j'}) + f(x_{i'}, y_j)$. Rearrange this inequality as $f(x_i, y_j) - f(x_{i'}, y_j) > f(x_i, y_{j'}) - f(x_{i'}, y_{j'})$; this says that the willingness to pay for the better woman x_i is higher for the better man y_j than for $y_{j'}$. So in the competition for partners j can outbid j' in the quest for i. Thus, any 'stable' outcome of the assignment game exhibits PAM. Alternatively, we could explore the performance of a competive market where agents from each side take 'partners' prices' as given. We will see that the same logic reveals that any Walrasian equilibrium of this market delivers PAM.[11]

Let u_i, $i = 1, 2, ..., N$ and v_j, $i = 1, 2, ..., N$ be the multipliers associated with respective constraints in (3). Then the dual of problem (2) is[12]

$$\min_{u,v} \sum_{i=1}^{N} u_i + \sum_{j=1}^{N} v_j \tag{4}$$

$$\text{s.t.} \quad u_i + v_j \geq f(x_i, y_j) \quad u_i \geq 0 \quad v_j \geq 0 \tag{5}$$

From the Duality Theorem of Linear Programming, the value to this problem is the same as that of problem (2)–(3) (and thus of (1)). Moreover, the α_{ij}'s of problem (2) are the multipliers of problem (4). It follows that if (α, u, v) solves (2)–(3) and (4)–(5), then (a) $\sum_{i=1}^{N} \sum_{j=1}^{N} f(x_i, x_j)\alpha_{ij} = \sum_{i=1}^{N} u_i + \sum_{j=1}^{N} v_j$; (b) for each pair (i, j) such that $\alpha_{ij} = 1 > 0$, constraint (5) binds, namely $u_i + v_j = f(x_i, y_j)$; and (c) for each pair (i, j) such that $\alpha_{ij} = 0$, we have $u_i + v_j \geq f(x_i, y_j)$.

We have thus found that the triple (α, u, v) optimally matches the two populations and provides a division of the match output between partners that exhausts the output. That triple is also a *stable matching* of the assignment game, for any pair not matched to each other cannot block the assignment in a profitable way given (c), since the sum of the utilities they obtain at their current matches more than exhaust the match output that would ensue if they rematch. Moreover, (c) also implies that no coalition of men and women can improve upon (α, u, v): hence, the solution of the dual problem characterizes the *core* of the assignment game.

It is now straightforward to decentralize the optimal/core matching as a Walrasian equilibrium. Let u_i, $i = 1, 2, ..., N$ be the 'price' of man i, and assume women take these prices as given. Then the problem of woman j is to choose the man i that maximizes $f(x_i, y_j) - u_i$. Now, by construction of the core allocation, notice that $v_j = f(x_i, y_j) - u_i$ if $\alpha_{ij} = 1$, that is, v_j is the level of utility that woman j obtains in the core allocation. Also, for any other woman i', so that $\alpha_{i'j} = 0$, we

[11] Analogous results hold for f strictly submodular and NAM.

[12] See Chvatal (1983), chapter 5 for a derivation and for the proof of the Duality Theorem of Linear Programming.

have that $v_j \geq f(x_{i'}, y_j) - u_{i'}$, or $f(x_i, y_j) - u_i \geq f(x_{i'}, y_j) - u_{i'}$. Hence, when facing prices u_i, $i = 1, 2, ..., N$, woman j will choose *exactly* the same partner as in the core allocation. One can do the same analysis from the men's perspective. Thus, the optimal matching can be decentralized as the outcome of a Walrasian equilibrium of the marriage market. Notice that the price of a man in this market depends only on his type, and not on the type of the woman he matches with. This is because these prices are formally equivalent to the utility that each man obtains in the core allocation. And since they are the multipliers of constraints (3), each can be interpreted as the shadow value of moving a man from the pool of singles to the matching market. Hence the dependence only on the man's type.

These are the main results under TU contained in the seminal papers listed above. As mentioned, we have made some simpifying assumptions to derive the results in the cleanest way. But it is immediate to relax some of them. For example, if we assume an unequal number of men and women, the only difference is that some agents on the large side of the market will remain single. Similarly, if we assume that agents can produce some output as singles, the only difference is that now a couple distributes the match surplus, defined as the match output minus the sum of the outputs they can produced as singles, and some agents may remain single at the optimal matching. We also assume that there is only one agent of each type. Allowing for an arbitrary discrete distribution of types on each side is easy to accommodate: e.g., one needs first to define a mapping between agents and their characteristics, and then PAM matches agents in decreasing order of their types, respecting the number of agents of each type in the population, until the populations are exhausted. A more interesting extension that is used in applications is that of a continuum of agents and continuously distributed characeristics. We now turn to this extension.

D. The Large Market Case. As is standard in economics, the continuum of agents idealization not only provides solid foundations for price-taking behavior, but also affords the use of calculus in the derivation of equilibria and their properties. We will illustrate this convenient feature below with some important economic applications of the frictionless matching paradigm.

Assume then a continuum of men and women, with each population having unit measure.[13] Each male is endowed with a type $x \in [0, 1]$ drawn from a cumulative distribution function (cdf) G, assumed for simplicity to be strictly increasing and continuously differentiable, with positive density g. Similarly, each woman has a type $y \in [0, 1]$, with H and h, respectively. We can define a (pure) matching as a function $\mu : [0, 1] \to [0, 1]$ that is 'measure preserving,' a property that ensures that the matching market clears.[14] For instance, under PAM μ satisfies $G(x) = H(\mu(x))$ for all x, or $\mu(x) = H^{-1}(G(x))$, which is strictly increasing with $\mu'(x) = g(x)/h(\mu(x)) > 0$. Under NAM, $G(x) = 1 - H(\mu(x))$ for all x, and thus $\mu(x) = H^{-1}(1 - G(x))$, with $\mu'(x) =$

[13] As before, the extension to unequal population sizes is immediate.

[14] Measure preserving means that for any measurable set A in $[0, 1]$, the measure of the set of types B in $[0, 1]$ matched with types in A under μ, must equal the measure of A, or $\int_A dF = \int_B dH$.

$-g(x)/h(\mu(x)) < 0$. Thus, the slope of the matching function reflects the relative 'number' of men and women of each type that are matched together.

As before, if a man with type x marries a woman with type y, then the match output is $f(x, y)$, which we assume twice continously differentiable for simplicity. A continuous version of the rearrangement inequality we used above (e.g., see Lorentz (1953) and Crowe, Zweibel, and Rosenbloom (1986)) shows that PAM is optimal if and only if f is supermodular, and NAM is optimal if and only if f is submodular. Also, the Shapley-Shubik linear programming derivation of the optimal assignment, core allocations, and Walrasian equilibrium extends to the continuous model (Gretsky, Ostroy, and Zame (1992) and Gretsky, Ostroy, and Zame (1999)).

Instead of going over these technical results, we will focus on the derivation of the Walrasian equilibrium, and deduce the sorting pattern that ensues under super or submodularity assumptions of f. In the process, we will draw a simple connection between the matching model and a basic monotone comparative statics result (Milgrom and Shannon (1994), Topkis (1998)).

Consider a woman of type y facing a 'price function' u for a men, so that $u(x)$ is the price or wage of a man of type x. Her problem is to choose the x that maximizes her payoff:

$$\max_{x\in[0,1]} f(x, y) - u(x). \tag{6}$$

Now, if f is strictly supermodular (i.e., $f_{xy} > 0$), then the objective function satisfies the strict single crossing property in $(x; y)$.[15] Hence, in *any* solution to this problem women with higher y choose men with higher x.[16] Hence, if a Walrasian equilibrium exists, it *must* exhibit PAM. This provides an alternative view of the sufficiency of supermodularity for PAM.[17] From the measure preserving or market clearing property above, the only candidate for an equilibrium matching is $y = \mu(x) = H^{-1}(G(x))$ or $x = \mu^{-1}(y)$. For this to be the optimal choice for x, it must satisfy the first-order condition (one can show that the second-order condition is satisfied as well)

$$u'(x) = f_x(x, \mu(x)),$$

from which we obtain that

$$u(x) = u_0 + \int_0^x f_x\left(s, \mu(s)\right) ds, \tag{7}$$

where u_0 is a constant of integration that parameterizes the sharing of the match output. Hence, if

[15] A function $z : X \times [\underline{t}, \bar{t}] \to \mathbb{R}$, X a lattice, satisfies the strict single crossing property in (x, t) if for all $x'' > x'$ and $t'' > t'$, $z(x'', t') - z(x', t') \geq 0$ implies $z(x'', t'') - z(x', t'') > 0$.

[16] Apply Theorem 4' in Milgrom and Shannon (1994) (quasi-supermodularity in y trivially holds in this problem).

[17] In a CEO-firm assignment application, Tervio (2008) derives the Walrasian equilibrium of the model in a similar way, and points out the relationship with the incentive compatibility conditions in screening problems. What lies at a more basic level is the monotone comparative statics result alluded to above, since both the problem of each agent in a matching setting and the incentive compatibility problem are parameterized optimization problems (in one case by an agent's observable type, and in the other case by an agent's privately known type).

f is strictly supermodular, then (μ, u) given by $\mu(x) = H^{-1}(G(x))$ and (7) constitute a Walrasian equilibrium and exhibits PAM. Clearly, each female y in equilibrium obtains $v(y) = f(\mu^{-1}(y), y) - u(\mu^{-1}(y))$. A similar analysis can be done for f strictly submodular and NAM.

2.2 Applications of Frictionless Matching with Transferable Utility

A. The O-Ring Production Function. A well-cited application of Becker's marriage model is Kremer (1993). This explores a "weakest link" production model that very naturally generates the supermodularity that Becker assumes. There are n tasks, each of them performed by a worker. Each worker i has a characteristic $x_i \in [\underline{x}, 1]$, $0 < \underline{x} < 1$, drawn from a continuous density g. This characteristic represents the probability with which the worker performs the task successfully. Unless the n workers succeed in their tasks, the product is worthless. That is, the expected output of a firm is $nB \prod_{i=1}^{n} x_i$, where B is the output per worker if all the tasks are performed successfully. There is a large number of workers of measure one and a large number of identical firms that hire n workers each in a competitive labor market. Firms take the wage function $w : [\underline{x}, 1] \to \mathbb{R}_+$ as given. For simplicity, we assume that labor is the only factor of production.

Notice the following features of this matching model. First, although there are two sides, firms and workers, only one side is heterogeneous. Second, each firm hires multiple heterogeneous workers, so the nontrivial matching problem is the combination of workers in each firm.

Firms choose the skill of each worker x_i to maximize expected profits:

$$\max_{\{x_i\}_{i=1}^{n}} nB \prod_{i=1}^{n} x_i - \sum_{i=1}^{n} w(x_i).$$

Notice that $\partial^2 nB \prod_{i=1}^{n} x_i / \partial x_j \partial x_k > 0$ for all j, k. Hence, expected output is strictly supermodular in $(x_1, x_2, ..., x_n)$ and the equilibrium exhibits PAM, which in this case means that all the workers employed by any given firm have the same x. The first order condition for each x_i evaluated at $x_i = x$ for all i yields $w'(x) = nBx^{n-1}$, and thus $w(x) = Bx^n + w_0$, where w_0 is a constant of integration, pinned down by the zero profit condition for all the firms, so $w_0 = 0$. In short, the Walrasian equilibrium of the model consists of each firm hiring workers of the same skill x and paying them according to the wage function $w(x) = Bx^n$, which divides the (expected) output of the firm equally among its workers.

Kremer (1993) goes on to show how the model sheds light on several stylized facts, including that firms hire workers of different skills and produce different quality products, and the fact that there is positive correlation among wages of workers in different occupations within a firm.

B. CEO-Firm Assignment Model. Tervio (2008) and Gabaix and Landier (2008) develop a matching model of firm size and CEO talent, and calibrate it using US data to analyze CEO pay.

The model has a continuum of CEOs and firms, each of measure one, with talent x distributed according to G and firm size y distributed according to H. Each side is 'ranked' in the unit interval as follows: a CEO with talent x is given rank (or quantile) $i \in [0,1]$ such that $i = \overline{G}(x) = 1-G(x)$, so there is a smooth relationship between i and x denoted by $X(i)$, with $X'(i) < 0$. The closer i is to zero the higher its rank. Similarly, $j = \overline{H}(y)$ and $Y(j)$, with $Y'(j) < 0$. If CEO i is matched with firm j, then the firm's revenue is $CY^d(j)X(i)$, $C > 0$, $d > 0$, and the CEO gets paid $w(i)$.

In a Walrasian equilibrium, firm j maximizes $CY^d(j)X(i) - w(i)$ over $i \in [0,1]$. Since the objective function satisfies the strict single crossing property, the equilibrium exhibits PAM with $i = j$ for all j. The first order condition yields $w'(i) = CY^d(j)X'(i)$, and hence (recall that higher i means less talent and thus less pay since $X'(i) < 0$, so the lowest wage is $w(1)$)

$$w(i) = \int_i^1 CY^d(s)X'(s)ds + w(1).$$

This wage function along with PAM constitute the Walrasian equilibrium of the model.

To calibrate the model, Gabaix and Landier (2008) assume that H is Pareto with parameter $1/\tau$, so $\overline{H}(y) = (\underline{y}/y)^{1/\tau}$, from which we deduce $Y(j) = \underline{y}j^{-\tau}$. For talent, they use Extreme Value Theory to posit that the 'spacing function' $X'(i)$ satisfies $X'(i) = -Ki^{\nu-1}$ for some constants K and ν, which holds for many standard distributions.

These assumptions yield the following closed form solution for the equilibrium wage function:

$$w(i) = \frac{CK\underline{y}^d}{\tau d - \nu}\left(i^{-(\tau d-\nu)} - 1\right) + w(1).$$

Focusing on the case with $\tau d > \nu$, Gabaix and Landier (2008) (see also Tervio (2008)) calibrate the model and analyze several features of CEO pay and its increase in recent years in the US. In particular, they show that the model exhibits a 'superstar' property (Rosen (1981)): small differences in talent can have drastic impact in pay, especially at the top (i.e., for CEOs whose rank is close to zero).[18] Also, the increase in size of large firms in recent years can account for a large fraction of the increase in CEO pay.

C. Matching Principals and Agents. Most predictions in contract theory, say with moral hazard, are based on the principal-agent model, where one principal hires an agent to perform a task, but the agent's actions are unobservable and the contract is based on a stochastic signal (e.g., output) correlated with those actions. By construction, this useful paradigm abstracts from the fact that the interaction of the principal and the agent could be the outcome of a matching market where heterogeneous principals and agents match. Relaxing this assumption can lead to some insightful modifications of standard predictions. In a nice empirical contribution, Ackerberg

[18] This follows since $w(i)$ is strictly decreasing and convex, with $w''(i)$ going to infinity as i goes to zero, as one can check by differentiation. Thus, CEOs matched with large firms receive increasingly larger pay near the top.

and Botticini (2002) convincingly argued that accounting for endogenous matching is important in testing contract theory, since it can bias many of the relevant coefficients. Using data from Renaissance Tuscany, they find strong evidence for matching between landlords with crops of different riskiness and tenants with different levels of wealth (proxy for risk aversion), which affects the contract form used (share contracts or fixed rent contracts).

Serfes (2005) provides a tractable illustration of the effects of matching heterogeneous principals and agents under moral hazard. He restricts attention to linear contracts and constant absolute risk aversion (CARA) utility function (i.e., using the standard Holmstrom and Milgrom (1987) justification), and this turns the model into a matching problem with TU.[19]

The characteristic x of a principal is the variance of his output, while the characteristic y of the agent is her coefficient of absolute risk aversion. The agent's preferences are given by a CARA utility function, $1 - e^{-y(I-(ke^2/2))}$, where I is income and the second term in the parenthesis is the disutility of exerting effort $e \geq 0$. When x matches y, the agent then exerts effort $e \geq 0$ and generates output $q = e + \varepsilon$, where $\varepsilon \sim N(0, x)$. The contract that rewards the agent is linear in output: $I(q) = I_0 + bq$, where I_0 is a base wage and b the contract's incentive power. It is well known (see Holmstrom and Milgrom (1987), Section 5) that the optimal contract sets $b = 1/(1 + kyx)$ and that the principal's expected profit is $f(x, y) = 1/(2k(1 + kyx))$.[20]

Notice that a prediction from the model is that b is decreasing in x, that is, there is a negative relationship between risk and incentives. The evidence on this prediction is weak: the data exhibits either a positive or an insignificant relationship. Embedding the principal-agent problem in a matching model can account for this finding if NAM turns out to be optimal: for in this case, if principals with high x are matched with agents with low y, that is $y = \mu(x)$ and μ is strictly decreasing, then $b = 1/(1 + k\mu(x)x)$ could increase in x if matching is endogenous.

Can NAM be the optimal/equilibrium matching? Differentiation yields that $f_{xy} < 0$ if and only if $yx < 1/k$, and this holds for all values $x \in [\underline{x}, \overline{x}]$ and $y \in [\underline{y}, \overline{y}]$ if $\overline{y}\,\overline{x} < 1/k$. So NAM emerges if this condition holds.[21] And if $\mu(x)x$ decreases in x (which depends on G and H), then b increases in x, as the evidence shows.

D. Matching in Large Firms. Especially in the labor market, realism demands models where firms can hire mulitple workers. The one-to-one matching paradigm therefore misses an important and realistic feature of actual labor markets. In principle, one could reinterpret it as many-to-one version where a firm consisting of a series of independent jobs that do not affect any other job's productivity. Often however, there are complementarities between jobs.

The importance of many-to-one matching was clear to Gale and Shapley (1962) and their

[19]We will see in the next section that, without this assumption, the presence of moral hazard leads to NTU.

[20]The expression for I_0 is not important since, due to the absence of wealth effects in CARA, its role is to meet the agent's reservation utility once b has been optimally set. Notice also that we have set f equal to the principal's profit; this is because, by the TU property, both differ by just a constant.

[21]Similarly, PAM ensues if $\underline{y}\underline{x} > 1/k$.

college admissions problem. A more relevant set up for labor market applications is one with a large number of firms and transfers (wages). The O-ring technology in Kremer (1993) is a particular case, where firms employ a fixed number n of workers. A general framework for analyzing large firms is the model in Kelso and Crawford (1982), which develops a many-to-one matching model of the firm with a general technology and no restrictions on the number or the type of workers a firm can hire. Existence of equilibrium, however, is not guaranteed as the following simple example illustrates. Consider two workers $y = 1, 2$ and two firms $x = 1, 2$. Assume that $f(1,1) = 4$, $f(1,2) = 1$, and with some abuse of notation, if firm 1 hires both workers output is $f(1, \{1,2\}) = 10$; similarly, $f(2,1) = 8$, $f(2,2) = 5$, and $f(2, \{1,2\}) = 9$. Output is zero without a worker. Then all possible allocations are $\{\{1,2\}, \{\emptyset\}\}$ (firm 1 hires both workers and firm 2 is empty), $\{\{2\}, \{1\}\}, \{\{\emptyset\}, \{1,2\}\}$, and $\{\{1\}, \{2\}\}$. Each of those allocations will be blocked. Consider for example the first allocation, which generates a total output of 10 to firm 1 and zero to firm 2. Any wage to worker 1 must be at least 8, which is what firm 2 is willing to offer, the output it produces when hiring worker 1. Likewise, the wage to worker 2 must be at least 5. But this total wage bill exceeds the output of 10 in firm 1. Hence this allocation is not stable. Similar arguments go through for the other allocations.

The source of the non-existence is that complementarities between workers are too strong. Firm 1 generates more output from hiring both workers than from each worker individually. In other words, when worker one is in firm 1 and worker two is hired, worker 1's productivity goes up and she can command a higher wage than when being hired alone. Kelso and Crawford (1982) derive a sufficient condition for existence, the *gross substitutes* condition. In words, this condition asserts that if wages increase for some types of workers, then the firm will not drop from its (optimally chosen) labor force those workers whose wages did not increase. A simple example that satisfies this condition is a production function that is additively separable in worker types. Then the productivity of a worker does not depend on the composition of her co-workers. It is important to point out that the gross substitutes condition is only a sufficient condition, but it is one that does not leave much room for complementarities (see Hatfield and Milgrom (2005)). Kelso and Crawford (1982) also offer an algorithm to find the equilibrium allocation and corresponding wages. It is a variation on the Gale-Shapley strict NTU deferred acceptance algorithm. Because there are transfers, it uses the features of an ascending bid auction to determine the wages.[22]

Constrained by the gross substitutes condition, it has proved difficult to develop models of large firms where both size and composition and endogenous, and there exist complementarities.

Without the composition issue, Lucas (1978) focuses on the *intensive margin* decision of how many workers to hire, a decision that he calls the management's span of control. More productive management hires a larger work force, thus spanning its control further. This model has been used to shed light on the distribution of firm size.

[22]Hatfield and Milgrom (2005) generalize the model of Kelso and Crawford (1982) and analyze package auctions.

In reality, however, management at a firm faces a more complex trade-off. It decides the number of people as well as the worker types to hire. This *extensive margin* decision about workers' composition is precisely the focus of matching models. The problem is that in the canonical matching model of the labor market (Becker (1973)) a firm consists of exactly one job.

Eeckhout and Kircher (2012) develop a model that incorporates these two margins (size and composition) while avoiding the gross substitutes condition. They assume that match output is given by a production function $F(x, y, l, r)$, where x denotes the firm type, y the worker type, l the labor force, and r the resources the firm has available. The function F exhibits constant returns in l, r, so we can write $F(x, y, l, r) = rF(x, y, \theta, 1)$ where $\theta = \frac{l}{r}$. Normalizing the resources to 1 and setting $F(x, y, \theta, 1) \equiv f(x, y, \theta)$, allow to write the problem of a firm x as follows:

$$\max_{y,\theta} f(x, y, \theta) - \theta w(y).$$

From the first and second-order conditions of the problem, Eeckhout and Kircher (2012) showed that the condition for PAM is

$$f_{xy} f_{\theta\theta} - f_{y\theta} f_{xy} + \frac{f_y f_{\theta x}}{\theta} \leq 0. \tag{8}$$

Since $F(x, y, \theta, 1) = f(x, y, \theta)$, this inequality can be written as $F_{xy} F_{\theta\theta} - F_{x\theta}(F_{y\theta}/\theta)$. The homogeneity of degree one of F implies that $-F_{lr} = \theta F_{ll}$ and $F_y = \theta F_{yl} + F_{yr}$ (if F is homogeneous of degree one in l, r then so is F_y). Hence, we obtain the following equivalent condition for PAM:

$$F_{xy} F_{lr} \geq F_{xl} F_{yr}. \tag{9}$$

Reversing the inequality yields the condition for NAM. We prove this result in an elementary way in the next section, using a different approach than Eeckhout and Kircher (2012).

For an intuition of (9), consider the left side: it involves the standard complementarity between the characteristics of the workers and the firm, increased by the complementarities between the total quantity of labor and resources of the firm. Regarding the right side, it is the product between the complementarity between firm type and size of the labor force, and that between worker type and firm resources. If this product is negative, e.g. $F_{yr} < 0$, then the condition for PAM is weaker than in the one-to-one model, while the opposite is true if it is positive.

2.3 Frictionless Matching with Non-Transferable Utility

A. Background. When partners cannot transfer utility one for one, we say that there is non-transferable utility (NTU). Many economic environments of interest fall in this category, such as risk sharing problems or matching problems where moral hazard is present.

One extreme case of NTU, henceforth called the *strict* NTU case, is when partners cannot transfer utility at all, e.g., the output $f(x, y)$ from a match between x and y is divided according to some fixed sharing rule. Actually, matching models without transfers have been extensively studied since Gale and Shapley (1962).[23] In their two-sided matching model, preferences are formulated as ordinal rankings over the partners on the other side of the market, and an equilibrium is defined in terms of stability. A matching is *stable* if there exists no *blocking pair* of agents, preferring to be matched to each other rather than to their respective partners in the candidate stable allocation. The fundamental result is that a stable matching exists. The existence proof is constructive by means of the *Deferred Acceptance (DA) Algorithm*. One side of the market, say women, can make offers to their preferred man, who temporarily retains his best choice. Each woman who has not been retained then makes an offer to her second most preferred man. Again, men retain their most preferred women, possibly dropping an earlier retention. This process continues until no more women are left who prefer any man over remaining single. This yields existence of a stable matching, and it highlights also that there may be multiple stable matchings.

Using a cardinal representation of preferences in terms of $f(x, y)$, Becker (1973) showed that if $f_x > 0$ and $f_y > 0$, i.e., each agent prefers a partner with a higher type, then PAM emerges under strict NTU. Too see this, suppose, as Becker (1973), that x obtains a positive share $\alpha(x)$ of the match output in *all* possible marriages, and y obtains a positive share $\beta(y)$. These shares need not add up to one if x and y marry, may be due to enforcement costs or public goods (see Becker (1973)). Consider now two men and two women, with types $x' > x''$ and $y' > y''$, who are matched in a NAM way, that is, x' with y'' and x'' with y'. Then x' and y' can block the matching and offer to rematch, since $\alpha(x')f(x', y') > \alpha(x')f(x', y'')$ and $\beta(y')f(x', y') > \beta(y')f(x'', y')$, due to the monotonicity of f in each argument.[24] So while Becker did not cite Gale and Shapley (1962), he intuitively grasped their pairwise stability notion.

Recently, Legros and Newman (2010) have shown that monotonicity in partner's type is not necessary for PAM. Indeed, the necessary and sufficient condition for PAM in this setting is that preferences exhibit 'co-ranking': given any two men and women, either the top man and woman prefer each other, or the bottom man and woman do. In the example above, this means that it has to hold for any two pairs of men and women, and this condition is consistent with f not being increasing in partner's type (see Legros and Newman (2010) for an example).

Thus far we have a clear understanding of assortative matching in both the TU and the strict NTU cases. In each case there are necessary and sufficient conditions solely on f that deliver PAM or NAM. What about the important intermediate case where there are transfers but preferences are not linear in money? Another way to formulate the question is as follows. Notice that the Pareto frontier of payoffs achievable by a pair of matched agents is linear in the TU case, and collapses to a point in the strict NTU case. What about typical intermediate cases where agents

[23] See Roth and Sotomayor (1990) for a thorough overview.

[24] A similar analysis can be done for NAM if one of the derivatives is positive and the other one is negative.

can transfer utility but not at a constant rate and thus the Pareto frontier is decreasing but neither linear nor a single point? Legros and Newman (2007) address this case. We now provide a detailed summary of their main insights and several illustrative applications.

B. The Basic Model. There are two populations, men and women, who differ in their types $x \in [0,1]$ and $y \in [0,1]$. For simplicity, we assume that they have the same size, which can be finite or a continuum, and their autarchy payoff is normalized to zero. We identify agents by their types, so that agents of the same type behave alike and receive the same payoff in equilibrium.[25]

The utility frontier of each pair of agents x and y is $\phi(x, y, v)$, which is the maximum utility that x can generate when matched with y who receives utility v. Since in equilibrium no agent receives less than their autarchy payoff, the maximum that x can ever obtain when matched with y is $\phi(x, y, 0)$. We assume that $\phi(x, y, 0) > 0$ and that $\phi(x, y, \cdot)$ is strictly decreasing in v whenever $\phi(x, y, v) > 0$. This Pareto frontier is the primitive of interest.[26] The function $\psi(y, x, \cdot)$ is the 'inverse' of $\phi(x, y, \cdot)$, and $\psi(y, x, u)$ is the maximum utility of y when matched with x who receives utility u. So for $u \in [0, \phi(x, y, 0)]$, we have that $\psi(y, x, u)$ solves $\phi(x, y, \psi(y, x, u)) = u$.

The equilibrium concept is the core of this assignment game. That is, it consists of a matching function μ and payoff functions u and v, so that $u(x)$ is the utility of x and $v(y)$ is the utility of y, such that they satisfy: (i) feasibility with respect to μ, so for all x, $u(x) \leq \phi(x, y, v(\mu(x)))$ and $v(\mu(x))) \leq \psi(y, x, 0)$; and ($ii$) stability of μ with respect to u and v, so that there is no pair of agents with x and y such that $v > v(y)$ and $\phi(x, y, v) > u(x)$. Note that the TU model is subsumed by this formulation, for under TU $\phi(x, y, v) = f(x, y) - v$.

C. Generalized Increasing Difference. As a starting point, consider the earlier TU case and assume that f is supermodular in (x, y), so that PAM is optimal. As seen, in equilibrium if $x > x'$ and $y > y'$ then y will outbid y' in the competition for x, or $f(x, y) - f(x', y) \geq f(x, y') - f(x', y')$, which can be rewritten as

$$f(x, y) - [f(x', y) - u] \geq f(x, y') - [f(x', y') - u], \tag{10}$$

where u is the utility that x' obtains. Notice that (10), which represents y's willingness to pay for x in comparison to that of type y', holds for *any* level of utility u. This increasing difference condition implies that the higher the type of an agent is, the higher is the matching partner's type, and thus the optimal matching exhibits PAM.

Inspired by (10), type y's willingness to pay for type x is higher than the willingness to pay

[25] This 'equal-treatment' property is an easy equilibrium result in Legros and Newman (2007) (see p. 1078), but for simplicity we are going to assume it here. Notice that there is little loss of generality in doing this in the finite case when each agent on each side has a different type, as we assumed in our basic TU model, or when the distributions are the same, or when there is a continuum of agents with continuously distributed characteristics. In the finite case with general discrete distributions, some agents of a given type may match with different types on the other side if the number of each type differ. Then one needs to specify the mapping from agents to characteristics.

[26] In the applications below, the frontier is derived from assumptions on technology and preferences of the agents.

of y' in the NTU case when x' obtains u if and only if

$$\phi(x, y, \psi(y, x', u)) \geq \phi(x, y', \psi(y', x', u)). \tag{11}$$

Why is this y's willingness to pay for x compared to the willingness to pay of type y'? First, notice that (11) reduces to (10) in the TU case, for the left side is

$$\phi(x, y, \psi(y, x', u)) = f(x, y) - \psi(y, x', u) = f(x, y) - [f(x', y) - u],$$

and similarly for the right side. So (11) is a generalization of (10), and Legros and Newman (2007) appropriately named it the *generalized increasing difference* condition (GID). And if the inequality is reversed, they called it *generalized decreasing difference* condition (GDD). Second, the left side of (11) asserts that if x' obtains u when matched with y, and thus y obtains $\psi(y, x', u)$, then if y fixes this level of utility $\psi(y, x', u)$ but matches with x instead, then the maximum utility that x can enjoy is given by the left side of (11). The right side has a similar interpretation but for y', asserting that y's willingness to pay for x exceeds that of type y' when x' receives utility u.

The main result in Legros and Newman (2007) is that if GID holds whenever $x > x'$, $y > y'$, and $u \in [0, \phi(x', y, 0)]$ (i.e., u 'feasible' in a match between x' and y), then either there is PAM in equilibrium, or any equilibrium is payoff equivalent to one with PAM (so PAM is without loss of generality). Similarly, if GDD holds whenever $x > x'$, $y > y'$, and $u \in [0, \phi(x', y', 0)]$, then either there is NAM in equilibrium, or any equilibrium is payoff equivalent to an equilibrium with NAM. These conditions are necessary if we want the result to hold for any distributions of agents' types. That is, GID and GDD are *sufficient* and *necessary* conditions for PAM and NAM, respectively. Moreover, they are *distribution-free* conditions.

This is a powerful result that subsumes TU as a special case, and significantly enlarges the set of economic applications for which we can pin down sorting patterns.

Instead of going through their proof, we will shed light on their result by using some simple notions of monotone comparative statics. (The paragraphs that follow are unavoidably more technical than the rest, and the reader may prefer to skip them and go directly to the differential version of GID we derive below, which is the one we will use in the applications.)

Consider the GID condition (11): since the Pareto frontier is strictly decreasing in its third argument, notice that (11) continues to hold if we replace, on the right side, u by $u' \leq u$, so that

$$\phi(x, y, \psi(y, x', u)) \geq \phi(x, y', \psi(y', x', u')). \tag{12}$$

Now let us interpret the meaning of condition (12) from the perspective of men with types x and x'. The condition asserts that if x' prefers to match with y and enjoy a level of utility u rather than to match with y' and obtain $u' \leq u$, then so does x, for he obtains $\phi(x, y, \psi(y, x', u))$ when matched with y (given that he must give y a level of utility $v = \psi(y, x', u)$), which is bigger than the utility $\phi(x, y', \psi(y', x', u'))$ that he obtains when matched with y' (given that he must deliver $v' = \psi(y', x', u')$ to her). Formally, all this can be written succintly as

$$\phi(x', y, v) \ \geq (>) \ \phi(x', y', v') \Rightarrow \phi(x, y, v) \ \geq (>) \ \phi(x, y', v'). \tag{13}$$

So if the set of pairs of (y, v) is a lattice, which can be guaranteed by imposing on this set the lexicographic order (i.e., $(y, v) \geq (y', v')$ if either $y > y'$ or $y = y'$ and $v > v'$), then (13) says that ϕ satisfies the *single crossing property* in $((y, v); x)$ (see Milgrom and Shannon (1994)).

We can now shed new light on why PAM ensues under GID. For by Theorem 4 in Milgrom and Shannon (1994), the solution to the problem $\max_{y,v} \phi(x, y, v)$, if unique, will have the property that the choice of y will be increasing in x.[27]

We can go further and provide simpler ways to check for GID. Assume from now on that ϕ is twice continuously differentiable. We know from Theorem 3 in Milgrom and Shannon (1994) that (13) is *equivalent* to the *Spence-Mirrlees* condition, which requires that that the marginal rate of substitution between y and v, i.e., $-\phi_y/\phi_v$, be increasing in x, or $\phi_{xy}\phi_v - \phi_{xv}\phi_y \leq 0$. Graphically, the indifference curves in the (y, v) space cross only once as a function of x.

Assume $\phi_y > 0$ from now on, so that ϕ is increasing in partner's type (and similarly for ψ). Then the equivalence between the single crossing property and the Spence-Mirrlees condition implies that GID is equivalent to (recall that $\phi_v < 0$)

$$\phi_{xy}(x, y, v) \geq \frac{\phi_y(x, y, v)}{\phi_v(x, y, v)} \phi_{xv}(x, y, v), \tag{14}$$

which provides a tractable differential version of the Legros and Newman (2007) condition, which we call, for clarity, SM-GID (SM for Spence-Mirrlees). And if we reverse the sign, it is easy to verify that we obtain an analogous SM-GDD.

Inequality (14) reveals the crucial tension in the NTU case between the two flavors of complementarities in this setting: (i) complementarity in partners' types in production (ϕ_{xy}), and (ii) complementarity in own type and partner's utility (ϕ_{xv}), which reflects whether it becomes easier or more difficult to transfer utility to a partner as one's type increases. In the TU case, the latter is zero and only the former matters, and (14) collapses to the well-known supermodularity condition $\phi_{xy} \geq 0$, which is equal to $f_{xy} \geq 0$ in the TU case.

As in Legros and Newman (2007), the weak inequality (14) does not preclude the existence of other equilibria that do not exhibit PAM but are payoff equivalent to it. But if either the single

[27] The following comments are in order. First, the theorem requires also that $\phi(x, \cdot, \cdot)$ be quasi-supermodularity in (y, v), and this holds under the lexicographic order. Second, there could be multiple solutions, in which case the set of solutions would be increasing in x in the so-called strong set order. And although there could be selections from the optimal correspondence that are *not* increasing, there is always a monotone increasing one, which resembles the assertion in Legros and Newman (2007) that any equilibrium is payoff equivalent to a PAM one.

crossing property or (14) holds strictly, which is a stronger than GID (see Edlin and Shannon (1998)), then *any* equilibrium will exhibit PAM (an analogous remark holds for NAM).

In short, simple notions of monotone comparative statics not only shed light on the meaning of GID but they also provide a simple way to obtain a differential version of this condition that will prove easy to check in the applications below.

We immediately obtain from (14) the *sufficient* conditions for positive and negative sorting: PAM ensues if both $\phi_{xy} \geq 0$ and $\phi_{xv} \geq 0$, and NAM if both are nonpositive. Intuitively, if there are complementarities in types, this is a force towards PAM. Now this can be offset by a force against PAM if it becomes more difficult to transfer utility to a partner as one's type increases. But if it becomes easier to transfer utility to a partner, then the two effects reinforce each other and PAM obtains. An analogous argument holds for NAM.

We have assumed that the Pareto frontier is strictly decreasing. Hence, the extreme case of strict NTU is not covered by the analysis. Legros and Newman (2010), however, showed that the co-ranking condition is equivalent to the GID condition. Therefore, GID is a single encompassing condition that subsumes both the TU and strict NTU cases.

2.4 Applications of Frictionless Matching with Nontransferable Utility

A. Matching Principals and Agents. Let us revisit this application but now without the CARA and linear contract assumptions (see Legros and Newman (2010), Section 5.2). The characteristic y of an agent is her level of initial wealth, which affects her attitudes toward risk. Assume that the agents's utility function is additively separable in income and disutility of effort, given by $V(y+I)-e$, where V is strictly increasing and strictly concave, with coefficient of absolute risk aversion $-V''/V'$ that decreases in y (decreasing absolute risk aversion), and e is the disutility of exerting effort. We will denote by Z the inverse function of V, i.e., $Z = V^{-1}$. We assume $e \in \{0,1\}$. Agent's effort is unobservable; what is observable is output $q \in \{\overline{q}, \underline{q}\}$, $\overline{q} > \underline{q}$. If the agent exerts $e = 0$, then $q = \underline{q}$ with probability one. If she exerts $e = 1$, then the probability of $\overline{q}$ is x, the characteristic of the principal. That is, principals differ in the riskiness of the distribution of output under high effort. We assume that $\overline{q} - \underline{q}$ is large enough so that principals always want to implement $e = 1$.

A contract consists of a pair $(\underline{I}, \overline{I})$ of contingent wages, where the first is the compensation when $q = \underline{q}$ and the second when $q = \overline{q}$. When principal x matches with y and y's reservation utility is v, the contracting problem is $\phi(x,y,v) = \max_{\underline{I},\overline{I}}\ x\left(\overline{q} - \overline{I}\right) + (1-x)\left(\underline{q} - \underline{I}\right)$ subject to the participation constraint $xV(y+\overline{I}) + (1-x)V(y+\underline{I}) - 1 \geq v$ and the incentive constraint $xV(y+\overline{I}) + (1-x)V(y+\underline{I}) - 1 \geq V(y+\underline{I})$. It is easy to show that both constraints must bind at the optimum: if either is slack, then wages can be suitably reduced in a way that strictly improves the principal's expected profit. Solving the two binding constraints in the two unknowns yields

$\overline{I} = Z(v + x^{-1})$ and $\underline{I} = Z(v)$. Therefore

$$\phi(x, y, v) = x\left(\overline{q} - Z\left(v + \frac{1}{x}\right)\right) + (1 - x)\left(\underline{q} - Z(v)\right). \tag{15}$$

Notice that ϕ is nonlinear in v (it is actually strictly concave in v), thus confirming the NTU property of the model. Differentiating (15) twice yields $\phi_{xy}(x, y, v) = 0$ and

$$\phi_{xv}(x, y, v) = \frac{1}{x}Z''\left(v + \frac{1}{x}\right) - Z'\left(v + \frac{1}{x}\right) + Z'(v). \tag{16}$$

Assume that Z' is convex; then the first term in (16) dominates the last two and hence $\phi_{xv} \geq 0$. Since $\phi_{xy} = 0$, we obtain PAM (see (14)).[28] That is, agents with high initial wealth and thus low levels of risk aversion are matched with principals whose distribution of output is safer (higher success probability). What utility functions V lead to a convex Z'? Since $Z' = 1/V'$, straightforward algebra shows that Z' is convex if and only if V satisfies the following condition: the coefficient of absolute risk prudence $-V'''/V''$ is less than or equal to three times the coefficient of absolute risk aversion $-V''/V'$. This condition is satisfied for many standard utility functions, including CARA and $V(I) = I^d$ with $d \geq 0.5$, and it commonly emerges in principal-agent models with moral hazard when trying to understand the impact of wealth effects. If instead Z' is concave, then $\phi_{xv}(x, y, v) \leq 0$ and there is NAM. This holds for, say, $V(I) = I^d$ with $d < 0.5$.

B. Marriage and Risk Sharing. Consider a marriage market where men and women with different levels of wealth (proxy for risk aversion here) marry to share risk. Let x be the wealth of a man and y the wealth of a woman. If x marries y, then they share the risk embedded in a gamble $q \in [0, 1]$ that has a continuous distribution Γ (this can be interpreted as stochastic household income). The utility function of men is $\log(1 + x + I)$ and of women is $\log(1 + y + I)$, where I is income. Efficient risk sharing solves the following problem:[29]

$$\phi(x, y, v) = \max_{I(\cdot)} \int_0^1 \log(1 + y + q - I(q))d\Gamma(q) \qquad \text{s.t.} \qquad \int_0^1 \log(1 + x + I(q))d\Gamma(q) \geq v, \tag{17}$$

where we have used the property that $I(q)$ is the share for the man and $q - I(q)$ goes to the woman for each realization of q. Intuitively, the constraint binds at the optimum.

Let ζ be the Lagrange multiplier associated with (17). Maximizing pointwise, we obtain that, for each q, marginal utilities are equalized, or $(1 + y + q - I(q))^{-1} = \zeta(1 + x + I(q))^{-1}$, and hence $I(q) = (-(1 + x) + (1 + y + q)\zeta)/(1 + \zeta)$. Inserting this equation into (17) yields an expression

[28] As Legros and Newman (2010) point out, the sorting pattern comes solely from the NTU property of the problem, as there are no complementarities between x and y. This is an artifact of assuming that v does not depend on y. If instead one assumes $v(y)$, then $\phi_{xy}(x, y, v) = v'(y)\phi_{xv}(x, y, v)$, but the same result emerges.

[29] This is a general version of an example in Legros and Newman (2010), Section 5.1 (see also Chiappori and Reny (2005) and Schulhofer-Wohl (2006)). We use our differential version of their condition to readily check for NAM.

for ζ. Substituting that into $I(q)$, and then into the objective function yields:

$$\phi(x,y,v) = \log\left(1 - e^{v - \int_0^1 \log(2+x+y+q)d\Gamma(q)}\right) + \int_0^1 \log(2+x+y+q)d\Gamma(q) \tag{18}$$

We will now show that the optimal sorting entails NAM, i.e., wealthy women with low levels of risk aversion marry poor men who are highly risk averse. Intuitively, the more risk averse has a higher demand for insurance and thus is willing to pay more for it. Hence, a highly risk averse man can outbid a less risk averse one for a wealthy woman.

To prove this result, one needs to verify that $\phi_{xy} < (\phi_y/\phi_v)\phi_{xv}$. It is easy to check that $\phi_{xy} < 0$ while $\phi_{xv} > 0$, so the quick sufficient condition for PAM or NAM does not hold. But some algebra reveals that $\phi_{xy} < (\phi_y/\phi_v)\phi_{xv}$. Thus, the optimal sorting pattern is NAM.

C. Matching in Large Firms. As seen, Eeckhout and Kircher (2012) analyze a matching problem where firms hire multiple workers. We presented their condition for PAM (NAM) without a formal proof. Although that paper has TU, it turns out that their main sorting condition can be derived in a simple way using the SM-GID (SM-GDD) condition.

Recall that in their set up, a firm of type x solves the following problem:

$$\max_{y,\theta} f(x,y,\theta) - \theta w(y) = \max_{y}\left[\max_{\theta} f(x,y,\theta) - \theta w\right],$$

where we have omitted the argument of w since it will be irrelevant for the sorting condition (recall that GID holds for all feasible levels of the utility a partner obtains). The first-order condition of the inner maximization problem is $f_\theta(x,y,\theta) = w$, from which we obtain the maximizer $\theta(x,y,w)$, and it is easy to verify that $\theta_x = -f_{\theta x}/f_{\theta\theta}$. Now we can write the maximization over y as

$$\max_{y} f(x,y,\theta(x,y,w)) - \theta(x,y,w)w \equiv \max_{y} \phi(x,y,w),$$

which looks like an NTU problem where the ability to transfer w (which plays the role of v in our derivation of GID) to a partner depends on the types via θ. Thus, we can apply SM-GID (SM-GDD) to pin down the condition for PAM (NAM).

Using $f_\theta = w$ and $\theta_x = -f_{\theta x}/f_{\theta\theta}$, straighforward differentiation yields the following expressions (we omit the arguments of the functions to simplify the notation)

$$\phi_y = f_y, \quad \phi_w = -\theta, \quad \phi_{xy} = f_{xy} - \frac{f_{y\theta}f_{x\theta}}{f_{\theta\theta}}, \quad \text{and} \quad \phi_{xw} = \frac{f_{x\theta}}{f_{\theta\theta}}.$$

Simple algebra confirms that $\phi_{xy} - (\phi_y/\phi_w)\phi_{wx} \geq 0$ if and only if (recall that $f_{\theta\theta} < 0$)

$$f_{xy}f_{\theta\theta} - f_{y\theta}f_{xy} + \frac{f_y f_{\theta x}}{\theta} \leq 0,$$

which is *exactly* the condition for PAM we stated before without proof. Reversing the inequality yields the necessary and sufficient condition for NAM, thus completing the proof of the result.

The way we have proved this result provides an illustration of a *trick* that we will use below in another application, this time with search frictions. The trick consists of finding in a matching problem the NTU problem 'embedded' in it, and then apply Legros and Newman (2007) to derive the condition for PAM or NAM. It highlights the importance of their condition for analyzing sorting patterns in a variety of matching problems. Alternatively, one can interpret it as yet another illustration of the power of monotone comparative statics techniques.

D. Additional Topics. There are many applications involving one population of agents, such as collaboration between partners in law firms, team members in consulting or sports, gay marriage, etc. Although existence of stable matchings might be problematic (see Roth and Sotomayor (1990)), in many instances one can divide the pool of agents into two sides and match them as if they come from a two-sided problem. An issue with this approach is that such a division depends on the equilibrium allocation one is looking for. Kremer and Maskin (1996) point out another feature that is unique to the one-sided model. The production function may be such that identical agents perform different tasks with different productivities. In a model of managers and workers (drawn from the same population) with a technology that exhibits complementarities between manager and worker types but also imperfect substitability (a manager who is twice as productive as a worker does not produce exactly twice the amount of output), they show that changes in the technology can lead to a sharp increase in wage inequality as well as in segregation of workers by skill. Moreover, the optimal matching need not be PAM.

In the models we covered, the output of each pair depends only on the characteristics of the pair. In many economic applications, it is natural to expect interdependence of the match values across matches. Consider for example a setting where firms first match with workers and then in a subsequent stage engage in a strategic interaction a la Cournot in the output market. When hiring heterogeneous workers who reduce the cost of production, the workforce composition of each firm will not only directly affect the firm's output, but it will also affect the competitor's output through the competition in the output market. Other examples include direct externalities such as knowledge spillovers, copying, patent races, etc. Despite its economic relevance, the analysis of matching problems with externalities has received little attention in the matching literature. The seminal matching paper by Koopmans and Beckmann (1957) describes a variation of the matching problem between locations and plants in the presence of transportation costs, and show that competitive equilibria may fail to exist. More recently, matching with externalities has been analyzed by Sasaki and Toda (1996), who provide notions of stability with externalities in both the TU and strict NTU case, and Chade and Eeckhout (2014), who analyze the role of externalities on the optimal and equilibrium matching patterns. Not surprisingly, in the presence of externalities the equilibrium outcome need not coincide with the planner's optimal allocation.

In a simple setting with two-person teams, they show that there is positive sorting when there are complementarities in worker types producing private benefits to the firm. This is inefficient and the planner prefers negative sorting when the differential external effect from negative sorting is sufficiently strong to outweigh this force towards positive sorting.

In this survey, we focus on matching problems where agents have one-dimensional characteristics. An interesting extension is to allow agents to have multidimensional characteristics, as in a marriage market where men and women are heterogeneous in education, income, attractiveness, etc., or in a labor market where firms have many heterogeneous tasks and workers differ across several skill dimensions. The mutldimensional problem is technically hard, and little is known about conditions for sorting. A state-of-the-art reference in existence and uniqueness of equilibrium in the multidimensional matching model with TU is Chiappori, McCann, and Nesheim (2010) (see also the references cited therein), who use powerful tools from the optimal transport literature to also shed light on the relationship between matching models and hedonic pricing models. Regarding matching patterns, Lindenlaub (2014) has recently developed a notion of PAM for multidimensional problems and, using a model where workers have both manual and cognitive skills and firms have jobs demanding both skills, she has applied it to the analysis of technological change in the US in recent years and its effects on the wage distribution.

Most of the matching literature assumes that agents characteristics are primitives of the model. A relatively small literature has explored ex-ante investments followed by a matching stage. A standard question addressed in these papers is whether the prospect of a better match induces agents to invest ex-ante, thus mitigating the hold up problem. Another one is to understand how imperfections at the matching stage combines with the investment problem ex-ante and generate inefficiencies. One of the earliest references on this topic that analyzes these issues is Cole, Mailath, and Postlewaite (2001), and a recent one is Noldeke and Samuelson (2014).

3 Foundations of Search Theory

3.1 Why Search Frictions?

The matching paradigm as we have described it has some unrealistic economic properties. In a standard Walrasian model, either a small change in the supply of some endowment, or a slight increase in the number of individuals with some preference, has only a small effect on the price. But in a pairwise matching setting, slight imbalances have tremendous effects. For instance, consider a marriage market with homogeneous men and women. Then with slightly more men than women, all matching rents go to women, but with slightly women than men, all the rents go to men. Or assume a world with heterogeneous people available for matches. There can be implausibly discontinuous matching allocations. For suppose that match payoffs are $f(x, y) = 1 + \varepsilon xy$, with $|\varepsilon| > 0$ is incredibly small. Depending on whether $\varepsilon \gtrless 0$, we either have positive or negative assortative matching, respectively. Also, the Walrasian auctioneer fiction is a far worse description of the matching market, since the items for sale are all quite unique. Admittedly, organizing this as a market is the challenge facing online matching services.

To fully understand how search frictions distort equilibrium market outcomes, we first need to learn some of the decision theory basic tools of search theory. Below, we explore single-agent search theory, and illustrate it with some economic applications. Search in microeconomic models is usually modeled in two ways: sequential, where the decision maker samples options over time until she decides to stop, and non-sequential or simultaneous, in which all the options are sampled at once and then the best one is chosen. In all cases, search theory explores the role of option value in choices: *where* to search or *how long* to search. Just as in finance theory, an option value is increasing in the riskiness of the choices, since the extreme events yield the surplus.

3.2 Simultaneous Search

The seminal paper by Stigler (1961) started the literature on search in economics. His is a model of simultaneous search, where a consumer samples prices from a distribution, and chooses how many searches to make. Each search costs $c > 0$. Specifically, suppose you are searching for a product, and *must buy it today.* In the morning you can call many (ex-ante) identical stores, and in the afternoon, after searching through their stock, they will call back with a price quote. Upon observing the prices sampled, the consumer buys the product in question from the firm that quoted the lowest price. The optimal sample size is an easy optimization problem in one variable and, for some distributions, it can be obtained in closed form.

Let the distribution of prices be given by a non-degenerate distribution $F(p)$ on $[0,1]$. A consumer chooses a fixed sample size n to minimize the expected total cost C (expected purchase cost plus search cost) of purchasing it. With n *independent* draws, the distribution of the lowest price is $F_n(p) = 1 - [1 - F(p)]^n$. Thus, if one plans to purchase K units, the expected total outlay is

$$P(n) = K \int_0^1 p dF_n(p) = K \int_0^1 [1 - F_n(p)]dp = K \int_0^1 [1 - F(p)]^n dp$$

Observe that this expected price falls in n, but at a diminishing rate. Thus, the second-order condition is met, and the optimal sample size n^* obeys the first-order condition: $P(n^* - 1) - P(n^*) \geq c > P(n^*) - P(n^* + 1)$. Easily, a large planned purchase K raises the marginal benefit of sampling, and thus induces greater search.

Stigler's fixed sample size search is tough to motivate, as it is almost always a contrived thought experiment. But there is one big occasion in life when we make such a one-shot search experiment: applying for college. This ignores the possibility of early admission at one school, which adds an interesting dynamic wrinkle, to which we return. Since the locations of prizes are known, but their realizations are not, this may be better thought of as an information friction.

Unlike in Stigler's model, one must choose the colleges to apply to, and not simply their number. For colleges vary by admission chances and career value.

Recently, Chade and Smith (2006) has extended the simultaneous search paradigm to include ex-ante heterogeneous options. In their model, the decision maker not only chooses the number of options to sample but also the composition of the sample. Each option generates a stochastic reward. After observing the rewards of each option, the decision maker chooses the largest one. Specifically, imagine a set of colleges $\{1, 2, \ldots, N\}$, with utilities $u_1 > u_2 > \cdots > u_N$. Since presumably easy colleges to gain entry to are not as valuable career options, they assume inversely-ranked admission chances $\alpha_1 < \alpha_2 < \cdots < \alpha_N$.

Assuming that each application costs $c > 0$, it then indicates how many colleges to apply to. In principle, the optimal portfolio might require searching through 2^N possibilities. Could college students actually be solving such a fantastically complex NP-hard problem? Chade and Smith (2006) prove that a simple *Marginal Improvement Algorithm* (MIA) yields the optimal portfolio, and it only takes about N^2 steps to find the best portfolio with N schools: At stage 1, one selects the school with greatest expected value. If that value exceeds c, then put college i in the tentative portfolio. At any stage $n+1$ in the recursion, one finds the school i_{n+1} yielding the greatest marginal benefit on the portfolio constructed so far. Add that school to the tentative portfolio if the incremental value is at least the cost c. Otherwise, stop.

That this algorithm works is surprising, because the problem is a static one, and not amenable to dynamic programming. Their proof — a joint mathematical induction on the number of options and cardinality of the portfolio set — establishes that one never wishes to remove a college that was added at an earlier stage. The MIA was a new member of a broadly-defined class of "greedy algorithms," in which a sequence of locally optimal choices leads to the global optimum.

For a minimal illustrative example, assume just three colleges with payoffs $u_1 = 1$, $u_2 = 0.8$, and $u_3 = 0.6$, and admission chances $\alpha_1 = 0.5$, $\alpha_2 = 0.8$, and $\alpha_3 = 1$. The expected payoffs $z_i \equiv \alpha_i u_i$ are therefore $z_1 = 0.5$, $z_2 = 0.64$, and $z_3 = 0.6$. With an application fee $c = 0.15$, the optimal portfolio includes college 2. The marginal benefit of the portfolio $\{1, 2\}$ over $\{2\}$ is

$$MB_{1|2} = [z_1 + (1-\alpha_1)z_2] - z_2 = z_1 - \alpha_1 z_2$$

since college 2 is only relevant in the event that one is rejected at college 1. On the other hand, in pondering the marginal benefit of $\{2, 3\}$ over $\{2\}$, we note that college 2 matters whenever one is accepted there. The marginal benefit is computed therefore in a different way:

$$MB_{3|2} = [z_2 + (1-\alpha_2)z_3] - z_2 = (1-\alpha_2)z_3$$

We conclude from the MIA that college 1 belongs to the optimal portfolio, since $z_1 - \alpha_1 z_2 = 0.18 > 0.12 = (1-\alpha_2)z_3$. Finally, one does not wish to add college 3 to this portfolio since

$(1-\alpha_1)(1-\alpha_2)z_3 = 0.06 < 0.15 = c$, and thus, the optimal college portfolio is $\{1,2\}$.

Using the algorithm, Chade and Smith (2006) prove that students apply more aggressively than they would if they were unaware of how their colleges jointly interact in their portfolio. For instance, the colleges with the two highest expected payoffs are the lower-ranked colleges 2 and 3. Students must account for the interactions of their college choices. They should not be told simply to apply to their best expected options.

This is best seen as a justification for why students pursue "stretch schools". For assume a world with just college i, and a very large number of identical lower ranked colleges j, thus with a lower payoff $u_j < u_i$. Assume that colleges j has a higher expected value $\alpha_j u_j > \alpha_i u_i$, so that the MIA starts with college j. While it may well continue to add copies of college j, college i is eventually chosen by the algorithm before exhausting all of the j-colleges. Let's see how a temptation to gamble upwards emerges. The marginal benefit of adding college j copies vanishes geometrically fast, and so falls below $\alpha_i(u_i - u_j) - c > 0$, for small application costs $c > 0$. But the marginal benefit of adding college i to a portfolio of n colleges j equals

$$\alpha_i u_i - \alpha_i u_j(1-(1-\alpha_j)^n) - c > \alpha_i(u_i - u_j) - c$$

For large enough n, adding the stretch application to college i is the best course of action.

By the same token, safety schools can only be understood if success chances in the options are not independent. For instance, suppose that a common unknown shock may affect all college evaluations. If a student is unaware, e.g., that math requirements have shot up, then as insurance, he might wish to apply more broadly down in rank. This remains an interesting research avenue.

Modeling search frictions in this simultaneous way appears in some equilibrium search models of price dispersion such as Burdett and Judd (1983), directed search models with one or multiple applications and ex-ante identical firms such as Burdett, Shi, and Wright (2001a), Albrecht, Gautier, and Vroman (2006), and search problems with multiple applications and heterogeneous options such as Chade, Lewis, and Smith (2011) for college admissions and Kircher (2009) and Kircher and Galenianos (2009) for labor markets.

3.3 Sequential Search

Amusingly, while Stigler's paper introduced price search — complaining about information as the "slum dwelling in the town of economics" — his model was almost immediately abandoned. After McCall (1965) introduced sequential search to economics, the simultaneous search model was essentially ignored until Chade and Smith (2006). For in Stigler's model, if the searcher could decide sequentially on whether to continue, he does better. Indeed, he would always have available the fixed sample size commitment policy, simply by ignoring what he has seen until sampling n^* stores, and then picking the best so far. But if given the option to recall a past search, he might well wish to stop either earlier or later. McCall (1970) re-worked his 1965 model for wage search, assuming that a worker samples a wage from a distribution in each period, and decides whether to continue the search, or stop and work at that wage. The optimal strategy is fully summarized by a reservation wage $\bar{w}$ above which the worker stops searching, and below which he continues.[30] Since he is now indifferent when faced with his reservation wage, and will in the future be, his new optimality condition requires:

$$\bar{w} = \int_0^\infty \max(\bar{w}, w) dF(w) - c \quad \Rightarrow \quad c = \int_{\bar{w}}^\infty [1 - F(w)] dw \tag{19}$$

Since the problem is stationary, a wage once rejected is forever rejected. An option to return to a previously declined option is worthless. This expression admits some immediate predictions. For instance, if one interprets c as foregone unemployment benefits, then the reservation wage rises in these benefits. And by standard stochastic dominance reasoning, a mean preserving spread of the wage distribution F likewise raises the reservation wage, since the max operator is convex.

We instead pursue a richer model that has occupied a higher ground in microeconomics since it allows for ex ante heterogeneous options: the Pandora's Box of Weitzman (1979). Suppose that a decision maker faces a finite number of heterogeneous options, each represented by a unique probability distribution $F_k(w)$ over prizes. Opening box k costs $c_k > 0$, and incurs a time discounting factor $\delta_i \in (0, 1]$, due to delay. He may ultimately accept only one prize. Options are independent and the decision maker can canvass them sequentially. At any point in time, she can decide to stop the search and keep the best reward observed thus far. So an optimal strategy requires specifying the order to explore options and a stopping rule.

Uncertain options in life should be undertaken as long as one is sufficiently optimistic. Uncited in Weitzman (1979) was the earlier solution of the more complex multi-armed bandit problem in Gittins and Jones (1974) and later Gittins (1979) — the so-called *Gittins Index*.[31] This index encapsulates this optimism, and *solely* reflects the uncertainty of that option for the case when option realizations are independent. It is the fixed prize $\bar{w}_k$ that leaves the decision maker indifferent about choosing prize k, and paying to open box k, knowing that the prize awaits him if he wants it. This intuitively captures the contingent decision making afforded the decision maker. Then:

$$\bar{w}_k = \delta_i \int_0^\infty \max(\bar{w}_k, x) dF_k(x) - c_i \tag{20}$$

Solving the problem by induction and dynamic programming, Weitzman showed that the optimal selection and stopping rules were then straightforward: at each stage, the decision maker

[30] Morgan and Manning (1985) endogeneized the sample size at each stage of the search process, thereby blending sequential and simultaneous search.

[31] A multi-armed bandit is a finite action infinite horizon Bayesian experimentation problem. When the payoffs of taking each action are independent of all others, the optimal strategy is given by Gittins indices.

samples the option with the largest value, and stops when the reward observed is larger than the reservation values of all the remaining options. In a sense, one can think of McCall (1965) as a special case of Weitzman (1979) when the options are identical and there is a large number of them. The reservation wage equation (19) emerges with homogeneous options and costs, and no discounting — namely, $F_i = F$, $c_i = c$, and $\delta_i = 1$. For even though Weitzman allows recall of past options, this freedom is inessential in a stationary setting, and thus the assumption of no recall of past wages is not a real restriction.

Keeping in mind our predictions about how reservation wages change, we can see that the searcher will first explore options with lower costs, higher means, and higher variances. Weitzman gives an example where the first options explored appear to be dominated, with lower mean and higher cost. They are valuable provided they have a high enough variance.

Let's revisit the college application problem of §3.2 with college-specific application costs. Assuming that one could apply in sequence to colleges, one would optimally employ the Weitzman rule. With our binary payoff distribution, the equation (20) for the reservation value $\bar{w}_i$ of college i reduces to:

$$\bar{w}_i = (1-\alpha_i)\bar{w}_i + \alpha_i u_i - c \;\Rightarrow\; \bar{w}_i = u_i - c/\alpha_i = (z_i - c)/\alpha_i$$

So sequential decision making is governed not by the expected net gains z_i, but instead by the expected net gain divided by the probability of success. Chade and Smith (2006) prove that sequential decision making encourages individuals to act more aggressively than simultaneous choices. One should pursue less likely options first. For instance, in the example of §3.2, college $i = 2$ is the first applied to; however, the Gittins' indexes are

$$\bar{w}_1 = u_1 - c/\alpha_1 = 1 - 0.15/0.5 = 0.7 > 0.6125 = 0.8 - 0.15/0.8 = u_2 - c/\alpha_2 = \bar{w}_2$$

The economic lesson is that sequential decision making always pushes toward risk-taking behavior. Intuitively, early admission encourages more aggressive applications by students. Harvard ended early admissions in 2006 and brought it back last year. But this decision seems unwise for a university at the top of the student preference order, since a more aggressive applications strategy can only encourage weaker students to apply.

These problems have long been explored in operations research under the rubric of Optimal Stopping. These consist of dynamic problems where a decision maker chooses a time — conditional on sequentially observed random variables — at which he takes an action to maximize his expected payoff. As expected, their solution typically involves heavy use of dynamic programming tools. There are excellent references on the subject, ranging from elementary to advanced, such as DeGroot (1970), Chow, Robbins, and Siegmund (1971), Shiryaev (1978), Ross (1983), Ferguson (Undated), and Peskir and Shiryaev (2006).

Modeling search as a sequential process is standard in much of economics. For instance, it is widely used in macroeconomic models of the labor market (e.g., Shimer, Rogerson, and Wright (2005)), in the literature of matching with vanishing frictions (e.g., Osborne and Rubinstein (1990)), and in assortative matching models with search (see §4).

One might think of search theory as optimal decision making when one knows the payoff distribution, but not the realizations. But in a key extension of the sequential search problem, certainty about the payoff distribution is relaxed. In this way, we must struggle with learning the distribution too. For instance, Rothschild (1974) explored the implications of learning the distribution while searching. This has been revisited many times and in many guises, and recently by Adam (2001), and Gershkov and Modovanu (2010).

Finally, in equilibrium applications later, a continuous time search model with exponential arrivals replaces the discrete time model. Assume that the arrival rate is $\rho > 0$. Subtract $\delta_i w_k$ from both sides of (20), and think of $\delta_i = 1 - r\ dt$, and c_i as a flow search cost. Then (20) becomes

$$r\bar{w}_k = \rho \int_0^\infty \max(x - \bar{w}_k, 0) dF_k(x) - c_i \tag{21}$$

This admits the intuitive statement that the return on the value equals the sum of the dividend $-c_i$ and the expected capital gains, namely, the expected surplus of prizes x over the value $\bar{w}_k$.

3.4 Web Search

Weitzman's model introduced a possibly nonstationary model of search with heterogeneous options. Let us turn for a moment to a special case that is also one of the modern frontiers of search that we all confront, in which people meet via the world wide web — as is common in many pairwise matching markets. Varian (1999) mused on practical advice from Weitzman (1979) for offering books to a rushed consumer in an airport bookstore. The take-out message is that one should err on the side of higher variance options first. Varian obviously intended this as a parable for web search engines. So inspired, Choi and Smith (2013) specialize Wiezman's search model to understand how people behave in the most typical of modern search experiences — web search engines. Equally well, it could be used by internet matching services, like match.com. Specifically, they assume a world in which all payoffs W hail from a Gaussian distribution. Furthermore, upon receiving a keyword query, the search engine first ranks

$$W = \alpha X + \sqrt{1 - \alpha^2} Z. \tag{22}$$

Think of X and Z as the independent common value and idiosyncratic components, each standard normal random variables. After entering the keyword and reading the search engine results page, the user learns the sequence of realized common components $\{x_k\}$, where k is the web page ranking. To learn the idiosyncratic component, the user must click on the web site and read it.

The search engine accuracy α represents how effective is the search engine in using the information, possibly via cookies, collected from the users to reduce noise, and renders more predictable web searches. When $\alpha = 0$, the problem is a stationary search problem. In that case, the user employs the same cutoff for all periods, and will never use the recall option unless the last period is reached. When $\alpha = 1$, the websites are perfectly sorted, and the user will stop at the first result. In this case, the recall option is likewise unused. For intermediate $0 < \alpha < 1$, the user faces a non-stationary search problem with decreasing cutoffs, and so might well recall an earlier draw. This is intuitively the world most of us find ourselves in while searching the internet.

Even though the search problem is highly nonstationary, the options stochastically worsen as one proceeds through the list: $x_1 > x_2 > \cdots$. In this model, a user clicks on the k^{th} web site iff the best draw so far lies below $\bar{w}_k = \alpha x_k + \zeta(\alpha)$, where αx_k reflects the common component, and $\zeta(\alpha)$ measures the search optionality — namely, the net benefits of the idiosyncratic randomness. The threshold $\zeta(\alpha)$ obeys a reservation equation in the spirit of (19), conveniently invariant to the web site rank k:

$$c = \int_{\zeta(\alpha)}^{\infty} \left[1 - \Phi\left(\frac{y}{\sqrt{1-\alpha^2}} \right) \right] dy.$$

One can compute that the implied optionality measure $\zeta(\alpha)$ monotonically falls in accuracy α — as a concave and then convex function of α, with extreme values $\zeta(0) > -c = \zeta(1)$.

Choi and Smith (2013) assume that individuals can quit any search and exercise an *outside quitting option* u, or continue searching. Search engines are keenly interested in the chance that one never quits searching, so that the search engine secures a successful match. This chance increases in accuracy iff $u < \zeta(\alpha)$. In other words, there may be conflict of interest between online shopping sites (for say Amazon) and consumers. In particular, when the outside options is high or the price is low, a shopping webs site secures higher sales with a noisier search engine.

3.5 Search by Committee

We finally explore an intriguing application of search theory as it applies to the search for job candidates. Of the two seminal papers here Albrecht, Anderson, and Vroman (2010) and Compte and Jehiel (2010), we focus on the characterization results in the former [AAV], since it sheds clear insights on how partner search is affected by fixed search costs. Consider a department seeking to hire job candidates that, for simplicity, arrive sequentially, one per period. The cost of search is impatience: the payoff is discounted by $0 < \delta < 1$. Each member i of hiring committee of N members observes a random private value $W^i \in [0, 1]$, independently drawn from the cdf F. After observing her private value, each committee member casts a yea or nay vote, to hire the current candidate or continue the search. Search ends with the current option if at least $M \leq N$ members vote to hire the current candidate, and continues otherwise.

For a taste of the theory, consider first a committee of one. It votes to hire a candidate if and only if $W \geq \bar{w}$, given the reservation value $\bar{w} = \delta V(\bar{w})$, where V is a fixed point $V = TV$ of the Bellman operator $TV(w) = (1 - F(w))\mathrm{E}[W|W \geq w] + \delta V(w)$. Next, assume a size $N = 2$ committee of Mr. j and Ms. k. In the symmetric equilibrium, everyone still employs a reservation value w, and secures payoff $V_M(w)$. For $M = 1$, this value solves the Bellman recursion $T_1V_1 = V_1$, where

$$T_1V(w) \equiv (1 - F(w))\mathrm{E}[W|W \geq w] + F(w)(1 - F(w))\mathrm{E}[W|W \leq w] + F(w)^2\delta V(w)$$

For Mr. j's payoff exceeds the threshold w with chance $1-F(w)$, and search ends; or is below with chance $F(w)$, whence search ends with chance $1 - F(w)$ if Ms. k's payoff exceeds the threshold w; otherwise, search continues. AAV prove that this common threshold is a Nash equilibrium.

Since the operators are ordered $T_1(v) < T(v)$ for all v, their fixed points obey $V_1 < V$, and their reservation values $\bar{w}_1 < \bar{w}$. Committee members are individually less picky: The single agent i rejects any candidate that committee member i rejects. This reflects a *stopping externality*, since the value reflects the bad event that Ms. k votes to stop when Mr. j has a low draw.

Next assume unanimity is needed: $M = 2$. The expected value $V_2(w)$ is now a fixed point of:

$$T_2V(w) \equiv (1 - F(w))^2\mathrm{E}[W|W \geq w] + F(w)(2 - F(w))\delta V(w)$$

with reservation wage $\bar{w}_2 = \delta V_2(\bar{w}_2)$. By the same logic, committee members are less demanding than solo searchers in equilibrium — namely, $\bar{w}_2 < \bar{w}$ — now because of a *continuation externality*, in the bad event that Ms. k votes to continue when Mr. j has a high draw. Since the stopping or continuation externalities obtain on any search committee, the committee is always less choosy than the solo searcher in any symmetric equilibrium. When $M = 1$, this reservation value ordering implies that the committee concludes search faster on average, since the continuation probability is lower: $F(\bar{w}_1)^N < F(\bar{w}_1) \leq F(\bar{w})$. But when $M > 1$, there is a tradeoff — a candidate must independently pass several lower thresholds. AAV show that when $M < N$, the committee concludes search faster then a solo searcher with enough patience or impatience: for $\delta \notin (\delta_L, \delta_H)$.

In the single agent problem, mean preserving spreads are unambiguously beneficial, since one can always discard low draws. In the committee search problem, AAV show via example that mean preserving spreads can lower welfare. Intuitively, a mean preserving spread can increase either the stopping externality (by making low draws more costly), the continuation externality (by increasing the chance that another member of the committee blocks you like), or both.

AAV also consider other comparative statics, some of which turn on log-concavity, and a smart understanding of order statistics. This literature adding strategic elements to the search problem is an inviting future direction, in light of the important of collective decision-making.

4 Matching and Search

4.1 An Introduction to Sorting in Search and Matching Models

In the Walrasian matching model, it is costless for agents to find potential partners, be they women searching for men, workers searching for jobs, or buyers searching for sellers. Building on the basic matching model, we now introduce search frictions. Search frictions are important, since they can explain not only equilibrium unemployment, but also mismatch. This can explain the lack of discontinuities, as seen in §3.1, as well as important phenomena such as unemployment, price dispersion of homogeneous products, and imperfect marriage sorting.

We distinguish between random search, modeling how markets only clear with delay, and directed/competitive search, where markets clear by queues and stockouts. Some of the questions we address are: How do market frictions affect match formation and sorting in marriage and labor markets as well as in models of bilateral trade? Who matches with whom in equilibrium? The literature frontier assumes a common evaluation of agents, without a hint that beauty is in the eye of the beholder,[32] and this remains a major direction of future research.

But the quest for enriching Becker's framework to account for search frictions has received an enormous amount of attention in recent years. The earliest such papers included two heterogeneous agent search models by Bergstrom and Bagnoli (1993), which we explore in §5.2, and Smith (1992). These papers assumed NTU, and were followed up by Burdett and Coles (1997) and Smith (1997, 2006). But the proper extension of Becker's model required TU, as was later assumed in Shimer and Smith (2000). Recently, the literature has explored the implications of replacing the assumption of anonymous search by that of directed search (e.g., Eeckhout and Kircher (2010a), Shi (2001)) where agents can identify where they send their applications to find potential partners.

With random search, there are several other modeling assumptions besides TU or NTU. Firstly, it is standard to assume continuous rather than discrete time, and to posit exponential arrivals of matching opportunities. Next, for models of partner search, it is common to model search cost as a time cost. Sometimes, it is modeled as a fixed cost that is paid each time an agent meets a potential partner. With a few salient exceptions, the unisex model is assumed for simplicity.

Crucially, one must take a stand on the nature of the *search technology* (Diamond and Maskin, 1979). With anonymous search, unmatched individuals meet one another in direct proportion to their mass in the unmatched pool. But what then is the constant of proportionality? In a *linear search technology*, potential partners arrive with constant rate $\rho > 0$ (for rendezvous), independent of the mass of unmatched individuals. At the opposite extreme is the *quadratic search technology*, in which unmatched individuals face a constant arrival rate of potential partners *ρ times the mass of unmatched agents.* For simplicity, we will refer to ρ as a *meeting rate.* For intuition, this arises

[32]This assumption has already been explored in a search and trading model in Smith (1995).

if invitations to meetings arrive at fixed rate ρ to everyone, but when either party is already matched, he misses the meeting. It is called quadratic because the mass of meetings rises as the square of the mass of unmatched individuals, rather than in linear proportion.

The quadratic search technology affords a crucial analytic advantage: players are unaffected by the matching decisions of those unwilling to match with them. By contrast, in order to hold the matching rate constant with a linear search technology, new individuals that enter one's matching set displace previous individuals. The "strategic independence" of the quadratic search technology allows the equilibrium to be solved, and is absolutely essential here in a nonstationary environment. We now illustrate this with the first heterogeneous agent search model that properly tracked the demographics Smith (1992). This also leads us into the sorting results.

4.2 Sorting with Random Search and Nontransferable Utility

A. Block Assortative Matching. Assume a world where everyone is summarized by a scalar type, and posit a uniform density on these types in $[0,1]$. Initially, everyone is unmatched, and so this model can capture a newly opening pairwise matching market. Everyone periodically meets match partners according to a quadratic search technology at rate ρ. Everyone wishes to enter a permanent match, and their payoff is the expected present value of the eventual match one enters. Intuitively, individuals will start off willing to match with the highest types, but then these will gradually vanish, and the expected value will monotonically fall. The logic of §3.3 implies that the marginal type that leaves one indifferent about continuing, equals the unmatched value $\bar{w}$.

Let us enrich the logic underlying (21). Now there is an additional capital loss owing to the falling value $\bar{w}' < 0$ (suppressing the time subscripts in this nonstationary analysis). Summing the capital loss and gain, we find that the return is now $r\bar{w} = \bar{w}' + \rho \int_{\bar{w}}^{1} (x - \bar{w}) u(x) dx$, where $u(x)$ is the type-x specific unemployment rate. We see here the richness of heterogeneous agent search models, since we must keep track of the unemployment rates of all types. But we can greatly simplify the problem, and capture the evolution of this threshold simply using two state variables — the total unmatched measure mass $\bar{u}$ across all types $[0,1]$, and the average type of an unmatched agent, which we write as $\chi/\bar{u}$. Rewrite the above law of motion for the unmatched value as:[33]

$$\bar{w}' = r\bar{w} + \rho[\bar{w}\bar{u} - \chi + \bar{w}^2/2] \tag{23}$$

Since types in $[\bar{w}, 1]$ match just among themselves, and they have mass $\bar{u}$, the quadratic search technology implies that their unmatched density falls at rate $\rho[\bar{u} - \bar{w}]$. Because the mass and first

[33] Indeed, $\bar{w}' = r\bar{w} - \rho \int_{\bar{w}}^{1} (x - \bar{w}) dx$. At any time, some types are below $\bar{w}$, and so have a uniform density, while those above $\bar{w}$ have partially matched. Thus, $\bar{w}' = r\bar{w} + \rho[\bar{w}\bar{u} - \chi - \int_0^{\bar{w}} (x - \bar{w}) dx] = r\bar{w} + \rho[\bar{w}\bar{u} - \chi + \bar{w}^2/2]$.

moment of agents below $\bar{w}$ is respectively $\bar{w}$ and $\bar{w}^2/2$, the law of motion for $(\bar{u}, \chi)$ is therefore:

$$\bar{u}' = -\rho[\bar{u} - \bar{w}]^2 \quad \text{and} \quad \chi' = -\rho[\bar{u} - \bar{w}][\chi - \bar{w}^2/2]$$

All told, the equilibrium is captured by this three-dimensional state $(\bar{w}, \bar{u}, \chi)$. One can see that the threshold $\bar{w}$ eventually must vanish; Smith (1992) argues that it does so at rate $O(1/t)$.[34]

It is natural to ask about a steady-state equilibrium, in a model with entry of new agents. The earliest heterogeneous agent sorting paper McNamara and Collins (1990) seeks a stationary equilibrium, but does not make any assumptions to ensure a constant population. Smith (1992) assumed that matches dissolved, so that a balance was achieved. Later several papers made an endogenous 'cloning assumption', positing that agents who leave the matching market are replaced by clones. Finally, the third option — first executed in Smith (1997, 2006) and Burdett and Coles (1997)— insists that an exogenous inflow of new unmatched agents of each type equals the flow of matches formed. This *balanced flow* condition is most realistic but hardest to implement. In steady-state, the number of matches involving a given type of agent must equal the number of matches destroyed in which an agent of this type was a partner (if replenishment is modeled via match destruction) or equal to the number of new singles of this type that enter the market (if replenishment is modeled via newborn singles).

In the steady-state equilibrium in Smith (1992) with uniform entry flow $e > 0$, the limit of the new threshold $\bar{w}_1$ is now positive. The balanced flow condition for types above this threshold now yields $e = \rho u_1(1 - \bar{w}_1)$, whereas the steady-state condition $\bar{w}_1' = 0$ in (23) yields an optimality condition:

$$r\bar{w}_1 = \rho u_1 \int_{\bar{w}_1}^{1} (x - \bar{w}_1)dx = \rho u_1[1 - \bar{w}_1]^2/2$$

Jointly, we can solve for the pair $(u_1, \bar{w}_1)$. Likewise, the sequence of thresholds $1 > \bar{w}_1 > \bar{w}_2 > \cdots$ and associated unmatched rates $u_2, u_3, \ldots$ are computed inductively, by strategic independence.[35]

We see that the first taste of sorting models with search frictions entailed jumps: Individuals match in ranked equivalence classes. Smith (1997, 2006) called this "block segregation", and highlights its counterfactual aspects. Burdett and Coles (1997) later took it as a parable of "marriage and class" in Britain. It is a very stark form of positive assortative matching (PAM). Confronted by such discontinuities, the question then arose — under what conditions are the matching sets continuous functions of the types? But with a continuum of types, it is not even clear what we should call positive sorting. For since every type must match with a positive mass of types, it is no longer possible to assert that the percentiles of matched men and women coincide. Shimer and Smith (2000) offered a formulation of PAM, and its natural opposite *negative assortative matching* (NAM), that simultaneously applies equally well to singleton matching sets

[34]In pursuant work, Daminano, Li, and Suen (2005) develop a general theory of how matching unravels.

[35]For by analogy, we have $r\bar{w}_{k+1} = \rho u_{k+1} \int_{\bar{w}_{k+1}}^{\bar{w}_k} (x - \bar{w}_{k+1})dx = \rho u_{k+1}[\bar{w}_k - \bar{w}_{k+1}]^2/2$, as $k = 1, 2, \ldots$.

or sets of positive measure. This definition asks that the matching set as a subset of $\mathbb{R}^2$ be a lattice. In other words, if (x_1, y_2) and (x_2, y_1) are willing to match, and $x_1 < x_2$, $y_1 < y_2$, then so are (x_1, y_1) and (x_2, y_2). Intuitively, any mismatches are explained by the thickness of the matching set. So when matching sets collapse to a singleton for each x, as in Becker's marriage model, this notion reduces to an increasing or decreasing function under PAM and NAM, respectively.

This definition implies that each type x matches with types in an intervals $[a(x), b(x)]$, with both $a(x)$ and $b(x)$ weakly increasing in x. Or with NAM, the functions a and b are decreasing in x. This definition has compelling economic implications. Most easily, the distribution of partners with whom x matches is increasing in x in the sense of first order stochastic dominance; as a result, the expected value of the partner with whom x matches is increasing in x under PAM, and decreasing under NAM. This is a testable implication for the data.

B. Strict Assortative Matching. Under strict NTU, a sufficient condition for PAM in the frictionless case is that the match output function is increasing in partner's type (Becker, 1973). We now explore under what conditions PAM ensues in the strict NTU case if there are search frictions. This literature has many participants, some with containing overlapping results: McNamara and Collins (1990), Smith (1992), Smith (1997, 2006), Morgan (1996), Bloch and Ryder (1999), Burdett and Coles (1997), Chade (2001), and Eeckhout (1999).

Smith (1997, 2006) characterizes sorting with search frictions modeled as impatience; from his analysis, we shall derive the analogous condition for fixed search cost case. He assumes a continuum of types $x \in [0, 1]$ distributed according to a cdf G and a positive density g. Time is continuous with an infinite horizon. Unmatched agents search for partners, which incurs a "time cost": they discount the future at rate $r > 0$. Being unmatched yields zero flow payoff, while a type x agent enjoys flow payoff $f(x, y) > 0$. Here, f is strictly increasing in partner's type (i.e., $f_y > 0$ everywhere). No side payments are allowed (NTU). Already, Gale and Shapley (1962) predict that PAM is the unique stable matching. Is that still true in a model with frictions?

Let unmatched agents randomly meet, according to a quadratic search technology with meeting rate $\rho > 0$. When agents meet, they approve a match if both earn nonnegative surplus. Matches vanish at *match dissolution rate* $\kappa > 0$, i.e. the match lasts past time t with chance $e^{-\kappa t}$.

Each type x agent chooses an *acceptance set* $A(x) \subseteq [0, 1]$ with whom she is willing to match. In turn, x is deemed acceptable by the types in the *opportunity set* $\Omega(x) = \{y | x \in A(y)\}$. Hence, the matching set of an agent with type x is $A(x) \cap \Omega(x)$. Of course, in the TU model, the matching decision is mutual, and so $A(x) = \Omega(x)$, whereas in the NTU model, $A(x)$ should be a higher set than $\Omega(x)$ — since one's preferences invariably surpass one's opportunities. The model is in steady-state with a constant unmatched density function u, satisfying $0 \leq u(x) \leq g(x)$ for all x.

We now must set up and solve a continuum of heterogeneous but interlaced dynamic programming problems. Let $v(x)$ be type x's expected present discounted unmatched value, and $v(x|y)$ the analogous value from being matched with y. In the Bellman equation, there is no dividend (zero payoff while unmatched), and an arrival rate of a capital gain equal to the expected match surplus:

$$rv(x) = \rho \int_{\Omega(x)} \max(v(x|y) - v(x), 0)u(y)dy. \tag{24}$$

Similarly, the matched value solves $rv(x|y) = f(x, y) + \kappa[v(x) - v(x|y)]$. Naturally, $v(x|y) > v(x)$ since $f(x, y) > rv(x)$. Since $f_y > 0$, type x accepts all types in an upper set $A(x) = [0, a(x)]$. Also, the threshold partner $a(x) > 0$ acts like a reservation wage, and solves the indifference condition $f(x, a(x)) = rv(x)$. So the opportunity set $\Omega(x) = \{y|x \geq a(y)\}$ is increasing in x. Substituting the expression for $v(x|y)$ and the threshold $a(x)$ into (24) yields the explicit recursion equation:

$$rv(x) = \frac{\rho}{r + \kappa} \int_{A(x)\cap\Omega(x)} [f(x, y) - rv(x)]u(y)dy \tag{25}$$

This expression reveals how the return on the unmatched value reflects the how matches dissolve at rate κ. Finally, an equilibrium might most succinctly be described as a triple (v, a, u), such that $v(x)$ obeys (25) for the acceptance set $A(x) = [0, a(x)]$, $a(x)$ solves $f(x, a(x)) = rv(x)$ given $u(\cdot)$, and $u(x)$ obeys the balanced flow condition (26) at every type x:[36]

$$\kappa\left(g(x) - u(x)\right) = \rho u(x) \int_{A(x)\cap\Omega(x)} u(y)dy. \tag{26}$$

Smith (1997, 2006) and Chade (2001) suggest a unified approach to exploring sorting under NTU; it applies to search with impatience, but extends to the case of fixed search costs. There is PAM when $a(x)$ is weakly increasing in x, then so is its inverse $b(x)$, and thus the opportunity set $\Omega(x) = [0, b(x)]$. Graphically, matching engulfs all types between two increasing bands $[a(x), b(x)]$.

Assume for now that $\Omega(x) = [0, b(x)]$ for all x, and that $b(x)$ is weakly increasing. For the highest types, this is true since $b(x) = 1$. Since $v(x)$ is an optimal value, (25) becomes

$$rv(x) = \max_a \frac{\int_a^{b(x)} f(x, y)u(y)dy}{\psi + \int_a^{b(x)} u(y)dy}. \tag{27}$$

where $\psi = (r+\kappa)/\rho$ encapsulates search frictions in a scalar constant. Assume an interior solution. The first-order condition that determines the optimal threshold $a(x)$ satisfies

$$\psi = \int_a^{b(x)} \left(\frac{f(x, y)}{f(x, a)} - 1\right) u(y)dy. \tag{28}$$

Since $y > a$, the integrand on the right side increases in x if $f(x, y)/f(x, a)$ is strictly increasing

[36] We do not discuss the complex argument for existence of equilibrium. See Smith (2006) or its distillation in Smith (2011). Burdett and Coles (1997) also shows equilibrium existence in a model with entry flows but without complementarities.

in x — namely, $f(x,y)$ is strictly log-supermodular. As a result, $a(x)$ is increasing, and so PAM ensues for the highest agents. But then the inverse $b(x)$ increases, and this logic works for all types x. Next, return to the earlier example with $f(x,y) = y$. In fact, Smith (1992) shows that the results held for multiplicative payoffs $f(x,y) = f_1(x)f_2(y)$. In the key observation that led to this general condition for PAM, both of these functions are log-modular, since $\log f = \log f_1 + \log f_2$.

Although we have not solved the fixed search cost search, a similar analysis yields the first-order condition:

$$c = \int_a^{b(x)} (f(x,y) - f(x,a))\, u(y) dy, \tag{29}$$

The integrand here increases in x if f is strictly supermodular in (x,y), for then $f(x,y) - f(x,a)$ is strictly increasing in x for all $y > a$. Once more, the optimal threshold $a(x)$ increases. Notice however that in this case, the fixed search case, knife-edge block-segregation PAM case arises with a modular production function $f(x,y) = f_1(x) + f_2(y)$ (Chade, 2001).[37]

All told, search frictions imposes a higher threshold for securing PAM, since higher types must have an incentive to hold out for higher types. This requires a complementarity, since higher types already bring higher payoffs to any match they form. With time cost of waiting, the search costs rise in proportion to the value, a stronger assumption is required — log-supermodularity rather than just supermodularity. While these assumptions on payoffs are clearly *sufficient* for PAM in the fixed search cost and the discounted case, either is *necessary* to ensure PAM for all unmatched distributions — otherwise, one could put a large mass on the failure type set, and violate PAM.

We have assumed a "unisex" matching model, but the logic for PAM holds with two-sided matching, with two populations, like men and women. In that case, the upper bound function $b(x)$ comes from the acceptance threshold of the other population.

The outlined argument yields a unique equilibrium, *given* the unmatched density function $u(x)$. But an equilibrium is really a triple $(v(x), a(x), g(x))$. Moving outside our model with a differentiable type distribution, Burdett and Coles (1997) provide a simple example with just two types, low and high, that exhibits multiple equilibria, once the unmatched density is accounted for. In the *non-selective* equilibrium, high type agents accept both high and low types. In the (*selective*) one, high type agents only accept matches with other high types. If less than half of types are high, then high types match with lower probability than the low types, and so comprise most of the unmatched pool; this raises the option value of waiting, and thereby induces them to choose a higher reservation type. The selection of types in the pool leads to multiplicity.

This result highlights the role of selection in generating multiple steady state equilibria. It has long been recognized that labor markets by themselves can generate multiplicity and therefore cyclical outcomes. Diamond (1982) shows that in a simple exchange economy, such multiple steady states arise, but his logic exploits the increasing returns to scale property of the quadratic

[37] Morgan (1996) also studies NTU matching with fixed search costs.

search technology. But Pissarides and Petrongolo (2001) finds evidence of constant returns in the observed matching technology in the labor market. What the result in Burdett and Coles (1997) establishes is that heterogeneous types and the endogeneity of the searching pool essentially creates local increasing returns for subsets of types. In a related two type model for transferable utility, Shimer and Smith (2001) analyze optimal policy and characterize the planner's assignment who is constrained by the search technology.[38] They find that even though there is no intrinsic uncertainty, optimal allocations may be nonstationary. This questions the focus on steady state models in search, and suggests that it may prove an intrinsic source of volatility. This is a difficult and inviting frontier of the search literature.

We have assumed nontrivial search frictions. The literature on search and trade, by contrast, focuses on the minimal frictions case, and seeks a foundation for the Walrasian outcome. To briefly touch on the analogous question here, Adachi (2003) assumes cloning and impatience, and shows that, for a general match output function f, the set of stationary equilibria converges to the set of stable matches as frictions vanish. So the stable matchings in Gale and Shapley (1962) are the limit of equilbria of the decentralized market with search frictions. Lauermann and Noldeke (2014) shows that without cloning, convergence of equilibrium matchings is guaranteed if and only if there is a unique stable matching. Otherwise, there exists a sequence of equilibria converging to unstable allocations.

4.3 Sorting with Random Search and Transferable Utility

We now turn to the other benchmark: matching with search frictions under TU. It is essential to consider transfers between matched partners in order to broaden the applicability of the search and matching models. For instance, while there are certainly non-transferable aspects to an employment relation, the wage is the central part that determines the terms of trade. Even in the marriage market, there are many transfers between partners, both monetary (like shared income or joint mortgage payment) and non-monetary (such as division of child care or household chores).

The present value of an unmatched type x accepting a match with type y is no longer exogenously determined, since the surplus is mutual. In a unisex model, it is most natural to split the match surplus equally. While called the Nash bargaining solution, this uses none of the richness of that concept given the linear surplus frontier. Easily, *match surplus* equals output less the returns on the unmatched value, or $s(x, y) = f(x, y) - rv(x) - rv(y)$. Whether surplus is non-negative is now mutual. Thus, we no longer need keep tract of an acceptance set and opportunity set, but a single *matching set* $M(x) = \{y|s(x, y \geq 0\}$. Analogizing the NTU Bellman equation (25):

$$rv(x) = \frac{1}{2}\frac{\rho}{r+\kappa}\int_{\Omega(x)} \max(f(x,y) - rv(x) - rv(y), 0)u(y)dy$$

[38] They analyze a setup with a joint match surplus, as in the case of search with TU (see below), rather than NTU payoffs. Because they analyze a planner's problem, the transfers of course do not matter.

It is no longer straightforward to guess a sufficient condition for PAM or NAM, as it was in the NTU case. For now the value functions are an important part of the ratio that appears in the integral. Having absorbed the log-supermodular logic of Smith (1997, 2006), Shimer and Smith (2000) showed that PAM provided: $f_x(x,0) \equiv 0, f_{xy} > 0, (\log f_x)_{xy} > 0, (\log f_{xy})_{xy} > 0$. That is, the conditions require supermodularity of f, and log-supermodularity of its marginal product f_x and of its rate of increase with respect to a partner's type.[39] Eeckhout and Kircher (2010a) show that the conditions in Shimer and Smith (2000) together with $f_x > 0$ and $f_y > 0$ imply log-supermodularity of the match surplus, i.e., $\log f_{xy} > 0$, thus providing a lower bound on the strong complementarities that yield PAM.

Besides providing sufficient conditions for PAM, Shimer and Smith (2000) also prove existence of equilibrium, something that is easy if either the cloning assumption is used or more involved but elementary in techniques if a finite number of types is assumed (see Manea (2014)), but technically challenging with a continuum of types, arguably the standard case in applications. The proof in Shimer and Smith (2000) assumes the quadratic meeting technology; later, Noldeke and Troger (2009) showed existence under the technically harder linear meeting technology.

Instead of time discounting, one could assume a fixed cost of search. In environments where search resolves swiftly, a direct search cost in monetary terms may be a more appropriate measure of search costs than the opportunity cost of time. Atakan (2006) shows that a sufficient condition for PAM with TU is that production $f(x,y)$ be supermodular.

In summary, PAM obtains under supermodularity in the benchmark TU model. But with search frictions, greater complementarity in types is needed, such as log-supermodularity and even more, depending on the precise form of frictions. An important question is whether either model just described can shed light on actual labor markets, or if a new one is needed. The form of search frictions is critical when identifying the complementarities between worker ability and firm productivity. There are currently literatures that take macro and micro search models with heterogeneous agents to the data on labor markets. This exercise, even for applying Becker (1973) to the data on labor markets, is futile without possibly profound modifications for frictions. Finding the conditions for sorting with more general search costs are open questions.

The effects of frictional mismatch is another frontier of this literature. For example, Burdett and Mortensen (1998) introduced on-the-job search, and Postel-Vinay and Robin (2002) used it to capture important aspects of the data. This creates an initial mismatch among firms and workers, and then a career ladder for workers — with or without complementarities.

Another research thread starting at Shimer and Smith (2000) questions the correlation between worker and firm fixed effects. Using a simple two period model with heterogeneous workers and firms, Eeckhout and Kircher (2011) find that wages are non-monotonic in job productivity. While output rises in job productivity, given the mismatch from search, a high productivity job has a

[39] For a careful distillation of the proof, see Smith (2011).

high option value of continuing vacancy. That job will match with a lower skilled worker if it receives a high share of the output. Wages then have an inverted U-shape.

The presence of equilibrium mismatch offers sufficient variation to identify the technology underlying the match value output function $f(x,y)$. In a two-step procedure, Eeckhout and Kircher (2011) first derive the search cost from the wage distribution earns across jobs, and then use the range of job types that the worker matches with to derive the degree of complementarity between workers and jobs. Hagedorn, Law, and Manovskii (2012) extend this model to a general setting akin to Shimer and Smith (2000), and provide a methodology to identify the model.[40]

4.4 Directed Search

A. Background. The search framework based on random meetings is an appealing modeling tool for introducing frictions in labor and marriage markets. Yet, when agents are heterogeneous, the random meetings process has the feature that high characteristics necessarily meet with low characteristics, even though they know they will never form a match. This is particularly costly when the distribution is very spread out (or the complementarities are very high), and as a result, the acceptance sets are very narrow. Consider for example the market for executives. In the random search framework, executives must randomly be paired with janitor jobs, to only reject those. This illustrates that random search is in a sense not using all the relevant information.

In particular, both buyers and sellers (or firms and workers) are not making use of any information conveyed by wages. The role of wages under random search is to determine the split of output. Yet, in a competitive economy, prices also have an informative role. They indicate the willingness to pay for a good. Unlike random search, where trading partners meet and then determine the price through bargaining, under directed search the order is reversed. Firms commit to a price and post it, and after observing the price buyers choose whom to trade with. This allows buyers to direct their search towards those sellers that offer better prices.[41]

The directed search framework is nonetheless far from being competitive because it embodies frictions. Multiple workers may turn up for one job, and the inability to coordinate implies that in equilibrium some of the workers remain unemployed and some of the firms do not fill their vacancies. This lack of coordination stems from the decentralized nature of price competition, as there is no Walrasian auctioneer who sets prices to clear the market.

In the literature, there are two different interpretations of the actual mechanism to model frictions. One involves the formation of a *queue* which gives rise to a trading probability. This is often referred to as competitive search following Moen (1997), who analyzes stationary equilibria

[40]See also the related work by Teulings and Gautier (2004), De Melo (2009) Bagger and Lentz (2014), Lamadon, Lise, Meghir, and Robin (2013) Bartolucci and Devicienti (2013).

[41]The directed search model, mostly for homogeneous agents or heterogeneous agents without complementarities, has extensively been analyzed since the late 1970s: to name a few, see Butters (1977), Peters (1984), Moen (1997), Acemoglu and Shimer (1999), Burdett, Shi, and Wright (2001b).

in a continuous time setting. The other is a static setup where frictions arise due to *stockouts*. Several workers turn up for one job (or multiple jobs, but with fixed capacity), which results in probabilistic rationing. Both approaches are similar in the sense that they generate a trading probability as a function of the ratio of applicants to jobs and a wage (transfer).[42]

Because of the search frictions, traders now value both the price and the probability with which trade occurs. It is precisely that tradeoff that governs the strategies: sellers that post lower prices will attract more potential buyers and will therefore sell with a higher probability. Buyers who aim for lower priced goods must accept lower probabilities of trade since there are more competing buyers. With two-sided heterogeneity and complementarity, this tradeoff plays an important role in the determination of the equilibrium sorting patterns.

B. Sorting and Directed Search. There is a large number of buyers each with characteristic $x \in [0,1]$ who want to buy a unit of a good, and a large number of sellers with characteristic $y \in [0,1]$. The value of a good consumed by x and bought from y is $f(x,y) > 0$, where f is twice continuously differentiable. If the buyer pays a price p, then her utility is $f(x,y) - p$ and that of the seller is p. Seller characteristics are observable. The distributions of characteristics x is $G(x)$ and of characteristics y is $H(y)$ are continuous with positive densities $g(x)$ and $h(y)$.

As in the case of random matching, the primitive of the frictions is a matching function $M(b,s)$ that gives the number of matches generated when there are b buyers and s sellers. We assume that the matching function exhibits constant returns to scale, so the total number of matches can be written as a linear function of the individual matching probability $M(b,s) = sm(\theta)$, where $\theta = b/s$ denotes the ratio of buyers to sellers. Similarly, the probability that a buyer meets a seller and trades is $q(\theta)$. Since trade is pairwise, we have $q(\theta) = m(\theta)/\theta$. The function m is strictly increasing, strictly concave, and with elasticity decreasing in θ. Because we assume seller characteristics are observable to buyers, there is effectively a separate submarket for each (y,p) pair, and each market with its own market tightness θ.[43]

The interaction proceeds in two stages. First, each seller posts a price p at which she is willing to sell the good. Second, buyers observe all the quality-price pairs (y,p) and choose which seller to visit in order to buy. In equilibrium, strategies must satisfy the usual best response properties plus a perfection condition (see the paper for details).

Seller y's problem is to choose a price p that maximizes his expected profit, subject to offering an induced wage-queue length pair that assures a worker a utility of $U(x)$. The latter is the market utility that the worker obtains in this competitive equilibrium with frictions.

Versions of this model have been analyzed by Shi (2001), Mortensen and Wright (2002),

[42]The difference between the two interpretations becomes real for more general mechanisms than mere price posting (Eeckhout and Kircher, 2010b). We discuss this below under the header of competing mechanisms.

[43]When all identical sellers of characteristic y post the same price and all buyers of characteristic x choose to visit the same seller, the market tightness is $\theta(x,y) = g(x)/s(y)$.

Eeckhout and Kircher (2010a), and Jerez (2012). We will derive the condition for PAM and NAM using the differential version of the Legros and Newman (2007) condition on the 'embedded NTU' problem. We illustrate this idea in the directed search model of Eeckhout and Kircher (2010a). A seller with characteristic y solves the following problem:

$$\max_{x,\theta,p} \; m(\theta)p \qquad \text{subject to} \quad q(\theta)(f(x,y)-p) = U(x)$$

where $U(x)$ is the highest utility that buyer x can obtain. Solving the constraint for p and using $q(\theta) = m(\theta)/\theta$ yields $\max_x \left[\max_\theta m(\theta)f(x,y) - \theta v\right]$, where we have replaced $U(x)$ by v since it will not affect the sorting result below.[44] The first order condition of the inner maximization problem is $m'(\theta)f(x,y) = v$, from which we obtain the optimal value $\theta(x,y,v)$. It is easy to show that (we suppress the arguments for simplicity)

$$\theta_x = -\frac{m' f_x}{m'' f}, \quad \theta_y = -\frac{m' f_y}{m'' f}, \quad \text{and} \quad \theta_v = \frac{1}{m'' f}. \tag{30}$$

The maximization problem with respect to x can now be written as $\max_x \; m(\theta(x,y,v))f(x,y) - \theta(x,y,v)v \; \equiv \; \max_x \; \phi(x,y,v)$. Notice that the objective function is now nonlinear in v, despite the fact that agents' preferences are linear in money. The reason is that the presence of search frictions and the seller's choice of θ turns the problem into an NTU one. Therefore, the equilibrium will exhibit PAM or NAM depending on wether $\phi_{xy} - (\phi_y/\phi_v)\phi_{vx}$ is positive or negative.

Using $m'(\theta)f(x,y) = v$, it is easy to verify that $\phi_y = mf_y, \phi_v = -\theta, \phi_{xy} = m'\theta_y f_x + mf_{xy}$, and $\phi_{vx} = -\theta_x$. Therefore, together with (30) we obtain

$$\phi_{xy} - \frac{\phi_y}{\phi_v}\phi_{vx} \gtreqless 0 \quad \Leftrightarrow \quad \frac{f_{xy} f}{f_x f_y} - \xi(\theta) \gtreqless 0, \tag{31}$$

where $\xi(\theta) = m'(\theta)(\theta m'(\theta) - m(\theta)))/(m''(\theta)m(\theta)\theta)$. Inequality (31) is *exactly* expression (14) in Eeckhout and Kircher (2010a), but derived here in a very short way by exploiting the NTU problem embedded in their model.

The rest of the proof proceeds as in Eeckhout and Kircher (2010a). Their assumptions on the search technology imply that $0 < \underline{\xi} < \xi(\theta) < \overline{\xi} < 1$. Hence, there is PAM if $f_{xy}f/(f_x f_y) \geq \overline{\xi}$; letting n solve $1 - (1/n) = \overline{\xi}$, then PAM obtains if

$$\frac{f_{xy}(x,y)f(x,y)}{f_x(x,y)f_y(x,y)} \geq 1 - \frac{1}{n},$$

which is the condition for f to be n-root supermodular, or $f^{1/n}$ supermodular. This condition

[44] This is because if the condition holds for all values of v, it clearly holds for those values $v \in U([0,1])$. See Legros and Newman (2010), Section 7.1.

is necessary if, as usual, one wants the result to hold for all distributions of characteristics. A similar analysis can be done for NAM. We refer the reader to their paper.

Notice that the equilibrium exhibits PAM if the match payoff function satisfies $f_{xy} \geq \xi(\theta)\, \frac{f_x\, f_y}{f}$. This condition relates the degree of supermodularity in match value f_{xy} to the elasticity of the matching function $\xi(\theta)$. The higher the elasticity of substitution, the stronger the degree of supermodularity required in order to obtain PAM. Indeed, the precise condition is that $f(x, y)$ be root-supermodular, where the root is equal to $1/(\xi(\theta) + 1)$.

To gain some intuition about the economics behind this condition, consider first the case of no complementarity in the match value function, $f_{xy} = 0$. This is also the case analyzed in Mortensen and Wright (2002). Then if the elasticity of substitution is not equal to zero, there will be negative assortative matching. This is because the matching function has the role of providing trading security. Firms of high characteristics would ideally like to trade with a high probability, which can be achieved with a high worker to firm ratio θ. Since low worker characteristics trade off a higher wage for a lower trading probability, this high ratio is obtained from matching few high firm characteristics with many low worker characteristics. Likewise for matching many low firm characteristics with few high worker characteristics. This is negative assortative matching. In order to revert the sorting pattern to positive assortative matching, there must be sufficiently strong complementarities in the match value.

To understand the role of the elasticity of substitution, consider the constant elasticity of substitution (CES) matching function, where $\xi(\theta)$ is constant. If the elasticity is zero (Leontief) then supermodularity (and monotonicity) are sufficient. In this case, the matching probability changes drastically in the ratio of workers to firms and it is impossible to substitute workers for firms in the matching technology. Instead, when the elasticity is maximal and equal to one, then f must be log-supermodular.[45] In this case, the matching probability varies little with the worker to firm ratio, and therefore there is a strong force towards negative assortative matching. To overcome this force, a strong enough degree of supermodularity in match values is needed.

C. Market Segregation. Market segmentation is an alternative to directed search. It assumes random matching, but allows heterogeneous agents to set up separate trading posts. Jacquet and Tan (2007) argue that even under random matching, there are gains for individuals to set up segregated markets. For example, starting from the perfect segregation equilibrium, suppose all agents in the upper class were to be able to meet only among themselves. Provided they can coordinate, it is in all agents' benefit to set up those segregated markets.

Clearly, once a market is segregated, the initial reservation strategy is no longer a best response since the distribution of singles has changed (in fact, it is truncated). Since the pool has improved, the implication is necessarily an increase in the reservation strategy, conditional on acceptance.

[45] Because the matching probability is bounded by one, the elasticity is bounded too.

With the ability to coordinate, agents can segregate further, setting up and even finer partition of characteristics in each market. By induction, one might expect this to lead to perfect segregation as in the frictionless model: every characteristic matches only with exactly one characteristic on the other side of the market. Jacquet and Tan (2007) show that this reasoning is wrong. While it would be ideal to match with one characteristic only, any agent cannot commit to rejecting a partner who is slightly below the ideal partner. Because of search frictions, it takes time to wait for the ideal partner if the current candidate is only marginally lower, and thus the loss in match value compensates the cost of search. Therefore each of the segregated markets consist of a non-degenerate distribution of characteristics as they attract lower characteristics who know that the higher characteristics will not reject them. There is public good component due to non-excludability. But notice that if somehow access to the segregated market could be made exclusive, then only the ideal characteristics would be let in.

D. Competing Mechanisms. There is a close relationship between directed search and competing mechanism design. McAfee (1993) and Peters (1997) argue that if one allows the set of feasible mechanisms to include more than just price posting, then price posting need not be optimal. Consider the basic model with heterogeneity on the buyer side, but without sorting. To parallel the setup of the competing mechanism design literature, we discuss the model of sellers and buyers rather than firms and workers. Buyers are heterogeneous in their private valuation for the homogenous goods that sellers have for sale. Sellers are heterogeneous in their outside option value for the good they sell. Within a broad class of direct mechanisms (e.g., auctions, price posting, bargaining, etc.), all sellers simultaneously choose the mechanism that maximizes their profits. Buyers observe the announced mechanisms knowing that sellers are committed to their announcements. Then buyers choose which seller to visit. As in the directed search literature, the focus is on symmetric buyer strategies where a buyer visits with equal probability any one of the sellers who announce the same mechanism. This implies that with positive probability there is no trade, and as a result there are frictions due to coordination failure.

McAfee (1993) shows that there exists an equilibrium in second-price auctions with a reserve price equal to the firm's outside option value for the good. Buyers visit each seller with a probability that decreases with the reserve price. With identical outside options, buyers visit each seller with equal probability. Compared to the optimal second-price auction by a monopolistic auctioneer where the reserve price is set strictly above the outside option value, the reserve price is minimized here. This is due to competition among sellers, much like Bertrand price competition, in order to attract as many buyers as possible. In the competitive auction, the number of participants is determined endogenously and is inversely related to the reserve price.

McAfee (1993) further shows that the second-price auction is always a weak best response to any arbitrary set of strategies (choices of mechanisms) by the other sellers. While this does not preclude that equilibria in strategies other than second-price auctions exist, it is a strong endorsement for the use of second-price auctions. Because of the continuum of agents in his model, McAfee (1993) makes an assumption on the absence of strategic interaction under deviations. Subsequently, Peters and Severinov (1997), and Peters (1997) provide foundations for that large market assumption, by considering the limit of the subgame perfect equilibrium of a finite game as the number of agents grows large.

McAfee's results establishes that auctions are superior to posted prices when there is competition between auctioneers. Eeckhout and Kircher (2010b) show that this result depends crucially on the fact that auctioneers can extract rents ex post from buyers by having them participate simultaneously in the auction. In the queueing interpretation of directed search as in Moen (1997), the firm only faces one bidder at a time and it makes no sense to run an auction since the bidder would always pay the reserve price, which is effectively a posted price. This sheds light on the fact that the real world prevalence of posted prices over auctions hinges in part on the nature of the search frictions. If those frictions do not permit the auctioneer to round up sufficient applicants to bid in the mechanism, she is better off simply posting a price. The outcome is of course substantially different, since under posted prices, it is optimal for buyers to sort ex ante. Each buyer chooses a different posted price-trading probability pair thus revealing her characteristic. Instead, under competing auctions, the buyers visit the sellers randomly and are screened ex post.

In a labor market setting, Shi (2002) and Shimer (2005) consider different version of competition in demand-contingent prices. This has an auction flavor since the price paid depends on the composition of the ex-post demand, i.e., the number and characteristic of agents that turn up. When workers' characteristics are observable, Shimer shows that in equilibrium it is optimal to announce schedules such that agents of different characteristics apply for the same job. This ensures that high productivity workers obtain a job with the highest probability. In turn, low productivity workers only obtain a job if no high-characteristic workers show up. This allocation is similar to that of a second-price auction. As Shi (2002) argues, such ex post rankings is not only realistic in many market settings, but is also important for the efficient allocation of resources.

E. Directed search and large firms. The directed search model lends itself also to the analysis of large firms in the presence of frictions. Under random search, Smith (1999) analyzes the role hiring in large firms that have decreasing returns to employment. A key issue is how wages are set, and Stole and Zwiebel (1996) have proposed a sequential Nash bargaining procedure between the firm and the marginal worker where the firm's outside option is profits with one less worker. Such bargaining is inefficient and leads to overemployment with firms larger than what a planner would like. Kaas and Kircher (2014) propose a directed search model with large firms and show that price posting with coordination frictions yield a constrained efficient division of the surplus. These search models can handle realistic environments where firms hire multiple workers with a non-additive technology. This is useful for applied work where most data is generated from matching workers to firms with multiple jobs, and where the firm size is endogenous.

Finally, Eeckhout and Kircher (2012) combine large firms, frictional labor markets and two-sided heterogeneity. This permits the study of sorting in the presence search frictions and large firms. That directed search model builds on the sorting model without frictions that we have outlined above. Each firm produces output $F(x, y, l_x, r_x)$ with a worker characteristic y, a firm characteristic x, l_y workers of characteristic y and r_y resources devoted to characteristic y. Total output is the sum over the output of each worker characteristic (again as above, there is an assumption of no complementarities between different skill characteristics x). Now the hiring decision – how many vacancies v to post – is governed by the matching function as well as by the expected wage a worker receives. We assume atomless workers, so that the actual number of jobs filled l is equal to the number of vacancies that are filled given search fictions: $l = m \cdot v$, where $m(\theta)$ is the firm's matching probability with θ equal to the ratio of workers per vacancy. Risk neutral workers care about the expected wage denoted by $w = \omega m(\theta)/\theta$, where ω is the realized wage payment to a given characteristic y. Finally, the cost of posting a vacancy is a constant c. The firms problem can thus be written as:

$$\max_{r_y, \omega_y, v_y} \int \left[F(x, y, l_y, r_y) - l_y \omega_x - v_x c\right] dx$$
$$\text{s.t. } \; l_y = v_y m(\theta_y); \quad \text{and} \quad \omega_y m(\theta_y)/\theta_y = w(y)$$

and r_y integrates to unity. This problem has an equivalent representation that simplifies the analysis: $G(x, y, s, r) = \max_v \left[F(x, y, vm(s/v), r) - vc\right]$, and solve $\max_{s_y, r_y} \int [G(x, y, s_y, r_y) - w(y)s_y] dx$. This formulation permits this search problem to be analyzed as a sorting problem without frictions, pretty much in the same way we analyzed it before.

5 Matching, Information, and Dynamics

Search frictions is the story of costly coordination — not knowing where a counterparty is, or how hard it is to match with him. Informational frictions expands the scope towards not knowing the match payoffs or types or other costs of individuals; this richer form of frictions promises to be the next frontier in matching models. We touch on some promising highlights of the work.

5.1 Static Models

A. Stability under Incomplete Information. A crucial assumption in Becker's matching model is that agents' types are publicly observable. This is not the case in many marriage and labor market applications. A natural question then is what constitutes a stable matching under incomplete information about agents' types, for checking whether a blocking pair exists, the agents involved must be able to compute their payoffs from rematching, and that requires some knowledge about their partner's type. There have been some attempts at formalizing a workable notion of stability, the most recent and relevant one for our purposes being the definition of stability in Liu, Mailath, Postlewaite, and Samuelson (2014), who analyze matching with one-sided incomplete information and TU.[46] Their incomplete information stability notion is in the spirit of rationalizability in game theory, rather than mechanism design. A interesting result they show is that a mild strengthening of supermodularity yields PAM under incomplete information.

B. Sorting with Signalling Costs. In some matching applications in labor and marriage markets, agents with private information about their characteristics try to signal them to the other side of the market before matching takes place. Intuitively, those signals may be costly to send, and such costs reduce the benefits of sorting. Hoppe, Moldovanu, and Sela (2009) analyze this issue in a static model with production complementarities and incomplete information about types. In their model, if x matches with y then the utility of each agent is xy minus any signalling cost. They consider two populations, men and women, who engage in the following contest: agents simultaneously send signals, which consist of a bid or amount of utility that they give away. After observing all the signals, a planner assortatively matches men and women by signal.

In one equilibrium, every one bids zero, and the planner randomly matches the two sides. The paper shows that there is another equilibrium in strictly increasing strategies, with positively sorted agents. They ask whether the random matching welfare dominates PAM, net of signalling costs. The paper provides conditions under which this is the case in both the case with a finite number of agents and with a continuum of them. In the latter simpler context, assume that two unit mass populations with the same type distribution G and density g on $[0, 1]$.

In the random matching equilibrium the expected total welfare is $2E[x]E[y] = 2(\int_0^1 xg(x)dx)^2$. Under perfect PAM, the total expected output is $2\int_0^1 x^2 g(x)dx$. If both sides use the same signalling quantity $\beta(x)$, then the equilibrium utility of type x equals:

$$\pi(x) = \max_z xz - \beta(z) = x^2 - \beta(x).$$

We proceed as in a first price auction, applying the revelation principle. Since every type x must optimally bid as if had type x, the Envelope Theorem yields $\pi'(x) = x$, and so $\pi(x) = \int_0^x sds$. Altogether,

$$\beta(x) = x^2 - \pi(x) = x^2 - \int_0^x sds = \int_0^x sds,$$

where the second equality follows from integration by parts. Hence, $\beta(x) = \int_0^x sds$ for all x constitutes the Bayesian equilibrium of the game that exhibits PAM. Total welfare is then

$$2\int_0^1 x^2 g(x)dx - 2\int_0^1 \left(\int_0^x sds\right) g(x)dx = \int_0^1 x^2 g(x)dx.$$

[46] For the NTU case and centralized matching, see Roth (1989) and Chakraborty, Citanna, and Ostrovsky (2010).

That is, equilibrium signalling costs consume half of output. Thus, comparing welfare under PAM versus random matching, we obtain

$$\int_0^1 x^2 g(x)dx - 2\left(\int_0^1 xg(x)dx\right)^2 = Var(x) - E[x]^2 = E[x]^2\left(CV(x)^2 - 1\right), \tag{32}$$

where $CV(x) = \sqrt{Var(x)}/E[x]$ is the coefficient of variation of x. Thus, if $CV(x) < 1$ random matching outperforms PAM, and the opposite hold if greater than one. It turns out (Barlow and Proschan (1996), Corollary 4.9) that if the hazard rate $g(x)/(1-G(x))$ is increasing in x, then $CV(x) < 1$ and thus random matching dominates PAM, while the opposite is true if it is decreasing in x. Hence, the class of distributions with monotone hazard rate functions yield strong predictions regarding the welfare under random matching compared to PAM.

C. College-Student Matching. An important and increasingly newsworthy problem revisits the matching of students and colleges first explored in Gale and Shapley (1962). The real world matching is fraught with search and information frictions ignored in Gale and Shapley (1962). How students and colleges react to these frictions determines the sorting of students across colleges — the subject of the college admissions model of Chade, Lewis, and Smith (2011).

In their model, two colleges 1 and 2 with capacities κ_1 and κ_2, and a unit mass of students with type x whose distribution has a positive density $g(x)$ over $[0,\infty)$. College capacity cannot accommodate all the students, i.e., $\kappa_1 + \kappa_2 < 1$. Capturing the search friction, students pay a separate application cost $c > 0$ for each college. Students uniformly prefer college 1 to college 2: Attending college 1 yields a utility 1, college 2 yields $u \in (0,1)$, and zero is the utility for not attending college. By fixing the payoffs of colleges, one might thus understand this as an accurate short to medium run description of the college world. Students maximize expected college payoff less application costs. Colleges maximize the integral quality of their student bodies.[47]

The results are easiest to grasp when students know their type, but colleges only observe a noisy conditionally independent signal of each applicant.[48] Signal outcomes σ are drawn from a continuous density $m(\sigma|x)$ with support on an interval of $\mathbb{R}$ (e.g., $[0,1]$), and cdf $M(\sigma|x)$. The density has the strict monotone likelihood ratio property (MLRP): $m(\tau|x)/m(\sigma|x)$ is increasing in x if $\tau > \sigma$. To ensure that very high types are almost never rejected, and very poor ones are almost always rejected, the signals must be able to reveal extreme types: So assume that $M(\sigma|x) \to 0$ as $x \to \infty$ and $M(\sigma|x) \to 1$ as $x \to 0$ for any interior σ.

Students costly choose a portfolio of college applications (as in the decision problem in §3.2). Students apply simultaneously to either, both, or neither college, so each type x chooses a portfolio $S(x)$ in $\{\varnothing, \{1\}, \{2\}, \{1,2\}\}$. In turn, colleges set admissions standards, consisting of a threshold

[47] This is a strict NTU model with no complementarities between college and student characteristics (each side to a match enjoys the type of the other side).

[48] Most of their insights carry over to a world in which students only see noisy signals of their true types.

signal $\underline{\sigma}_i$ such that college i admits students with signal realizations above the threshold.

An equilibrium is a triple $(S^*(\cdot), \underline{\sigma}^*_1, \underline{\sigma}^*_2)$ such that, given $(\underline{\sigma}^*_1, \underline{\sigma}^*_2)$, $S^*(x)$ is an optimal college application portfolio for each x, and given $(S^*(\cdot), \underline{\sigma}^*_j)$, standard $\underline{\sigma}^*_i$ maximizes college i's payoff.

An equilibrium exhibits sorting if college and student strategies are "increasing". This means that the better college is more selective $(\underline{\sigma}^*_1 > \underline{\sigma}^*_2)$ and higher type students are increasingly aggressive in their portfolio choice: The weakest apply nowhere; better students apply to college 2; even better ones "gamble" by applying also to college 1; the next tier up applies to college 1 while shooting an "insurance" application to college 2; finally, the top students just apply to college 1. Strategies that are monotone in this fashion ensure the intuitive result that the distribution of student types accepted at college 1 first-order stochastically dominates that of college 2.

We will exploit a simple graphical analysis of the student's problem for given college thresholds in Chade, Lewis, and Smith (2011). Consider a student with respective admission chances $0 \leq \alpha_1, \alpha_2 \leq 1$. His expected payoff of applying to both colleges is $\alpha_1 v + (1-\alpha_1)\alpha_2 u$. The marginal benefit MB_{ij} of adding college i to a portfolio of college j is then:

$$\begin{aligned} MB_{21} &\equiv [\alpha_1 + (1-\alpha_1)\alpha_2 u] - \alpha_1 = (1-\alpha_1)\alpha_2 u \\ MB_{12} &\equiv [\alpha_1 + (1-\alpha_1)\alpha_2 u] - \alpha_2 u = \alpha_1(1-\alpha_2 u) \end{aligned}$$

The plot of these two curves looks like Figure 1 when $c < u(1-u)$ and $c < u/4$, i.e. with applications not too costly. As a result, the student's optimal portfolio choice is to apply:

(a) Nowhere, if $c > \alpha_1$ and $c > \alpha_2 u$.

(b) Just to college 1, if it beats applying just to college 2 ($\alpha_1 \geq \alpha_2 u$), and nowhere ($\alpha_1 \geq c$), and to both colleges ($MB_{21} < c$, i.e. adding college 2 is worse).

(c) Just to college 2, if it beats applying just to college 1 ($\alpha_2 u \geq \alpha_1$), and nowhere ($\alpha_2 u \geq c$), and to both colleges ($MB_{12} < c$, i.e. adding college 1 is worse).

(d) To both colleges, if this beats applying just to college 1 ($MB_{21} \geq c$), or just to college 2 ($MB_{12} \geq c$).

This optimal decision rule neatly partitions the unit square into four application regions. corresponding to the portfolio choices (a)–(d), denoted Φ, C_2, B, C_1, shaded in the right panel of Figure 1. Region B consists of students who either apply to college 2 and send a *stretch* application to college 1, or who apply to college 1 and send a *safety* application to college 2.

Let us now endogenize the acceptance chances by considering the noisy admissions process. Notice that not all pairs of acceptance chances (α_1, α_2) are 'feasible,' since these chances are pinned down by the student's type and the college thresholds. Fix the thresholds $\underline{\sigma}_1$ and $\underline{\sigma}_2$ set by college 1 and college 2. Student x's acceptance chance at college $i = 1, 2$ is given by $\alpha_i(x) \equiv 1 - M(\underline{\sigma}_i|x)$. Since a higher type student generates stochastically higher signals, $\alpha_i(x)$ increases in x. We can then invert α_1 and define the following *acceptance function* that links

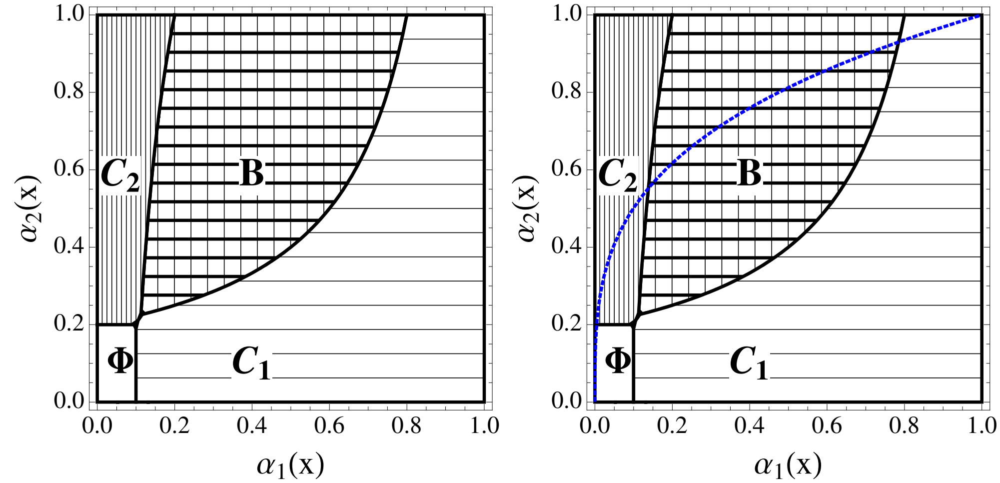

Figure 1: **Students Portfolio Problem** In the left panel, a student in the blank region Φ applies nowhere. He applies to college 2 only in the vertical shaded region C_2; to both in the hashed region B, and to college 1 only in the horizontal shaded region C_1. The right panel depicts the acceptance function $\psi(\alpha_1)$ for the case of exponential signals. As their caliber increases, students apply to nowhere (Φ), college 2 only (C_2), both colleges (B), and finally college 1 only (C_1). Student behavior is monotone in this case.

acceptance chances for each type x given colleges thresholds:

$$\alpha_2 = \psi(\alpha_1, \underline{\sigma}_1, \underline{\sigma}_2) = 1 - M(\underline{\sigma}_2 | \xi(\alpha_1, \underline{\sigma}_1))$$

For example, $\psi(\alpha_1) = \alpha_1^{\underline{\sigma}_2/\underline{\sigma}_1}$ given exponentially distributed signals, $m(\sigma|x) = (1/x)e^{-\sigma/x}$. This is increasing and strictly concave if $\underline{\sigma}_1 > \underline{\sigma}_2$ — and lies above the diagonal at the right panel in Figure 1. Although the acceptance function need not in general be concave, it does have a falling secant: α_2/α_1 is a decreasing function. This is a testable property of admissions, that as a student type increases, the ratio of his acceptance chances at college 2 to college 1 decreases.

The acceptance function and the application strategy respectively capture opportunities and preferences for student applications. By superimposing them in Figure 1 (right panel), we can visualize the choices that students make: The figure depicts a monotone application strategy, in which higher types apply more aggressively to college. And since $\underline{\sigma}_1 > \underline{\sigma}_2$ in the picture, it follows that this strategy profile, if it could be sustained in equilibrium, would exhibit the stochastic form of PAM described above, as casual intuition suggests.

Yet there are two possible sorting violations, both illustrated in Figure 2. The first occurs when stronger students do not apply more aggressively. For relatively high types may apply just to college 2, while some lower types also send stretch applications to college 1. This is depicted in the left panel of Figure 2, where application sets are $\Phi, \{2\}, \{1,2\}, \{2\}, \{1,2\}, \{1\}$ as student type rises. This can be an equilibrium if college 1 is not "sufficiently better" than college 2, for then one can find signal densities with the strict MLRP that engenders this non-monotone behavior.

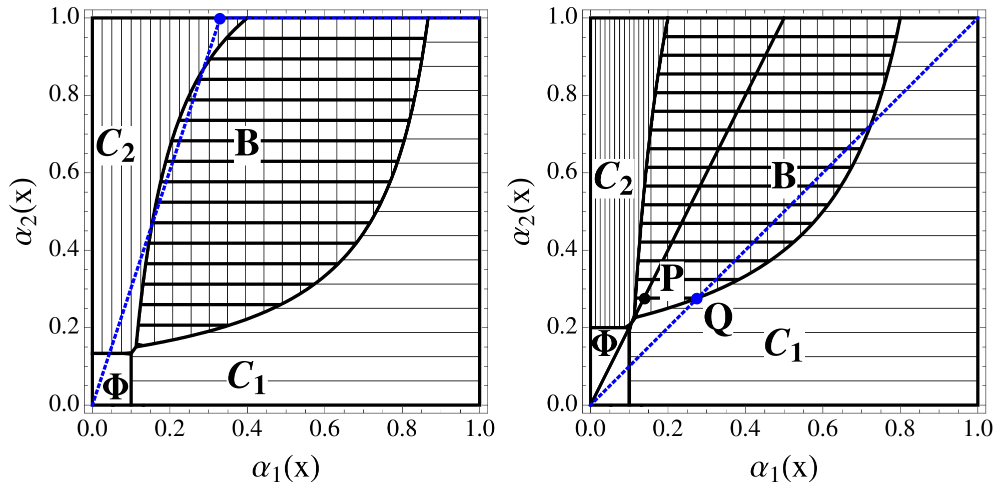

Figure 2: **Non-Monotone Behavior.** In the left panel, student behavior is non-monotone, since there are both low and high types who apply to college 2 only (C_2), while intermediate ones insure by applying to both. In the right panel, equal thresholds at both colleges induce an acceptance function along the diagonal, $\alpha_1 = \alpha_2$. Student behavior is non-monotone, as both low and high types apply to college 1 only (C_1), while middle types apply to both.

The second violation occurs when the worse college sets a higher admissions threshold. To see this, assume that both colleges set the same thresholds. As seen in the right panel of Figure 2, the application sets transition through $\Phi, \{1\}, \{1, 2\}, \{1\}$ as the student type rises. In this case, college 2 attracts only safety applications. The paper shows that this is an equilibrium outcome for a large enough capacity of college 2 (for by making it small enough, college 2 enjoys a higher standard than college 1).[49] The paper shows that a sufficient condition for all equilibria to exhibit sorting is that college 2 is sufficiently worse than college 1 (specifically, $u \leq 0.5$), and college 1 is small enough in capacity relative to college 1: Graphically, in this case the acceptance function traverses the unit square high enough as to preclude the case in the right panel of Figure 2.

The paper also conducts equilibrium analysis in the spirit of supply and demand, where the supply is the college capacity, and the demand is the derived enrollment function at each school. In this metaphor, the acceptance thresholds act like prices that equilibrate the two college markets. Comparative statics reflect not only a "standards effect" by existing applicants, but also a "portfolio effect" — as relaxed standards encourage applications. The latter yields surprising results: for example, a capacity increase at the worse college can reduce admission standards at the better college, via portfolio reallocation effects triggered by the students applications.

[49] The sorting failures can be drastic. For instance, consider the right panel of Figure 2. If $g(x)$ concentrates most of its mass on the interval of low calibers who apply just to college 1, then the average caliber of students enrolled at college 1 will be strictly smaller than that at college 2.

5.2 Dynamic Models

A. Sorting with Evolving Reputations. Anderson and Smith (2010) asks the question of whether Becker's assortative matching of types extends to reputations. For in many economic settings, parties to a match do not know their characteristics and learn them over time as they observe the output produced in a match.[50] Matching then solves two distinct objectives: on the one hand, it serves to exploit complementarities in production between the partners; on the other hand, it provides information about agent's attributes that may allow them to improve their continuation payoffs in future matches. Anderson and Smith (2010) explore the trade-off between these two goals. They show that despite production complementarities, PAM generally fails at high discount factors due to the importance of information. They argue that it is neither an equilibrium or an optimum that agents with identical current reputations always match.

The paper presents a general matching model with evolving human capital. They first show that a Pareto optimal steady state and a Walrasian equilibrium exists, and proves the welfare theorems. We illustrate this finding in their simpler motivational two period partnership model.

Assume a continuum of agents of two underlying *true types*, high or low, i.e., $\theta \in \{\theta_\ell, \theta_h\}$. No one knows his own type, but merely the probability $x \in [0,1]$ of a high true type — called his *reputation*. Output is stochastic, and can assume a finite number of positive values $q_1, \ldots, q_N$. The chance of each output q_i is h_i, m_i and ℓ_i, respectively, from a match between two high types, a low and high type, and two low types. Then the chance of production q_i from a match between two agents with reputations x and y is

$$p_i(x,y) = xyh_i + [x(1-y) + y(1-x)]m_i + (1-x)(1-y)\ell_i.$$

Let $H = \sum_i q_i h_i$, $M = \sum_i q_i m_i$, and $L = \sum_i q_i \ell_i$. Then the *expected output* is

$$f(x,y) = \sum_i q_i p_i(x,y) = xyH + [x(1-y) + y(1-x)]M + (1-x)(1-y)L,$$

Assuming $H + L - 2M > 0$, production is strictly supermodular in reputations, since $f_{xy} = H + L - 2M > 0$. In a one-shot model with transferable utility, PAM arises by Becker (1973).

Assume now a two period matching model. Let agents discount future payoffs by a common factor δ. In the second and last period, output is strictly supermodular, and so the matching exhibits PAM. As a result, the equilibrium wage of x is half of the output for the agent, $w(x) = f(x,x)/2$. Easily, $w''(x) = 2f_{xy} > 0$, and so the wage is convex in reputation.

But in the first period, matching plays both a production and an information role. To pin down the expected continuation payoff for an agent with current type x, notice that if he matches

[50] Early examples of matching models with learning about the match are Jovanovic (1978) and Jovanovic (1984). Although these papers derive very useful insights on the dynamics of turnover, they do not include ex-ante heterogeneity and thus they do not shed light oncomplementarities and sorting patterns.

with y, then after observing q_i in the first period x updates his belief that his type is high to $z_i(x,y) = p_i(1,y)x/p_i(x,y)$.[51] It follows that the expected continuation payoff for x is $\psi(x|y) = \sum_i p_i(x,y)w(z_i(x,y))$. Thus, the present value of a match between agents with types x and y is

$$v(x,y) = (1-\delta)f(x,y) + \delta(\psi(x|y) + \psi(y|x)).$$

If ψ were supermodular in (x,y), then v would be supermodular, and PAM would ensue, per Becker (1973). We will next show that PAM fails with sufficient patience, or large enough $\delta < 1$.

First, we make a useful preliminary observation. If $\delta = 1$, then $v(x,y) = \psi(x|y) + \psi(y|x)$ and only the continuation payoff matters for matching. We can show that ψ is convex in x and in y.[52] We now ask whether the matching exhibits PAM. To see this, consider three pairs $(0,0)$, $(1,1)$, and (x,x), where $x \in (0,1)$. Convexity of $\psi(x|y)$ in y implies that either $\psi(x|0) > \psi(x|x)$ or $\psi(x|1) > \psi(x|x)$. Easily, $\psi(0|x) = \psi(0|0)$ and $\psi(1|x) = \psi(1|1)$, for there is no informational value for these extreme types from matching with any other type. So either $\psi(x|0) + \psi(0|x) > \psi(x|x) + \psi(0|0)$ or $\psi(x|1) + \psi(1|x) > \psi(x|x) + \psi(1|1)$. Thus, PAM fails since re-matching x agents with either 0 or 1 raises total payoffs. Since this holds for $\delta = 1$, by continuity PAM fails for δ sufficently high. Inutitively, when δ is high enough the learning value of matching outweighs the complementarities in production. And since by matching with a 'corner' type all the variability in output is due to the interior x (this happens at least in one of the corners 0 and 1), it follows that matching x and the corner types as in PAM is dominated by mixing them.

So sufficiently forward-thinking behaviour leads to a failure of PAM. Unfortunately, as the discount factor rises to one in an infinite horizon model, the continuation value becomes linear. For intuitively, in the perfect patience limit, almost all production arises when one perfectly know all types; this means that output of type x is the linear weighted average $xH + (1-x)L$, that results from PAM given the true types. In the infinite horizon version, Anderson and Smith (2010) find a robust PAM failure: as the number N of output levels explodes, PAM fails near both high enough and low enough types with probability tending to one. The proof relies on a careful examination of the asymptotic behavior of the continuation value function, finding that the second derivative explodes near 0 and 1.

A key implication is that partnerships of identical types (either both θ_ℓ or both θ_h) eventually break up. Intuitively, as information accumulates over time, the probability that anyone is a high type approaches 0 or 1, and at that point, the above PAM failure kicks in, the match dissolves.

B. Sorting and Evolving Types. Inspired by the changing reputational types in Anderson and Smith (2010), Anderson (2014) explores the dynamics that arise when individuals are changed

[51] This is just an application of Bayes' rule: the denominator is the probability of q_i while the numerator is the prior probability x that his type is high times the probability of q_i if his type is indeed high and he matches with y.

[52] Information about one's matching partner induces mean zero noise in the posterior reputation, and is obviously valuable. Since mean zero gambles always have positive expected value, Pratt (1964) implies convexity.

by the association with their match partners. Assume an initial distribution over human capital G.[53] A matching μ is *feasible* when the measure of all matched types weakly below x equals $G(x)$. For a taste of his conclusions, assume a two period model, with types changing after period one: Specifically, if types (x, y) match in period one, then type x transitions to a new type $z \leq s$ with probability $\mathcal{T}(s|x, y)$. Given any feasible matching μ in period one, the distribution over human capital in the final period $H(x|\mu)$ can be naturally defined, given $\mathcal{T}$. Assume symmetric, supermodular output $f(x, y)$, so that PAM is optimal in the final period.[54] Given the final wage $w(x) \equiv f(x, x)/2$, the period-one continuation value is:

$$V(\mu) \equiv \int w(x)dH(x|\mu) \tag{33}$$

For a high enough discount factor, PAM is initially optimal when it maximizes (33) across all feasible matchings. Since the wage $w(x)$ is increasing, PAM maximizes (33) if and only if the continuation distribution under PAM first order stochastically dominates the continuation distribution $H(x|\mu)$ (i.e. minimizes $H(x|\mu)$) across all feasible matchings μ. Lorentz (1953) argues that this holds when $\mathcal{T}(s|x, y)$ is submodular in (x, y) for all s. With deterministic transitions, where (x, y) matched implies x updates to $\tau(x, y)$, we can write $\mathcal{T}(s|x, y) = \mathbb{I}_{s \geq \tau(x,y)}$. A salient special case in which $\mathcal{T}$ is submodular is $\tau(x, y) = \min\{x, y\}$. This is the "Bad Apples" case in the peer effects literature, in which the greater type is pulled down to the lesser one.

In the two period model, the continuation value is exogenous. In order to extend these PAM characterization results to the infinite horizon model, Anderson (2014) first characterizes the Planner's preferences over human capital distributions. These value characterization results require additional assumptions on the transition distribution. For example, the Planner's value rises in the increasing convex order over human capital distributions when $f(x, y)$ is individually convex in x and in y and $\int_z^1 \mathcal{T}(s|x, y)ds$ is individually concave in x and in y.

In a related model, Jovanovic (2014a) explores a dynamic matching model with imperfect information, where agents do not know their types, and are randomly matched in the first period. (This precludes any matching role for information in the first period.) They then observe the output produced, equal to the product of their true types. Finally, they decide whether to rematch ('recombine') in the second period. He shows that if the output produced is publicly observed, as in Anderson and Smith (2010), then all agents recombine in a PAM way in the second period. For signals enter in a complementary fashion in the expected product of the second period. But if output is only observed by the pair, then only those with low output recombine in the second period (adverse selection), and the overall matching exhibits negative correlation among pairs.

[53] Jovanovic (2014b) explores a related idea in an overlapping generations setup with two-period lives to study assortative matching and growth in the presence of mismatch due to shocks.

[54] Anderson and Smith (2010) establish the welfare theorems for this dynamic matching model. In particular, PAM is optimal *iff* PAM is a market outcome.

C. Marriage Markets and Age Gaps. Bergstrom and Bagnoli (1993) may be the first paper to incorporate incomplete information into a dynamic matching market. To shed light on the empirical regularity observed in most countries and across time that women on average marry older men, they develop an infinite horizon overlapping generations marriage market model with heterogeneous types, incomplete information about men's types, in which men and women time their entry into the marriage market. In equilibrium, males use their entry date into the matching market to signal their type, and men with higher types tend to marry later in life.

Assume a heterogeneous continuum of men and women. The type of a man is $x \sim G$ on $[0, 1]$ and of a woman is $y \sim H$ on $[0, 1]$. In an important novelty for the matching literature, this paper introduces the assumption of log-concavity, a property satisfied by many common distributions.[55] The cdf H is log-concave in y. Utility is nontransferable: If man x ever marries woman y, then he enjoys a positive flash utility $f_1(x)f_2(y)$, where $f_1'(x) \geq 0$ for all x, $f_2'(y) > 0$, $f_2''(y) \geq 0$, and additionally $f_2'''(y) \leq 0$ for all y. In turn, a woman of type y enjoys a match utility $\beta_1(y)\beta_2(x)$ if she ever marries a man of type x, with $\beta_1'(y) \geq 0$ and $\beta_2'(x) > 0$. Bergstrom and Bagnoli (1993) analyzed the simpler case with $f_1(x) = \beta_1(y) = 1$, $f_2(y) = y$, and $\beta_2(x) = x$.

An equal number of men and women are born in each period. Everyone lives for two periods and their only decision is whether to enter the marriage market in period 1 or period 2. Delaying marriage entails a fixed cost $c_1 > 0$ for men and $c_2 > 0$ for women. A woman's type is publicly observable, while a man's type is his private information in period 1, and publicly observable in period 2. As a result, the period that a man chooses to marry signals his type. Divorce is ignored, since the match payoff is one-time only.

A centralized matchmaker matches agents as follows — which also delivers the unique stable assignment. In each period, the planner assortatively matches men of age 2 and any women who choose to marry, until exhausting the supply of women, or of men of age 2 whose types exceed the expected value of the type of males of age 1. Specifically, man x marries woman $\mu(x)$, with $\mu'(x) > 0$. The rest of the women are assigned to age 1 males, whose true types are as yet hidden — and age 1 men are randomly matched to a random woman in period 1. Any unmatched agents of age 1 reappear in the marriage market when they reach age 2. The population has constant size, with men and women of age 1 and 2 always present in the market, and the same number of men and women entering period 2, so that everyone eventually matches.

An equilibrium must specify the agents' marrying strategies (age 1 or age 2). First of all, observe that women have no incentive to delay. Given the demographic stationarity, they secure the same expected payoff from marriage, but incur a fixed search cost c_2 only in period 2. But men solve a timing problem: In the spirit of a reservation wage, there is a cutoff value for men: high types wait until age 2, and low types enter at age 1. To see this, let women of types $C \subset [0, 1]$

[55] In an underground classic that was published more than a decade later, they then authored the log-concavity encomium Bergstrom and Bagnoli (2005). The importance of this property generally in economics had previously been introduced in Proposition 1 of Heckman and Honore (1990).

seek to marry age 1 men. Then a type x man strictly prefers to delay marriage until age 2 when

$$-c_1 + f_1(x)f_2(\mu(x)) \geq \frac{\int_C f_1(x)f_2(s)dH(s)}{\int_C dH(s)} \tag{34}$$

Easily, if this inequality holds for any type x, then it also holds for any higher type, thereby confirming the cutoff value property.[56] Hence, there is a threshold $\bar{x}$ such that men with $x \leq \bar{x}$ choose to marry at age 1 and those with types $x > \bar{x}$ choose to marry at age 2. When interior, the threshold solves indifference equation, namely (34) with equality, namely:

$$f_1(\bar{x}) \left(f_2(\mu(\bar{x})) - \frac{\int_0^{\mu(\bar{x})} f_2(s)dH(s)}{H(\mu(\bar{x}))} \right) = c_1. \tag{35}$$

In the purported equilibrium, in every period, age 1 women with high types marry age 2 men with high types, assortatively, whereas age 1 women with low types marry age 1 men with low types, but randomly. This is their story of the marriage age gap between men and women.

Does this equilibrium exist and is it unique? The answer is yes if (35) has a unique solution. First, the left side of (35) vanishes in the limit $\bar{x} \downarrow 0$ by l'Hopital's rule. Since f_2 is increasing, the left side of (35) exceeds c_1 at $\bar{x} = 1$, for small enough $c_1 > 0$. Existence follows by continuity. Uniqueness follows if the left side of (35) is strictly increasing in $\bar{x}$. For this, Bergstrom and Bagnoli (1993) introduce a log-concavity assumption. Since $f_1'(\bar{x}) \geq 0$, it suffices to show that the term in parenthesis is increasing in $z = \mu(\bar{x})$. Now, integration by parts reveals that

$$f_2(z) - \frac{\int_0^z f_2(s)dH(s)}{H(z)} = \frac{\int_0^z f_2'(s)H(s)ds}{H(z)} = f_2'(z)\frac{\xi(z)}{\xi'(z)},$$

where $\xi(z) = \int_0^z f_2'(s)H(s)ds$. So it suffices that $f_2'(z)\xi(z)/\xi'(z)$ is increasing in z, and since $f_2''(z) \geq 0$, it suffices that ξ/ξ' increases in z. This holds when $\xi''\xi - \xi'^2 < 0$ or, equivalently, when ξ is strictly log-concave in z. Since $f_2''' \leq 0$, we have f_2' is concave and thus strictly log-concave. If we assume a log-concave distribution H, then $f_2'H$ is strictly log-concave, and so too is the integral, as log-concavity is preserved by integration. All told, there is a unique equilibrium.

D. Matching and the Acceptance Curse. In many matching applications, such as the college admissions problem or marriage, the characteristics of agents on one or both sides of the market are only observed with noise prior to matching. Chade (2006) considers an NTU matching market with random search, where agents know their types but they only observe a noisy signal of potential partners they meet. After observing the signal, an agent updates his belief about the partner's type and then chooses whether to accept or reject. Intuitively, agents set a threshold for

[56]As in Smith (1997, 2006), multiplicative preferences in this fashion ensure identical cardinal preferences over match partners, and thus greatly simplifies analysis.

the signal realization and accept a partner when the signal observed exceeds a threshold. If both accept, they marry and leave the market, while in any other case they continue the search. Under the standard MLRP condition on the signal distribution, higher signal realizations convey better news about a partner's type. The twist here is that agents must also account the information in the event that the partner agrees to match. And if agents on the other side of the market grow more choosey as their types increase, then being accepted leads one to downgrade the posterior estimate of the potential partner's type. Chade suggestively called this the *acceptance curse*, since it is akin to the winner's curse effect in auction theory (Milgrom and Weber, 1982).

The model is in steady state over an horizon infinite in discrete time, with matched agents replaced by clones. Using the marriage market metaphor, there are continuum populations of men and women. The density of men's types $x \in [0,1]$ is $g(x)$, and of women's types $y \in [0,1]$ is $h(y)$. The per period utility of each agent is 0 if single, and the type of the spouse if matched (NTU). Every period, men and women randomly meet. When a man x meets a woman y, she observes a signal $\sigma \in [0,1]$ drawn from $m(\sigma|x)$, and he observes a signal $\tau \in [0,1]$ drawn from a conditional density $n(\tau|y)$, where m and n satisfy the strict MLRP. After observing the signals, both announce simultaneously accept or reject; if they both accept, they marry and exit the market, otherwise they continue searching next period. Agents discount the future by $\delta \in (0,1)$.

A stationary strategy for a woman of type y or a man with type x is a fixed set of signals that led either to accept. Intuitively, these are upper intervals of signals, $\sigma \geq \underline{\sigma}(y)$ and $\tau \geq \underline{\tau}(x)$, by the MLRP. Focus on a woman of type y facing a population of men. Let $m(\sigma) = \int_0^1 m(\sigma|x)g(x)dx$ be the unconditional density of signal σ, and $k(x|\sigma) = m(\sigma|x)g(x)/m(\sigma)$ the posterior density on x, given the signal realization σ. Then the chance $a(y|\sigma)$ that y's current partner accepts, conditional on σ, equals:

$$a(y|\sigma) = \int_0^1 \int_{\underline{\tau}(x)}^1 n(\tau|y)k(x|\sigma)d\tau dx,$$

Next, let $f(y|\sigma)$ be the expected discounted utility from marriage given the signal realization σ *and* the information contained in the event that she is accepted by the current partner. Formally,

$$f(y|\sigma) = \mathbb{E}\left[\int_0^1 \frac{x}{1-\delta} \frac{\left(\int_{\underline{\tau}(x)}^1 n(\tau|y)d\tau\right) k(x|\sigma)}{a(\sigma, y)} dx\right].$$

Consider a woman y seeing a signal σ. That man accepts with probability $a(y|\sigma)$. In this event, the woman decides whether to accept and leave the market, securing a discounted expected payoff $f(y|\sigma)$, or reject and continue searching, and thereby earn expected discounted payoff $\delta\Psi(y)$. If the man does not accept, which occurs with probability $1 - a(y|\sigma)$, then the woman continues to search. Her Bellman equation is thus:

$$v(y|\sigma) = a(y|\sigma)\max\{f(y|\sigma), \delta\Psi(y)\} + (1 - a(y|\sigma))\delta\Psi(y). \tag{36}$$

where $\Psi(y) = \int_0^1 v(y|\sigma)m(\sigma)d\sigma$ is the optimal continuation value, and $\underline{\sigma}$ solves $f(\underline{\sigma}(y), y) = \delta\Psi(y)$.

Similarly, the optimal strategy of a man of type x is a threshold $\underline{\tau}(x)$. Thus, the search for a stationary equilibrium reduces to finding a pair of functions $(\underline{\sigma}(\cdot), \underline{\tau}(\cdot))$ that are mutual best responses. The downward recursive construction in §4 under complete information is inapplicable here since any type may be accepted by any other, owing to signal noise. Chade (2006) shows that the model can be reinterpreted as a two-player game with incomplete information with a continuum of types and actions, and then one can appeal to a theorem in Athey (2001) to show that there exists a equilibrium in increasing strategies.

Finally, observe that the acceptance curse emerges: For since $f(\sigma, y) \leq \mathbb{E}[X/(1-\delta)|\sigma]$, the event of being accepted is a discouraging signal for a woman of type y. Nevertheless, stochastic sorting still emerges: the distribution of types an agent can end up matched is ordered in the sense of first order stochastic dominance as a function of the agent's type. As a result, in equilibrium one's expected partner's type is increasing in the agent's type.

6 Conclusion

We have reviewed the main frameworks used in micro models of search and matching, focusing on the conditions for sorting (either positive or negative) both with and without search or information frictions. We have started from the benchmark frictionless assignment model both with TU and NTU. Despite its simplicity, the model explains many interesting economic phenomena ranging from the labor market to financial markets and issues in international trade. The many variations of the model allow for a simple relation between technology and the resulting sorting pattern.

We have subsequently introduced both search and information frictions. While this duly complicates many aspects of the analysis, it renders the setting more realistic. We have carefully reviewed the most important sorting results in this area, explaining in a unified way the logic underlying them, and we have also discussed the emerging applied literature on the subject.

We think these these models can be a building block for further theory and a spark for empirical work. This literature is at best in its infancy, and many open questions remain. These include a better understanding of many-to-one matching models, the role of externalities in matching, stochastic types, nonstationary models, the role of on-the-job search, multidimensional models. Of course, there are also trade theoretic models of search. That side is less well-explored, and has so far focused on low levels of search frictions. The search and matching framework naturally captures heterogeneity, which is the hallmark of economic exchange and the source of gains from trade. Formally modeling the choice of whom to trade should prove useful in all fields of economics.

References

Acemoglu, D., and R. Shimer (1999): "Holdups and Efficiency with Search Frictions," *International Economic Review*, 40(4), 827–49.

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

Adachi, H. (2003): "A Search Model of Two-Sided Matching under Nontransferable Utility," *Journal of Economic Theory*, 113(2), 182–198.

Adam, K. (2001): "Learning while Searching for the Best Alternative," *Journal of Economic Theory*, 101, 252–280.

Albrecht, J., A. Anderson, and S. Vroman (2010): "Search by Committee," *Journal of Economic Theory*, 145(4), 1386–1407.

Albrecht, J., P. Gautier, and S. Vroman (2006): "Equilibrium Directed Search with Multiple Applications," *Review of Economic Studies*, 73, 869–901.

Anderson, A. (2014): "Steady-State in Matching and Bargaining," under revision for the *Journal of Economic Theory*.

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. (2001): "Single Crossing Property and the Existence of Equilibria in Games of Incomplete Information," *Econometrica*, 69(2), 861–889.

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

Barlow, R., and F. Proschan (1996): *Mathematical Theory of Reliability*. SIAM, New York.

Bartolucci, C., and F. Devicienti (2013): "Better Workers Move to Better Firms: A Simple Test to Identify Sorting," Discussion paper, IZA Discussion Paper.

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

Bergstrom, T., and M. Bagnoli (1993): "Courtship as a Waiting Game," *Journal of Political Economy*, 101(1), 185–202.

Bergstrom, T., and M. Bagnoli (2005): "Log-concave Probability and its Applications," *Economic Theory*, 26(2), 445–469.

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

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

Burdett, K., and K. Judd (1983): "Equilibrium Price Dispersion," *Econometrica*, 51, 955–969.

Burdett, K., and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.

Burdett, K., S. Shi, and R. Wright (2001a): "Pricing and Matching with Frictions," *Journal of Political Economy*, 109, 1060–1085.

——— (2001b): "Pricing and Matching with Frictions," *Journal of Political Economy*, 109(5), 1060–1085.

Butters, G. (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44, 465–491.

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

——— (2006): "Matching with Noise and the Acceptance Curse," *Journal of Economic Theory*, 129(1), 81–113, forthcoming, *Journal of Economic Theory.*

Chade, H., and J. Eeckhout (2014): "Competing Teams," UCL mimeo.

Chade, H., G. Lewis, and L. Smith (2011): "Student Portfolios and the College Admissions Problem," Arizona State University and University of Wisconsin Working Paper.

Chade, H., and L. Smith (2006): "Simultaneous Search," *Econometrica*, 74, 1293–1307.

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): "Two-Sided Matching with Interdependent Values," *Journal of Economic Theory*, 145(2), 85–105.

Chiappori, P., R. McCann, and L. Nesheim (2010): "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness," *Economic Theory*, 42(2), 317–354.

Chiappori, P.-A., and P. Reny (2005): "Matching to Share Risk," Columbia University.

Choi, M., and L. Smith (2013): "The Economics of Web Search," PhD thesis, University of Wisconsin.

Chow, Y., H. Robbins, and D. Siegmund (1971): *Great Expectations: The Theory of Optimal Stopping*. Dover.

Chvatal, V. (1983): *Linear Programming*. Freeman, New York.

Cole, H., G. Mailath, and A. Postlewaite (2001): "Efficient Non-Contractible Investmentes in Large Economies," *Journal of Economic Theory*, 101(2), 333–373.

Compte, O., and P. Jehiel (2010): "Bargaining and Majority Rules: A Collective Search Perspective," *Journal of Political Economy*, 118(2), 189–221.

Crowe, J. A., J. A. Zweibel, and P. C. Rosenbloom (1986): "Rearrangement of Functions," *Journal of Functional Analysis*, 66, 432–438.

Daminano, E., H. Li, and W. Suen (2005): "Unraveling of Dynamic Sorting," *Review of Economic Studies*, 53, 1057–1076.

De Melo, R. L. (2009): "Sorting in the labor market: Theory and measurement," Ph.D. thesis, Yale University.

DeGroot, M. (1970): *Optimal Statistical Decisions*. McGraw-Hill.

Diamond, P., and E. Maskin (1979): "An Equilibrium Analysis of Search and Breach of Contract I: Steady States," *The Bell Journal of Economics*, 10, 282–316.

Diamond, P. A. (1982): "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, 90(5), 881–94.

Edlin, A., and C. Shannon (1998): "Strict Single Crossing and Strict Spence-Mirrlees Condition: A Comment on Monotone Comparative Statics," *Econometrica*, 66, 1417–1425.

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

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

Eeckhout, J., and P. Kircher (2010b): "Sorting versus Screening: Search Frictions and Competing Mechanisms," *Journal of Economic Theory*, 145, 1354–1385.

——— (2011): "Identifying Sorting–In Theory," *Review of Economic Studies*, 78(3), 872–906.

——— (2012): "Assortative Matching with Large Firms: Span of Control over More versus Better Workers," UCL mimeo.

Ferguson, T. (Undated): *Optimal Stopping and Applications*. Electronic Text, UCLA.

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.

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

Gershkov, A., and B. Modovanu (2010): "Optimal Search, Learning and Implementation," University of Bonn Working Paper.

Gittins, J. C. (1979): "Bandit Processes and Dynamical Allocation Indices," *Journal of the Royal Statistical Society, Series B*, 41, 148–177.

Gittins, J. C., and D. M. Jones (1974): "A Dynamical Allocation Index for the Sequential Design of Experiments," in *Progress in Statistics*, ed. by J. G. et al., pp. 241–266. North Holland, Amsterdam.

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

——— (1999): "Perfect Competition in the Continuous Assignment Model," *Journal of Economic Theory*, 88, 60–118.

Grossman, G. M., E. Helpman, and P. Kircher (2013): "Matching and sorting in a global economy," Discussion paper, National Bureau of Economic Research.

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

Guadalupe, M., V. Rappoport, B. Salanie, and C. Thomas (2014): "The Perfect Match: Assortative Matching in International Acquisitions," Discussion paper, Working paper.

Hagedorn, M., T. H. Law, and I. Manovskii (2012): "Identifying Equilibrium Models of Labor Market Sorting," mimeo.

HARDY, G. H., J. E. LITTLEWOOD, AND G. POLYA (1952): *Inequalities*. Cambridge University Press, United Kingdom.

HATFIELD, J. W., AND P. R. MILGROM (2005): "Matching with Contracts," *American Economic Review*, 95(4), 913–935.

HECKMAN, J. J., AND B. E. HONORE (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–1149.

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

HOPPE, H. C., B. MOLDOVANU, AND A. SELA (2009): "The Theory of Assortative Matching Based on Costly Signals," *Review of Economic Studies*, 43, 253–281.

JACQUET, N. L., AND S. TAN (2007): "On the Segmentation of Markets," *Journal of Political Economy*, 115(4), 639–664.

JEREZ, B. (2012): "Competitive equilibrium with search frictions : a general equilibrium approach," Discussion paper.

JOVANOVIC, B. (1978): "Job Matching and the Theory of Turnover," *Journal of Political Economy*, 87(5), 972–990.

——— (1984): "Matching, Turnover, and Unemployment," *Journal of Political Economy*, 92(1), 108–122.

——— (2014a): "Learning and Recombination," New York University.

——— (2014b): "Misallocation and growth," *The American Economic Review*, 104(4), 1149–1171.

KAAS, L., AND P. KIRCHER (2014): "Learning and Recombination," New York University.

KELSO, ALEXANDER S, J., AND V. P. CRAWFORD (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50(6), 1483–1504.

KIRCHER, P. (2009): "Efficiency of Simultaneous Search," *Journal of Political Economy*, 117, 861–913.

KIRCHER, P., AND M. GALENIANOS (2009): "Directed Search with Multiple Job Applications," *Journal of Economic Theory*, 114, 445–471.

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.

KREMER, M., AND E. MASKIN (1996): "Wage Inequality and Segregation by Skill," Working Paper.

LAMADON, T. (2014): "Productivity shocks, dynamic contracts and income uncertainty," Discussion paper, Working paper.

LAMADON, T., J. LISE, C. MEGHIR, AND J.-M. ROBIN (2013): "Matching, sorting and wages," Discussion paper, Mimeo.

LAUERMANN, S., AND G. NOLDEKE (2014): "Stable Marriages and Search Frictions," *Journal of Economic Theory*, 151(5), 163–195.

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

——— (2010): "Co-ranking Mates: Assortative Matching in Marriage Markets," *Economics Letters*, 106, 177–179.

LINDENLAUB, I. (2014): "Sorting Multidimensional Types: Theory and Application," NYU mimeo.

LISE, J., C. MEGHIR, AND J.-M. ROBIN (2013): "Mismatch, sorting and wage dynamics," Discussion paper, National Bureau of Economic Research.

LIU, Q., G. MAILATH, A. POSTLEWAITE, AND L. SAMUELSON (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82(2), 541–587.

LORENTZ, G. G. (1953): "An Inequality for Rearrangements," *The American Mathematical Monthly*, 60, 176–179.

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

MANEA, M. (2014): "Steady-State in Matching and Bargaining," MIT, mimeo.

MCAFEE, R. P. (1993): "Mechanism Design by Competing Sellers," *Econometrica*, 61(6), 1281–1312.

MCCALL, J. (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics*, 84, 113–126.

McCall, J. J. (1965): "The Economics of Information and Optimal Stopping Rules," *The Journal of Business*, 38(3), 300–317.

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

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62, 157–180.

Milgrom, P., and R. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), 1089–1122.

Moen, E. (1997): "Competitive Search Equilibrium," *Journal of Political Economy*, 105, 385–411.

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

Morgan, P., and R. Manning (1985): "Optimal Search," *Econometrica*, 53, 923–944.

Mortensen, D., and R. Wright (2002): "Competitive Pricing and Efficiency in Search Equilibrium," *International Economic Review*, 58, 1–20.

Noldeke, G., and L. Samuelson (2014): "Investment and Competitive Matching," Econometrica, forthcoming.

Noldeke, G., and T. Troger (2009): "Matching Heterogeneous Agents with a Linear Search Technology," University of Basel Working Paper.

Osborne, M., and A. Rubinstein (1990): *Bargaining and Markets*. Academic Press, San Diego.

Peskir, G., and A. Shiryaev (2006): *Optimal Stopping and Free-Boundary Problems*. Birkhauser-Verlag, Germany.

Peters, M. (1984): "Bertrand Equilibrium with Capacity Constraints and Restricted Mobility," *Econometrica*, 52(5), 1117–27.

——— (1997): "A Competitive Distribution of Auctions," *Review of Economic Studies*, 64(1), 97–123.

Peters, M., and S. Severinov (1997): "Competition among Sellers Who Offer Auctions Instead of Prices," *Journal of Economic Theory*, 75(1), 141–179.

Pissarides, C. A., and B. Petrongolo (2001): "Looking into the Black Box: A Survey of the Matching Function," *Journal of Economic Literature*, 39(2), 390–431.

Postel-Vinay, F., and J. M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Pratt, J. W. (1964): "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 122–136.

Robin, J.-M., and J. Lise (2013): "The Macro-dynamics of Sorting between Workers and Firms," Discussion paper, Working paper.

Rosen, S. (1981): "The Economics of Superstars," *American Economic Review*, 71, 845–858.

Ross, S. (1983): *Introduction to Stochastic Dynamic Programming*. Academic Press, London.

Roth, A. (1989): "Two-Sided Matching with Incomplete Information about Others's Preferences," *Games and Economic Behavior*, 1(2), 191–209.

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

Rothschild, M. (1974): "Searching for the Lowest Price when the Distribution is Unknown," *Journal of Political Economy*, 82, 689–711.

Sasaki, H., and M. Toda (1996): "Two-Sided Matching Problems with Externalities," *Journal of Economic Theory*, 70, 93–108.

Sattinger, M. (1993): "Assignment Models and the Distribution of Earnings," *Journal of Economic Literature*, 31(2), 831–880.

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.

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

——— (2002): "A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Biased Technology," *Review of Economic Studies*, 69(2), 467–491.

Shimer, R. (2005): "The Assignment of Workers to Jobs in an Economy with Coordination Frictions," *Journal of Political Economy*, 113, 996–1025.

SHIMER, R., R. ROGERSON, AND R. WRIGHT (2005): "Search-Theoretic Models of the Labor Market," *Journal of Economic Literature*, 43, 959–988.

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

——— (2001): "Nonstationary Search," mimeo.

SHIRYAEV, A. N. (1978): *Optimal Stopping Rules*. Springer, New York.

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

SMITH, L. (1992): "Cross-Sectional Dynamics in a Two-Sided Matching Model," .

——— (1995): "A Model of Exchange Where Beauty is in the Eye of the Beholder," working paper, MIT.

——— (1997): "The marriage model with search frictions," .

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

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

SØRENSEN, M. (2007): "How smart is smart money? A two-sided matching model of Venture Capital," *The Journal of Finance*, 62(6), 2725–2762.

STIGLER, G. (1961): "The Economics of Information," *Journal of Political Economy*, 69, 213–225.

STOLE, L. A., AND J. ZWIEBEL (1996): "Intra-firm Bargaining under Non-binding Contracts," *Review of Economic Studies*, 63(3), 375–410.

TERVIO, 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.

TOPKIS, D. M. (1998): *Supermodularity and Complementarity*. Princeton University Press, Princeton.

VARIAN, H. R. (1999): "Economics and Search," Berkeley Working Paper.

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

Vince, A. (1990): “A Rearrangement Inequality and the Permutahedron,” *The American Mathematical Monthly*, 97, 319–323.

Weitzman, M. (1979): “Optimal Search for the Best Alternative,” *Econometrica*, 47, 641–654.