

Sorting through Search and Matching Models in Economics[†]

HECTOR CHADE, JAN EECKHOUT, AND LONES SMITH*

Toward understanding assortative matching, this is a self-contained introduction to research on search and matching. We first explore the nontransferable and perfectly transferable utility matching paradigms, and then a unifying imperfectly transferable utility matching model. Motivated by some unrealistic predictions of frictionless matching, we flesh out the foundational economics of search theory. We then revisit the original matching paradigms with search frictions. We finally allow informational frictions that often arise, such as in college-student sorting. (JEL C78, D82, D83, I23, J12)

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.

In this research thread of the assignment and matching literature, a dominant theme is positive sorting—the best are matched with one another, as are the next best, and so on. Indeed, we observe firms spending significant resources to hire 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—including the use of 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

*Chade: Department of Economics, Arizona State University. Eeckhout: Department of Economics UPF–ICREA–Barcelona GSE and University College London. Smith: Department of Economics, University of Wisconsin–Madison. We are grateful to the editor and five anonymous referees for their helpful comments and suggestions. We also benefited from comments from Manolis Galenianos, Shoshana Grossbard, Belen Jerez, Stephan Lauermaann, Patrick Legros, Ilse Lindenlaub, Jeremy Lise, Qingmin Liu, Benny Moldovanu, Sam Schulhofer-Wohl, and Randy Wright. An early version of parts of this draft was developed for lectures at the Centre for Market Design in Melbourne, Western Ontario, and the Becker Friedman Center in Chicago.

[†]Go to <https://doi.org/10.1257/jel.20150777> to visit the article page and view author disclosure statement(s).

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 by Gale and Shapley (1962) that developed an intuitive algorithm for generating 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 (1781) and Kantorovic (1942) 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 offers a way to formalize decentralized trade. In many formulations, 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 that began 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 bedfellows. 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 theme that has emerged.

Pursuant to this merger of the search and matching literatures, an alternative approach to search theory developed. Rather than explicitly and separately model the dynamic matching and price negotiation processes, in directed search, firms first set prices and then buyers direct their search, and finally meetings materialize. One slice of the literature 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. Another approach instead explicitly models the stockouts that emerge—students apply for slots at colleges, and are generally rejected. Formally, they are told that no slot is available. Directed search often also exploits the role of prices: sellers post prices first, upon

¹Sattinger (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.

which buyers make their purchase decisions, taking matching frictions into account.

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. By focusing on its main analytic idea, possibly by way of example, we present each as a teachable unit. We then touch on salient applications, for in recent years matching models have been applied broadly in economics. Examples without frictions include marriage markets, hierarchies, international trade, finance, CEO compensation, foreign direct investment, and development.² Matching models with frictions afford analyses of unemployment in the presence of sorting such as mismatch, the transmission of labor market risk, and the impact of macroeconomic fluctuations.³ Throughout the review, we refer to some of these papers, and highlight open-research agendas as a roadmap for future work.

Overall, this survey explores how two economic literatures, one in optimization and another in equilibrium, merged to create a cohesive equilibrium story of frictional markets.

2. *Frictionless Matching and Sorting*

To study how search frictions and/or information frictions shape matching outcomes in

economic environments, we first analyze the benchmark case without frictions.

Many important problems can be thought of as pairwise matching or an assignment of two groups of heterogeneous elements, either 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, locations to plants, buyers to sellers, countries to goods, etc. A distinctive feature is that agents or objects on each side are indivisible and frequently heterogeneous.

An important modeling choice in this framework is how payoffs are shared within a match. Two polar choices are perfectly transferable utility (perfect TU), where agents can freely transfer payoffs between them at a constant rate, and nontransferable utility (NTU), where either no transfers are possible or 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 imperfectly transferable utility (imperfect TU), where payoffs are neither fully transferable nor exogenously given.⁴ In the rest of the section, we provide a detailed analysis of the TU and NTU paradigms, as well as several economic applications.⁵ Our focus is on the conditions under which assortative matching obtains. Given our interest in sorting, we

²See, among many others, Garicano (2000); Sørensen (2007); Antras, Garicano, and Rossi-Hansberg (2006); Grossman, Helpman, and Kircher (2013); Grossman and Maggi (2000); Tervio (2008); Gabaix and Landier (2008); Guadalupe et al. (2014); and Akerberg and Botticini (2002).

³See, for example, Lise, Meghir, and Robin (2013); Lamadon (2014); Lise and Robin (2017).

⁴We break with tradition and use the more suggestive terminology in Noeldke and Samuelson (2015) instead of the well-established one from cooperative game theory used by Legros and Newman (2007). That is, instead of transferable utility we call it perfectly transferable utility; instead of nontransferable utility we call it imperfectly transferable utility; and instead of strictly nontransferable utility we call it simply nontransferable utility.

⁵We abuse the terminology slightly by calling the NTU case “frictionless,” since this feature can be due to some friction that prevents full transferability. What we mean here is that there are no search frictions (agents observe all potential partners) and no incomplete information about partners (agents observe their characteristics).

only discuss the NTU case results that shed light on sorting patterns, 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 Sorting with Perfectly 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).⁶ They 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 provided solid game theoretic foundations to the problem, deriving the optimal assignment, core allocations, and competitive equilibrium in a unified way.⁷ 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.

2.1.1 *The Basic Model*

We derive the main insights using a simple instance of the matching model with perfect

TU (i.e., an assignment game, using the terminology of Shapley and Shubik 1972), 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 woman i has a characteristic (type) $x_i \in [0, 1]$ and each man j has a characteristic $y_j \in [0, 1]$; for simplicity, assume that $x_1 < x_2 < \dots < x_N$ and $y_1 < y_2 < \dots < y_N$. If woman x_i marries man y_j , then they produce a positive output $f(x_i, y_j)$. We can thus identify each agent with his or her type. We make the innocuous assumption that single agents produce zero output. Crucially, agents' preferences are linear in money (perfect TU), and thus partners can freely share the match output produced using transfers.⁸

We answer 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?

2.1.2 *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.⁹ Formally, the optimal matching is the solution to the following maximization problem:

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

where the maximization is taken over all possible permutations $\pi : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$. By a well-known result in

⁶In the mathematics literature, the Monge–Kantorovich optimal-transport problem subsumes many frictionless matching problems. For an authoritative treatise of this problem, see Villani (2009).

⁷An excellent source for these results in both the finite and continuous case is Gretsky, Ostroy, and Zame (1999).

⁸Actually, 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.

⁹If 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.

rearrangement inequalities (e.g., see Vince 1990 and the main sorting result in the appendix of Becker 1973), the identity permutation $\pi(i) = i$ for all i solves problem (1) if f is supermodular^{10,11} on $[0, 1]^2$. This condition is not only *sufficient* but also *necessary* if the result must hold for all distributions of types for men and women. In short, PAM is 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 pairs the woman and man with the best characteristics, the second best woman with the second best man, and so on.

It is easy to see why supermodularity is sufficient for PAM—independently of the distribution of men's and women's types. Under any other assignment, there are two women, say i and i' with $i' > i$, respectively matched with two men j and j' with $j > j'$. The total output of these couples $f(x_i, y_j) + f(x_{i'}, y_{j'})$ is lower than $f(x_i, y_{j'}) + f(x_{i'}, y_j)$, by supermodularity. Hence, the planner can increase total output by assortatively rematching them.

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 type, and so on.

A useful alternative linear programming formulation of the optimal assignment problem is:

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

subject to $\sum_{j=1}^N \alpha_{ij} \leq 1$ for all i , $\sum_{i=1}^N \alpha_{ij} \leq 1$ for all j , and $\alpha_{ij} \geq 0$ for all i, j . Since α_{ij} is not merely zero or one, the problem permits fractional assignment of men and women. Koopmans and Beckmann (1957) and Shapley and Shubik (1972), however, showed that there is an optimal solution with $\alpha_{ij} \in \{0, 1\}$. If f is supermodular, then $\alpha_{ij} = 1$ when $i = j$, and PAM ensues; if not, then one can find a profitable rematching. A similar analysis holds for NAM.

2.1.3 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 fix ideas, assume f is strictly supermodular, and thus the optimal assignment is PAM. Let $i > i'$ and $j > j'$; by strict supermodularity, $f(x_i, y_j) + f(x_{i'}, y_{j'}) > f(x_i, y_{j'}) + f(x_{i'}, y_j)$. This inequality implies $f(x_i, y_j) - f(x_i, y_{j'}) > f(x_i, y_{j'}) - f(x_{i'}, y_{j'})$, so that the willingness to pay for the higher woman x_i is higher for the higher man y_j than for $y_{j'}$. So when competing for partners, j can outbid j' in the quest for i . Consequently, any "stable" outcome of the assignment game exhibits PAM. Alternatively, we could explore the performance of a competitive market where

¹⁰A real-valued function f on a lattice $X \subseteq \mathbb{R}^n$ (e.g., $[0, 1]^2$) is *supermodular* if $f(x' \vee x'') + f(x' \wedge x'') \geq f(x') + f(x'')$ for all x' and x'' in X , where $x' \vee x'' = \max\{x', x''\}$ and $x' \wedge x'' = \min\{x', x''\}$. If f is twice continuously differentiable, then this is equivalent to $\partial^2 f(x) / \partial x_i \partial x_j \geq 0$ for all $i \neq j$. The function is submodular if $f(x' \vee x'') + f(x' \wedge x'') \leq f(x') + f(x'')$ for all x' and x'' in X , and this is equivalent to $\partial^2 f(x) / \partial x_i \partial x_j \leq 0$ for all $i \neq j$ if f is twice continuously differentiable. These concepts are strict if the inequalities are strict.

¹¹This 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 f_1, \dots, f_n are real-valued functions on an interval I , then $\sum_i f_i(b_{n-i+1}) \leq \sum_i f_i(b_{\pi(i)}) \leq \sum_i f_i(b_i)$ for all sequences $b_1 \leq b_2 \leq \dots \leq b_n$ and all π if and only if $f_{i+1} - f_i$ is increasing on I for $1 \leq i < n$.

agents from each side take “partners’ prices” as given. The same logic reveals that any Walrasian equilibrium of this market delivers PAM when f is strictly supermodular, and NAM when f is strictly submodular.

Let v_1, \dots, v_N and w_1, \dots, w_N be the multipliers associated with the constraints after (2). Then the dual problem is¹²

$$(3) \quad \min_{v,w} \sum_{i=1}^N v_i + \sum_{j=1}^N w_j,$$

subject to $v_i + w_j \geq f(x_i, y_j)$, $v_i \geq 0$, and $w_j \geq 0$. From the linear programming duality theorem, the value to this problem is the same as that of problem (2), and thus of (1). Moreover, the α_{ij} s of problem (2) are the multipliers of problem (3). It follows that if (α, v, w) solves (2) and (3), then (i) $\sum_{i=1}^N \sum_{j=1}^N f(x_i, y_j) \alpha_{ij} = \sum_{i=1}^N v_i + \sum_{j=1}^N w_j$; and (ii) $v_i + w_j = f(x_i, y_j)$ for each pair (i, j) such that $\alpha_{ij} = 1$; and (iii) $v_i + w_j \geq f(x_i, y_j)$ for each pair (i, j) such that $\alpha_{ij} = 0$.

The triple (α, v, w) optimally matches the two populations and provides a division of the match output between partners that exhausts output, if we interpret v_i and w_j as the wages of woman i and man j . That triple is also a *stable matching* of the assignment game, for no man and woman not originally matched can profitably block the assignment given (iii), since the sum of the utilities in their original matches more than exhausts the match output if they rematch. Moreover, (iii) implies that no coalition of men and women can improve upon (α, v, w) . Hence, the solution of the dual problem characterizes the *core* of the assignment game.

We now decentralize the optimal/core matching as a Walrasian equilibrium. Let women take the men’s wages w_1, \dots, w_N as given. Then woman i chooses the man j that

maximizes $f(x_i, y_j) - w_j$. By construction of the core allocation, $v_i = f(x_i, y_j) - w_j$ if $\alpha_{ij} = 1$. That is, v_i is the wage that woman i obtains in the core allocation. Also, for any other man j' with $\alpha_{ij'} = 0$, we have $v_i \geq f(x_i, y_j) - w_j$, or $f(x_i, y_j) - w_j \geq f(x_i, y_{j'}) - w_{j'}$. Hence, when confronted with men’s wages w_1, \dots, w_N , woman i optimally selects the same partner as in the core allocation.

Since one can perform this analysis for men, the optimal matching can be decentralized as a Walrasian equilibrium of the marriage market. The wage of a woman in this market depends only on her type, and not on that of the man she matches with. For these wages are formally the utility payoffs of each woman in the core allocation. Since these wages are the multipliers of the linear programming constraints, each can be interpreted as the shadow value of adding a woman to the matching market—hence the dependence only on the woman’s type.

This analysis is valid with an unequal number of men and women. In that case, some agents on the long side of the market will remain single. Similarly, if we assume that agents can produce some output as singles, then match surplus will be its output minus the sum of the singles outputs. In this case, some agents may remain single at the optimal matching. While we have assumed only one agent of each type, our analysis is valid with different discrete distributions of types on each side. In this case, PAM matches agents from the top types down, respecting the measure of agents of each type in the population until the populations are exhausted.

2.1.4 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 enables the use of calculus in the derivation of equilibria and their properties.

¹²See Chvatal (1983), chapter 5, for a derivation and for the proof of the duality theorem of linear programming.

We will illustrate this convenient feature below with some important economic applications of the frictionless matching paradigm.

Assume an equal unit mass continuum of men and women. Each female has a type $x \in [0, 1]$ drawn from a strictly increasing and continuously differentiable cdf G with positive density g . Similarly, each man has a type $y \in [0, 1]$, with cdf H and density h . We can define a (pure) matching as a function $\mu : [0, 1] \rightarrow [0, 1]$ that is *measure preserving*, i.e., matching equal measures of men and women. For instance, PAM requires that $G(x) = H(\mu(x))$ for all x . Hence, $\mu(x) = H^{-1}(G(x))$ is strictly increasing and $\mu'(x) = g(x)/h(\mu(x)) > 0$. Under NAM, $G(x) = 1 - H(\mu(x))$ for all x , and thus $\mu(x) = H^{-1}(1 - G(x))$, with $\mu'(x) = -g(x)/h(\mu(x)) < 0$.

The match output of a woman with type x with a man with type y is $f(x, y)$, now assumed twice continuously differentiable. A continuous version of the above rearrangement inequality (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 continuous models (Gretsky, Ostroy, and Zame 1992, 1999).

We now derive the Walrasian equilibrium and deduce the sorting pattern that ensues when production $f(x, y)$ is supermodular or submodular. In so doing, we draw a simple connection between matching models and a basic monotone comparative statics result (Topkis 1998).

Consider a man of type y facing a wage profile $v(x)$ for women x . He seeks the woman x that maximizes his payoff:

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

Now, if f is strictly supermodular (i.e., $f_{xy} > 0$), then the objective function satisfies the strict single-crossing property in (x, y) .¹³ Hence, in *any* solution to this problem, men with higher y choose women with higher x .¹⁴ So if a Walrasian equilibrium exists, then it *must* exhibit PAM. This provides an alternative view of the sufficiency of supermodularity for PAM.¹⁵ From the above measure-preserving (market-clearing) property, the only candidate for equilibrium matching is $y = \mu(x) = H^{-1}(G(x))$ or $x = \mu^{-1}(y)$. This must satisfy the first-order condition $v'(x) = f_x(x, \mu(x))$, and thus:

$$(4) \quad v(x) = v_0 + \int_0^x f_x(s, \mu(s)) ds,$$

where v_0 is a constant of integration. (One can show that global optimality holds.) Hence, if f is strictly supermodular, then $\mu(x) = H^{-1}(G(x))$ and (4) constitute a Walrasian equilibrium and exhibits PAM. Clearly, each man y in equilibrium obtains $w(y) = f(\mu^{-1}(y), y) - v(\mu^{-1}(y))$. A similar analysis can be done for f strictly submodular and NAM.

¹³A function $z : X \times [t, \bar{t}] \rightarrow \mathbb{R}$, 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 $z(x'', t'') - z(x', t'') > 0$.

¹⁴Apply theorem 4' in Milgrom and Shannon (1994) (quasi-supermodularity in y trivially holds in this problem).

¹⁵In 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).

2.2 Applications of Frictionless Sorting with Perfectly Transferable Utility

2.2.1 The O-Ring Production Function

In an application of Becker's marriage model, Kremer (1993) explores a "weakest link" production model that naturally generates supermodularity. There are n tasks, each performed by a worker. Each worker i has a type $x_i \in [\underline{x}, 1]$, $0 < \underline{x} < 1$, drawn from a continuous density g . This type is the probability that the worker successfully performs the task. Production happens when all n workers succeed in their tasks. That is, the expected output of a firm is $nB \prod_{i=1}^n x_i$, where $B > 0$ is the output per worker if all perform their tasks successfully. There is a unit mass of workers and a limited number of identical potential firms that each hires n workers from a competitive labor market. Firms take the wage function $\omega : [\underline{x}, 1] \rightarrow \mathbb{R}_+$ as given. For simplicity, we assume that labor is the only factor of production.

In the matching problem, each firm hires n possibly heterogeneous workers to maximize expected profits:

$$(5) \quad \max_{\{x_i\}_{i=1}^n} nB \prod_{i=1}^n x_i - \sum_{i=1}^n \omega(x_i).$$

Since $\partial^2(\prod_{i=1}^n x_i)/\partial x_j \partial x_k > 0$ for all $j \neq k$, the expected output is strictly supermodular in (x_1, x_2, \dots, x_n) and the equilibrium exhibits PAM: In other words, all the workers employed by any given firm have the same type x . The first order condition for (5) in x_i evaluated at $x_i = x$ yields $\omega'(x) = nBx^{n-1}$. Consequently, $\omega(x) = Bx^n + \omega_0$, where the constant ω_0 is pinned down by the firm's zero profit condition, or $\omega_0 = 0$. All told, each firm hires workers of the same skill x and pays them the wage $\omega(x) = Bx^n$, equally dividing the expected output among its workers.

Kremer (1993) shows how the model sheds light on several stylized facts, such as the

positive correlation among wages of workers in different occupations within a firm.

2.2.2 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. They assume a unit mass continuum of CEOs and of firms. The CEO talent x has a differentiable cdf G , and the firm size y has a differentiable cdf H . They (and Tervio 2008) identify each CEO of talent x with his quantile rank $i = G(x)$; since G is differentiable, there is a smooth relationship between i and $X(i)$, with $X'(i) > 0$. Likewise, associate firms with their quantile rank $j = H(y)$, where firm $Y(j)$ smoothly increases: $Y'(j) > 0$. They assume that the revenue when CEO i is matched with firm j is $CY^d(j)X(i)$, for constants $C, d > 0$, from which the CEO gets paid $\omega(i)$.

In a Walrasian equilibrium, firm j maximizes $CY^d(j)X(i) - \omega(i)$ over $i \in [0, 1]$. Since the objective function is strictly supermodular in (i, j) , the equilibrium exhibits PAM—to wit, $i = j$, for all j . The first-order condition (FOC) yields $\omega'(i) = CY^d(j)X'(i)$, and hence

$$\omega(i) = \int_0^i CY^d(s)X'(s) ds + \omega(0).$$

This wage function, along with PAM, constitutes the Walrasian equilibrium of the model (modulo the constant $\omega(0)$).

To calibrate the model, Gabaix and Landier (2008) posit a Pareto firm size distribution $H(y) = 1 - (y/y)^{1/\tau}$, which yields $Y(j) = \underline{y}(1-j)^{-\tau}$. Meanwhile, inspired by extreme value theory, they posit that the "spacing function" for CEO talents $X'(i)$ satisfies $X'(i) = K(1-i)^{\nu-1}$, for constants K and ν . The wage function is

$$\omega(i) = \frac{CK\underline{y}^d}{\tau d - \nu} \left((1-i)^{-(\tau d - \nu)} - 1 \right) + \omega(0).$$

Assuming $\tau d > \nu$, Gabaix and Landier (2008) calibrate the model and analyze several features of CEO pay and its increase in recent years in the United States (see also Tervio 2008). They show that the model exhibits a “superstar” property (Rosen 1981): small differences in talent can have a drastic impact in pay at the top—i.e., CEOs with rank close to one, here.¹⁶ Also, the increase in size of large firms in recent years can account for a large fraction of the increase in CEO pay.

2.2.3 Matching Principals and Agents

In the principal–agent model, the principal hires an agent to perform a task. Since the agent’s actions are unobservable, the contract is based on a stochastic signal, such as output, that is correlated with those actions. Akerberg and Botticini (2002) convincingly argue that accounting for endogenous matching of principals and agents is important when testing predictions of 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 (proxying for risk aversion), which affects the contract form used (share contracts or fixed-rent contracts).

Serfes (2005) explores a tractable matching model of 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 justification of Holmstrom and Milgrom 1987), and this turns the model into a matching problem with perfect TU. Without this assumption,

we will see in section 2.4 that moral hazard leads to imperfect TU.

The type x of a principal is the variance of her output, while the type y of the agent is his coefficient of absolute risk aversion. A match of principal x and any agent generates stochastic output $q = e + \varepsilon$, where $\varepsilon \sim N(0, x)$, and the agent’s effort $e \geq 0$ incurs disutility $ke^2/2$. With CARA utility function, agent y ’s expected utility is $1 - e^{-y(I - ke^2/2)}$, where I is income. The optimal contract is linear in output: $I(q) = I_0 + bq$, where I_0 is a base wage and b the incentive power. By Holmstrom and Milgrom (1987), the optimal contract sets $b = 1/(1 + kyx)$ and yields the principal expected profit $f(x, y) = 1/(2k(1 + kyx))$. (The base wage I_0 is irrelevant with CARA utility.) Finally, we can check that $f_{xy} < 0$ if and only if $yx < 1/k$, and this holds for all $x \in [\underline{x}, \bar{x}]$ and $y \in [\underline{y}, \bar{y}]$ if $\bar{y}\bar{x} < 1/k$. So NAM emerges if this condition holds, and PAM if $\underline{y}\underline{x} > 1/k$.

This principal–agent model predicts that b is decreasing in x , namely, 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 is optimal: for in this case, high variance x principals are matched with less risk averse agents $y = \mu(x)$. Since μ is strictly decreasing, the incentive power $b = 1/(1 + k\mu(x)x)$ could increase in x if matching is endogenous. If $\mu(x)x$ decreases in x , which depends on the distributions G and H , then b increases in x , as the evidence shows.

2.3 Frictionless Sorting with Imperfectly Transferable Utility

2.3.1 Background

When partners cannot transfer utility one for one, we say that there is imperfect TU.

¹⁶Differentiation reveals that $\omega(i)$ is strictly increasing and strictly convex, with $\omega''(i)$ going to infinity as i goes to one. Thus, CEOs matched with large firms receive increasingly larger pay near the top.

Many economic environments of interest fall into this category, such as risk-sharing problems or matching problems where moral hazard is present.

In one extreme case, NTU, partners cannot transfer utility at all. For instance, the output $f(x, y)$ from a match between x and y may be divided according to some fixed sharing rule; more generally, assume that if x matches with y , then x obtains utility $f_1(x, y)$ and y obtains $f_2(x, y)$, as done in Smith (1997). Actually, matching models without transfers have been extensively studied since Gale and Shapley (1962)—see Roth and Sotomayor (1990). 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 their respective partners in the candidate allocation. The fundamental result is that a stable matching exists. The existence proof is constructive by means of the *deferred acceptance 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 ones.

Using a cardinal representation of preferences, Becker (1973) noted that if each agent strictly prefers a partner with a higher type, then PAM emerges under NTU. In our notation, we need f_1 to be strictly increasing in y and f_2 in x . To see this, consider two women and two men, 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 $f_1(x', y') > f_1(x', y'')$ and $f_2(x', y') > f_2(x'', y')$, due to the monotonicity in partner's type. Similarly, NAM emerges if one of the partial derivatives is positive and the other one is negative. So while Becker did not cite Gale and Shapley (1962), he intuitively grasped their pairwise stability notion in this case.

Recently, Legros and Newman (2010) have shown that PAM does not require monotonicity in the partner's type. 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_i , $i = 1, 2$, not being increasing in partner's type (see Legros and Newman 2010 for an example).

Thus far we have explored assortative matching in both the perfect TU and the NTU cases. Notice that the Pareto frontier of payoffs achievable by a pair of matched agents is linear in the perfect TU case (with constant slope that is independent of the agents' types), and collapses to a point in the NTU case. What about typical intermediate cases where agents can transfer utility but not at a constant rate, so that the Pareto frontier is decreasing but neither linear nor a single point (see figure 1)? Legros and Newman (2007) address this case. We now provide a detailed summary of their main insights and several illustrative applications.

2.3.2 The General Model

There are two populations, women and men, indexed by 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

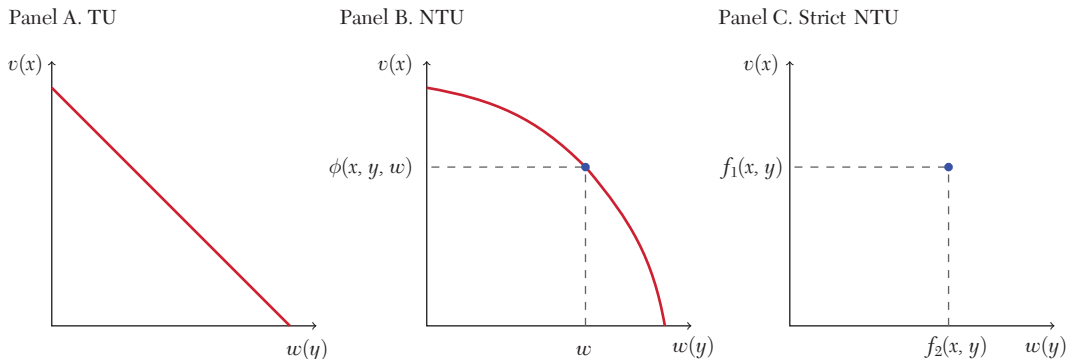


Figure 1. Examples of Pareto Frontiers for TU and NTU Cases

Note: We depict examples of the Pareto frontiers respectively for perfect TU; the intermediate imperfectly TU case introduced in Legros and Newman (2007) (with a decreasing nonlinear frontier); and NTU.

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—the “equal-treatment” property deduced in Legros and Newman (2007).

To capture a utility frontier for each pair of agents, let $\phi(x, y, w)$ be the maximum utility that x generates when matched with y , if y receives utility w . Since no agent receives less than their autarchy payoff in equilibrium, $\phi(x, y, 0)$ is the maximum that x can obtain when matched with y . We assume that $\phi(x, y, w)$ is strictly decreasing in w when positive. In later applications, we derive this Pareto frontier from assumptions on technology and preferences. Let $\psi(y, x, v)$ be the maximum utility of y when matched with x who receives utility v . This is the partial inverse of $\phi(x, y, \cdot)$, in the sense that $\phi(x, y, \psi(y, x, v)) = v$ for all $v \in [0, \phi(x, y, 0)]$.

The equilibrium concept is the core of this assignment game—namely, a matching function μ and utility functions v for types x and w for types y , that satisfy the following properties: (i) feasibility of v and w with respect to μ , so $v(x) \leq \phi(x, y, w(\mu(x)))$

and $w(\mu(x)) \leq \psi(y, x, 0)$ for all x and y ; and (ii) stability of μ with respect to v and w , so that there is no pair of agents with x and y with $w > w(y)$ and $\phi(x, y, w) > v(x)$. This subsumes the perfect TU model with $\phi(x, y, w) = f(x, y) - w$, and subsumes the NTU model with $\phi(x, y, w) = f_1(x, y) \mathbf{1}_{w=f_2(x,y)}$.

2.3.3 Generalized Increasing Differences

First consider supermodular production with perfect TU, so that PAM is optimal. If $x > x'$ and $y > y'$, then y can weakly outbid y' in the competition for x , or $f(x, y) - f(x', y) \geq f(x, y') - f(x', y')$. Rewrite this increasing differences condition as

$$(6) \quad f(x, y) - [f(x', y) - v] \geq f(x, y') - [f(x', y') - v],$$

where x' obtains utility v . Inequality (6) holds for *any* level of utility v . By this increasing difference condition, higher types choose higher-matching partners, and thus PAM is

optimal. Extending (6) to our richer class of match payoffs in the general NTU case, y can weakly outbid y' in the competition for x when the outside option is a match with x' who earns utility v if and only if

$$(7) \quad \phi(x, y, \psi(y, x', v)) \\ \geq \phi(x, y', \psi(y', x', v)).$$

This reduces to (6) with perfect TU, as $\phi(x, y, \psi(y, x', v)) = f(x, y) - \psi(y, x', v) = f(x, y) - [f(x', y) - v]$, and similarly the right sides of (6) and (7) coincide. There are *generalized increasing differences* if (7) holds whenever $x > x'$, $y > y'$, and v is feasible, namely, $v \in [0, \phi(x', y, 0)]$.

Legros and Newman (2007) prove that when (7) holds, all equilibria are payoff equivalent to PAM. Similarly, given the reverse condition, *generalized decreasing differences*, all equilibria are payoff equivalent to NAM. These conditions are necessary if PAM must hold for any type distribution. To wit, as with Becker's supermodularity and submodularity conditions, generalized increasing and decreasing differences are the necessary and sufficient *distribution-free* conditions for PAM and NAM, respectively. This powerful result nests perfect TU and NTU as special cases, and thus greatly enlarges the set of economic applications for which we can assert PAM and NAM.

Condition (7) can be usefully simplified. Label $w = \psi(y, x', v)$ and $w' = \psi(y', x', v)$. Then $v = \phi(x', y, w) = \phi(x', y', w')$, so that type x' obtains the same utility v either in a match with y , paying him w , or in a match with y' , paying him w' . Then (7) asserts $\phi(x, y, w) \geq \phi(x, y', w')$, so that $x > x'$ obtains more utility by matching with y than with y' , if she must pay them the same. Put differently, if x' is indifferent between (y, w) and (y', w') , then $x > x'$ prefers (y, w) . Formally, if $x > x'$ then $\phi(x', y, w) = \phi(x', y', w') \Rightarrow \phi(x, y, w) \geq \phi(x, y', w')$.

Finally, we can write this in a standard single-crossing form with an inequality premise—that for any (y, w) and (y', w') :¹⁷

$$(8) \quad \phi(x', y, w) \geq \phi(x', y', w') \\ \Rightarrow \phi(x, y, w) \geq \phi(x, y', w')$$

for all $y > y'$ and $x > x'$.

We next provide a new differential version of (8). Let ϕ be twice continuously differentiable. Then under the regularity assumptions in theorem 3 in Milgrom and Shannon (1994),¹⁸ the single crossing property (8) is equivalent to the *Spence–Mirrlees* condition, that the marginal rate of substitution $-\phi_y/\phi_w$ between y and w increases in one's type x . Since ϕ is twice differentiable, we have

$$(9) \quad \phi_{xy}(x, y, w) \\ \geq \frac{\phi_y(x, y, w)}{\phi_w(x, y, w)} \phi_{xw}(x, y, w)$$

since $\phi_w < 0$. In other words, a high type x is willing to pay more of w for an increment in his partner's type y . Consequently, the indifference curves in (y, w) -space single cross as x changes.

We are now equipped to give a simple smooth argument for why (9) leads to PAM with a continuum of agents. Mimicking logic familiar in screening models (Fudenberg

¹⁷ Indeed, (8) implies (7), since equality is a special case of the left side of (8). To see that (7) is equivalent to (8), recall that ϕ strictly falls in the partner's utility. Let $v_1 \equiv \phi(x', y, w) \geq \phi(x', y', w') \equiv v_2$, so that $w = \psi(y, x', v_1)$ and $w' = \psi(y', x', v_2)$. Since ψ is also decreasing in v , we have $\psi(y', x', v_2) \geq \psi(y', x', v_1)$. Along with (7), this yields

$$\phi(x, y, w) = \phi(x, y, \psi(y, x', v_1)) \\ \geq \phi(x, y', \psi(y', x', v_1)) \geq \phi(x, y', \psi(y', x', v_2)) \\ = \phi(x, y', w').$$

¹⁸ See also theorem 2.1 in Edlin and Shannon (1998), and theorem 3 in Athey, Milgrom, and Roberts (1998).

and Tirole 1991, chapter 7), assume a PAM allocation $y = \mu(x)$, with μ strictly increasing and smooth. Consider the problem that type x solves when matched with a type y agent who earns $w(y)$. If this is an equilibrium, the FOC for type x 's optimization $\max_y \phi(x, y, w(y))$ holds when evaluated at $y = \mu(x)$, or

$$(10) \quad \phi_y(\mu^{-1}(y), y, w(y)) \\ + \phi_w(\mu^{-1}(y), y, w(y))w'(y) = 0.$$

Next, the solution to (10) is a global maximum. For consider any other type, say $\hat{y} > y$. Then

$$\begin{aligned} & \phi(x, \hat{y}, w(\hat{y})) - \phi(x, y, w(y)) \\ &= \int_y^{\hat{y}} [\phi_y(x, s, w(s)) \\ & \quad + \phi_w(x, s, w(s))w'(s)] ds \\ &= \int_y^{\hat{y}} (-\phi_w(\mu^{-1}(y), s, w(s))) \\ & \quad \times \left(-\frac{\phi_y(\mu^{-1}(y), s, w(s))}{\phi_w(\mu^{-1}(y), s, w(s))} - w'(s) \right) ds \\ &\leq \int_y^{\hat{y}} (-\phi_w(\mu^{-1}(y), s, w(s))) \\ & \quad \times \left(-\frac{\phi_y(\mu^{-1}(s), s, w(s))}{\phi_w(\mu^{-1}(s), s, w(s))} - w'(s) \right) ds, \end{aligned}$$

where the inequality follows from $-\phi_y/\phi_w$ increasing in x and the last line vanishes by the FOC (10). Hence, $\phi(x, y, w(y)) \geq \phi(x, \hat{y}, w(\hat{y}))$, so x does not have incentives to deviate up. A similar argument shows that x does not have incentives to choose $\hat{y} \leq y$. Hence, choosing y is a global optimum for x .

In other words, *the differential inequality (9) is an easily-checked condition for PAM (NAM) in imperfect TU matching models.*

We repeatedly exploit it in later applications. This inequality reveals a tension between complementarity in one's own type and one's partner's type (ϕ_{xy}) or in a partner's utility (ϕ_{xw}). The latter reflects whether transferring utility to a partner becomes easier as one's type increases. In the perfect TU case, the second complementarity is absent, and (9) collapses to Becker's condition $f_{xy} \geq 0$, since $f_{xy} \equiv \phi_{xy}$ with perfect TU. Next, assume that ϕ increases in one's partner's type ($\phi_y > 0$). Since $\phi_w < 0$, inequality (9) and therefore PAM ensues if $\phi_{xy} \geq 0$ and $\phi_{xw} \geq 0$ while NAM obtains if $\phi_{xy} \leq 0$ and $\phi_{xw} \leq 0$. Intuitively, type complementarity abets sorting, whereas increasing difficulty of transferring utility to one's partner (namely $\phi_{xw} \leq 0$, recalling that $\phi_w < 0$) discourages sorting. But if transferring utility becomes easier with higher types ($\phi_{xw} \geq 0$), then the two effects reinforce each other and PAM obtains.

2.4 Applications of Frictionless Sorting with Nontransferable Utility

2.4.1 Matching Principals and Agents

We revisit this application from section 2.2, but now without assuming CARA and linear contracts. An agent's characteristic y is her initial wealth, which affects her risk attitude. Her utility is $V(y + I) - e$, where $e \in \{0, 1\}$ is the disutility of exerting effort. Per usual, we assume $V' > 0 > V''$, with a decreasing coefficient of absolute risk aversion $-V''/V'$. The agent's effort is unobservable, while his output $q \in \{\bar{q}, q\}$ is observable. If the agent exerts effort $e = 0$, then output is low $q = \underline{q}$ for sure. If she exerts effort $e = 1$, then output $q = \bar{q}$ with probability $x > 0$. Principals differ in x , a riskiness measure. We assume that $\bar{q} - q$ is large enough so that principals always want to implement $e = 1$.

A contract is a pair (\underline{I}, \bar{I}) of wages contingent on outputs \underline{q} and \bar{q} . If principal x

matches with agent y , whose reservation utility is w , the resulting contracting problem is $\phi(x, y, w) = \max_{\underline{I}, \bar{I}} x(\bar{q} - \bar{I}) + (1 - x) \times (\underline{q} - \underline{I})$ subject to an incentive constraint $xV(y + \bar{I}) + (1 - x)V(y + \underline{I}) - 1 \geq V(y + \underline{I})$ and a participation constraint $xV(y + \bar{I}) + (1 - x)V(y + \underline{I}) - 1 \geq w$. Both constraints bind at the optimum: if either is slack, then wages can be reduced to strictly raise the principal's expected profit. Solving the two binding constraints yields $\bar{I} = Z(w + x^{-1}) - y$ and $\underline{I} = Z(w) - y$, where $Z \equiv V^{-1}$. So:

$$\begin{aligned} \phi(x, y, w) &= x\left(\bar{q} - Z\left(w + \frac{1}{x}\right)\right) \\ &\quad + (1 - x)\left(\underline{q} - Z(w)\right) - y. \end{aligned}$$

Notice that ϕ is strictly concave in w , as in figure 1. Moreover, $\phi_{xy}(x, y, w) = 0$ and

$$\begin{aligned} (11) \quad \phi_{xw}(x, y, w) &= \frac{1}{x} Z''\left(w + \frac{1}{x}\right) - Z'\left(w + \frac{1}{x}\right) + Z'(w). \end{aligned}$$

If Z' is convex,¹⁹ then the first term in (11) dominates the last two, and so $\phi_{xw} \geq 0$. Since $\phi_{xy} = 0$, we obtain PAM by (9). That is, agents with high initial wealth and thus low risk aversion matched with principals with safer output distributions. As Legros and Newman (2010) point out, the sorting pattern emerges despite the lack of any complementarities between x and y . If instead Z' is concave, then $\phi_{xw} \leq 0$ and there is NAM.

¹⁹Since $Z' = 1/V'$, this is convex if and only if $-V'''/V'' \leq -3V''/V'$. Many standard utility functions satisfy this condition, including CARA (exponential) utility as well as $V(I) = I^\xi$ with $\xi \leq 0.5$. This commonly emerges in principal-agent models with moral-hazard and wealth effects. Conversely, Z' is concave, say, for $V(I) = I^\xi$ with $\xi > 0.5$.

2.4.2 Marriage and Risk Sharing

Consider a marriage market where men and women with different wealths (and thus different risk aversion levels) marry to share risk.²⁰ If a woman of wealth x marries a man of wealth y , then they share the risk embedded in a gamble whose payoff $q \in [0, 1]$ has a continuous distribution Γ . The utility function of women is $\log(1 + x + I)$ and of men is $\log(1 + y + I)$, where I is income. Efficient risk sharing solves the following problem:²¹

$$\begin{aligned} (12) \quad \phi(x, y, w) &= \max_{I(\cdot)} \int_0^1 \log(1 + y + q - I(q)) d\Gamma(q) \end{aligned}$$

subject to

$$\int_0^1 \log(1 + x + I(q)) d\Gamma(q) \geq w,$$

where $I(q)$ is the woman's share of q and $q - I(q)$ the man's share.

Intuitively, the constraint binds at the optimum. Maximizing pointwise, we obtain $I(q) = (-(1 + x) + (1 + y + q)\zeta)/(1 + \zeta)$, where ζ is the Lagrange multiplier. Inserting $I(q)$ into (12) and solving for ζ yields

$$\begin{aligned} (13) \quad \phi(x, y, w) &= \log\left(1 - e^{v - \int_0^1 \log(2+x+y+q)d\Gamma(q)}\right) \\ &\quad + \int_0^1 \log(2 + x + y + q) d\Gamma(q). \end{aligned}$$

²⁰For a survey of the literature, see the book by Browning, Chiappori, and Weiss (2014).

²¹This is a general version of an example in Legros and Newman (2010), section 5.1 (see also Chiappori and Reny 2015 and Schulhofer-Wohl 2006). We use our differential version of their condition to readily check for NAM.

We claim that NAM emerges, i.e., wealthy women who are less risk averse marry poor men who are more risk averse. Intuitively, a more risk averse individual is willing to pay more for insurance. To wit, a highly risk averse man can outbid a less risk averse one for a wealthy woman.

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

2.4.3 Matching in Large Firms

The one-to-one matching paradigm misses an important feature of actual labor markets. Labor market realism demands that firms can hire many workers. In principle, one could reinterpret a firm as a series of independent jobs that do not affect any other job's productivity. Often, however, there are complementarities between jobs.

Gale and Shapley (1962) highlighted the importance of many-to-one matching with their college admissions problem. A more relevant setup for labor market applications is one with a large number of firms and transfers (wages). In the O-ring technology in Kremer (1993), firms employ a given fixed number of workers.²² Kelso and Crawford (1982) develop a general many-to-one matching model of firms to any number or type of workers. Existence of equilibrium, however, is not guaranteed, as the following simple example illustrates. Consider two workers $x = 1, 2$ and two firms $y = 1, 2$. Let $f(y, \{X\})$ be the output of firm y matched with a set of workers $\{X\} \in \{\{1\}, \{2\}, \{1, 2\}, \{\emptyset\}\}$.

Assume the following technology for producing output:

$$f(1, \{\emptyset\}) = 0, f(1, \{1\}) = 4, f(1, \{2\}) = 1, f(1, \{1, 2\}) = 10 \\ f(2, \{\emptyset\}) = 0, f(2, \{1\}) = 8, f(2, \{2\}) = 5, f(2, \{1, 2\}) = 9.$$

Then all possible allocations with full employment are blocked. For instance, if firm 1 hires both workers, then the total output is ten at firm 1 and zero at firm 2. Worker 1 must earn at least eight, namely, what firm 2 is willing to offer when it hires only him. Likewise, worker 2 must earn at least five. But this total-wage bill overexhausts firm 1's total output. So this allocation is not stable, and all other allocations are also blocked.

The nonexistence is driven by worker complementarities. Specifically, if firm 1 hires worker 2 when it already employs worker 1, then worker 1's productivity rises, and she can thus command a higher wage. The *gross substitutes* condition in Kelso and Crawford (1982) precludes this possibility, and secures for existence. It asserts that if wages increase for some workers, then the firm will not drop from its labor force any worker whose wage did not increase. Additively separable production functions easily obey this condition, since a worker's productivity does not depend on her coworkers. Gross substitutes is sufficient and almost necessary, as it leaves little room for complementarities (Hatfield and Milgrom 2005 and Hatfield and Kojima 2008). Kelso and Crawford (1982) also provide an algorithm that finds the equilibrium allocation and wages. It is a variation on the deferred acceptance algorithm of Gale and Shapley (1962) (see section 2.3).²³

Modeling firms with an endogenous size and labor force composition with complementarities has proven difficult, given the gross substitutes condition. For instance, in

²²For another application where firms hire teams with a fixed number of experts, see Chade and Eeckhout (forthcoming).

²³Hatfield and Milgrom (2005) apply their ascending-bid auction to analyze package auctions.

a model of the span of control by management, Lucas (1978) focused on the *intensive margin* decision of how many workers to hire, but ignored composition. A more productive management hires a larger work force, increasing its span of control. This model shed light on the distribution of firm size. In reality, however, management at a firm also faces an *extensive-margin* decision about workers' composition; this is precisely the focus of matching models.

Eeckhout and Kircher (2012) develop a tractable model with size and composition margins. We present a slightly simplified version that yields their sorting condition. Assume match output is given by a production function $F(x, y, l_x, r_x)$, where y is the firm type, x the worker type, l_x the labor force size of type x , and r_x the resources the firm dedicates to workers of type x . That is, a firm of type y hires a quantity or measure l_x of workers of a *common* type x at a wage $\omega(x)$ per worker.²⁴ Since a firm chooses both the type and the number of workers, the model embeds size and composition margins: It is inspired by Becker's TU pairwise matching model, but instead allows for variable sizes of one side of the match. It is further assumed that F is strictly concave in l and r , and exhibits constant returns in l, r , so that $F(x, y, l, r) \equiv rf(x, y, \theta)$, where $f(x, y, \theta) \equiv F(x, y, \theta, 1)$, and where $\theta = l/r$ is the labor resources ratio. The problem of firm y reduces to the maximization problem: $\max_{x, \theta} f(x, y, \theta) - \theta\omega(x)$. Let us solve first the maximization in θ , for each x . The FOC $f_\theta(x, y, \theta) = \omega(x)$ yields the unique maximizer $\theta(x, y, \omega(x))$. Recalling that F is strictly concave in its third argument, we have $f_{\theta\theta} < 0$. Define $\psi(y, x, \omega) \equiv f(x, y, \theta(x, y, \omega)) - \theta(x, y, \omega)\omega$. Then the

optimization problem reduces to maximizing $\psi(y, x, \omega(x))$. Since this has the NTU structure analyzed in section 2.3, PAM or NAM ensues depending on whether $\psi(y, x, \omega)$ has $-\psi_x/\psi_\omega$ globally increasing or globally decreasing in y for all (y, x, ω) , i.e., $\psi_{xy} - (\psi_x/\psi_\omega)\psi_{\omega y} \geq 0$ from (9). Now, $\psi_x = f_x$, $\psi_\omega = -\theta$, and $\psi_y = f_y$ by the envelope theorem. Also, for fixed ω in $\psi(x, y, \omega)$, it follows from implicitly differentiating the FOC $f_\theta(x, y, \theta) = \omega$ that $\theta_y = -f_{y\theta}/f_{\theta\theta}$ and $\theta_x = -f_{x\theta}/f_{\theta\theta}$. Therefore, differentiating $\psi_\omega = -\theta$ yields $\psi_{\omega y} = -\theta_y$, while differentiating $\psi_x = f_x$ gives $\psi_{xy} = f_{xy} + f_{x\theta}\theta_y = f_{xy} - f_{x\theta}f_{y\theta}/f_{\theta\theta}$. Next,

$$\psi_{xy} = f_{xy} + f_{y\theta}\theta_x = f_{xy} - \frac{f_{y\theta}f_{x\theta}}{f_{\theta\theta}}$$

and

$$\psi_{x\omega} = -\frac{f_{x\theta}}{f_{\theta\theta}}.$$

Since $f_{\theta\theta} < 0$, the condition for PAM reduces to $f_{xy}f_{\theta\theta} - f_{y\theta}f_{x\theta} + (f_x f_{y\theta}/\theta) \leq 0$, for

$$\begin{aligned} (14) \quad 0 &\leq \psi_{xy} - (\psi_x/\psi_\omega)\psi_{\omega y} \\ &= f_{xy} - \frac{f_{y\theta}f_{x\theta}}{f_{\theta\theta}} + (f_x/\theta)\frac{f_{y\theta}}{f_{\theta\theta}} \\ &= \left(f_{xy}f_{\theta\theta} - f_{y\theta}f_{x\theta} + \frac{f_x f_{y\theta}}{\theta}\right)/f_{\theta\theta}, \end{aligned}$$

which is the inequality in Eeckhout and Kircher (2012). The reverse inequality yields NAM.

We now express this in terms of $F(x, y, l, r)$. Since F is homogeneous of degree one in (l, r) , F_l is homogeneous of degree zero, and so $lF_{ll} + rF_{lr} = 0$ by Euler's theorem, i.e., $-F_{lr} = \theta F_{ll}$. Additionally, F_x is homogeneous of degree one in (l, r) , whence $F_x = \theta F_{xl} + F_{xr}$, by Euler's theorem. So the

²⁴Eeckhout and Kircher (2012) provide a justification for focusing on a single type of worker by positing a production function that is additively separable in worker types.

inequality $f_{xy}f_{\theta\theta} - f_{y\theta}f_{x\theta} + (f_x f_{y\theta}/\theta) \leq 0$ can be rewritten as

$$\begin{aligned} F_{xy}F_{ll} - F_{yl}\left(F_{xl} - \frac{F_x}{\theta}\right) \\ &= F_{xy}F_{ll} + F_{xr}\frac{F_{yl}}{\theta} \\ &= -F_{xy}\frac{F_{lr}}{\theta} + F_{xr}\frac{F_{yl}}{\theta} \leq 0, \end{aligned}$$

which rearranges to the key finding of Eeckhout and Kircher (2012), namely, that $F_{xy}F_{lr} \geq F_{xr}F_{yl}$ delivers PAM. The left side of this inequality involves the standard cross-interactive term between the worker and firm types, scaled by the cross-interactive term of labor and resources of the firm. Meanwhile, the right side is the product of the cross-interactive terms of firm type and labor force size, and that of the worker type and firm resources. This analysis highlights the importance of the condition in Legros and Newman (2007) for analyzing sorting patterns in matching problems.

2.4.4 Further Topics

Many applications involve just one population of agents, such as collaboration between partners in law firms, team members in consulting or sports, gay marriage, etc. Although the existence of stable matchings might be problematic (Roth and Sotomayor 1990), sometimes one can divide agents into two sides and match them as if they came from a two-sided problem. Kremer and Maskin (1996) explore such a model in which identical agents might be able to perform different tasks with different productivities. If managers and workers are drawn from the same population and they are complementary but managers play a more important role, then they show that technology changes can aggravate wage inequality and the segregation

of workers by skill. Moreover, the optimal matching need not exhibit PAM.

The models analyzed above assume that output f depends only on the types of matched pairs. In some applications, the value of the match to a pair also depends on the entire matching, which gives rise to a problem with externalities. Sasaki and Toda (1996) and Pycia and Yenmez (2015) have analyzed matching with externalities, introducing notions of stability with externalities in both the perfect TU and NTU cases, and studying their implications. More recently, Chade and Eeckhout (2015) analyze the impact of externalities on the optimal and equilibrium matching patterns using a two-stage model of teams where teams are formed and later compete.

We have focused on matching by agents with scalar characteristics. The multidimensional problem—say where men and women differ in education, income, attractiveness, etc., or in a labor market where firms have many heterogeneous tasks and workers differ across several skill dimensions—is technically harder. Chiappori, McCann, and Nesheim (2010) explore existence and uniqueness of equilibrium in the multidimensional perfect TU matching model. They use tools from the optimal transport literature, linking matching models and hedonic pricing models. Lindenlaub (2014) provides a notion of sorting for multidimensional problems and studies a matching model where workers have both manual and cognitive skills and firms have jobs demanding both skills. Using US data, she analyzes technological change and its effects on the wage distribution.

Most of the matching literature assumes that agents' characteristics are primitives of the model. A small literature explores 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, thereby mitigating the hold-up

problem. Another one is to understand how imperfections at the matching stage combine with the investment problem *ex ante* and generate inefficiencies. Early work here includes Makowski and Ostroy (1995), Felli and Roberts (2016), and Cole, Mailath, and Postlewaite (2001). Noldeke and Samuelson (2015) and Bhaskar and Hopkins (2016) explore the efficiency of pre-matching investments, and Chade and Lindenlaub (2015) derive comparative statics of risk on pre-matching investments.

Search frictions or learning, analyzed later, can lead to mismatch and so rematch. In Chade and Eeckhout (2016), agents match based on observable characteristics that index the distributions of payoff-relevant attributes that are revealed after the match. For example, employers may sort workers by education, hoping this signals later productivity, and this leads to mismatch. They provide an empirical application of the model to the assignment of CEOs to firms.

3. Foundations of Search Theory

3.1 Why Search Frictions?

The matching paradigm as we have described it has some unrealistic economic properties. For one thing, it predicts no unmatched agents, except due to obvious imbalance. For another, it says nothing about mismatch among those that do match. Finally, its predictions are excessively volatile in a counterfactual way. To see this last claim, consider 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 can sometimes have dramatic effects. For instance, consider a marriage market with homogeneous men and women. With slightly more men than women, all matching rents

go to women, but with slightly more women than men, the opposite holds.²⁵ Or assume a world with heterogeneous people available for matches. There can be implausibly discontinuous matching allocations. Suppose that match payoffs are $f(x, y) = 1 + \varepsilon xy$, with $|\varepsilon| > 0$ and incredibly small. Depending on whether $\varepsilon \gtrless 0$, we either have positive or negative assortative matching, respectively. So the frictionless predictions by Becker of both the matched and unmatched agents are counterfactual in different ways. Additionally, the Walrasian auctioneer fiction is a far less accurate description of the actual matching process, since the Walrasian fiction rings less true for a market with a massive number of essentially unique items for sale. Indeed, organizing this as a market is the very challenge facing online matching services.

To fully understand how search frictions distort equilibrium market outcomes, we first need to learn some basic decision theoretic 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 nonsequential or simultaneous, where all the options are sampled at once and then the best one is chosen. In all cases, search theory explores how option value governs choices: *where* to search or *how long* or *how much* to search. Just as in finance theory, an option value is increasing in the riskiness of the choices, since extreme events yield the surplus.

3.2 Simultaneous Search

The seminal paper by Stigler (1961) started the literature on search in economics. In this model of simultaneous search, a consumer samples prices from a distribution,

²⁵For a recent reference on this issue, see Ashlagi, Kanoria, and Leshno (2017).

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 will purchase K units, the expected total outlay is

$$\begin{aligned} P(n) &= K \int_0^1 p dF_n(p) \\ &= K \int_0^1 [1 - F_n(p)] dp \\ &= K \int_0^1 [1 - F(p)]^n dp. \end{aligned}$$

Observe that $[1 - F(p)]^n$, and thus the expected cost $P(n)$, falls in n , but at a diminishing rate. Thus, the second-order condition is met, and the optimal sample size n^* obeys the discrete first-order condition: $P(n^* - 1) - P(n^*) \geq c > P(n^*) - P(n^* + 1)$. Easily, a larger planned purchase K raises the marginal benefit of sampling, and thus induces weakly more searches n^* .

Stigler's fixed sample size search is tough to motivate, as it is almost always a contrived thought experiment (as above). But there is one major 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 will 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 because colleges vary by admission chances and career value.

Recently, Chade and Smith (2006) extended the simultaneous-search paradigm to allow for ex ante heterogeneous options. The decision maker chooses not only the number of options to sample, but also the sample composition. 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 $\{1, 2, \dots, N\}$, with payoffs $v_1 > v_2 > \dots > v_N$. Since better colleges are presumably harder to secure entry to, assume inversely ranked admission chances $\alpha_1 < \alpha_2 < \dots < \alpha_N$.

Chade and Smith (2006) then deduce the optimal portfolio of any given size $n \leq N$ —and thus solve the richer problem of the optimal portfolio when all college application costs are $c > 0$. (Their analysis does not help if application costs vary.) In principle, the optimal portfolio might require searching through all possible n -subsets, or in the richer problem, all 2^N portfolios. Could college students actually be solving such a fantastically seemingly complex NP-hard problem? The authors 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, since the problem is static and not amenable to dynamic programming. One could easily imagine that a college optimal for one portfolio size might not remain so for a larger portfolio. The proof that this never happens—a joint mathematical induction on the number of options and cardinality of the portfolio set—shows why one never wishes to remove a college added at an earlier stage. The MIA is a member of a 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 $v_1 = 1$, $v_2 = 0.8$, and $v_3 = 0.6$, and admission chances $\alpha_1 = 0.5$, $\alpha_2 = 0.8$, and $\alpha_3 = 1$. The expected payoffs $z_i \equiv \alpha_i v_i$ 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 adding college 1 to a portfolio $\{2\}$ is

$$\begin{aligned} MB_{12} &= [z_1 + (1 - \alpha_1)z_2] - z_2 \\ &= z_1 - \alpha_1 z_2 \end{aligned}$$

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 adding 3 to a portfolio $\{2\}$, we note that college 2 matters whenever one is accepted there. The marginal benefit is computed therefore in a different way:

$$\begin{aligned} MB_{32} &= [z_2 + (1 - \alpha_2)z_3] - z_2 \\ &= (1 - \alpha_2)z_3. \end{aligned}$$

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. They should not blindly apply to their best expected options. For instance, the lower-ranked colleges 2 and 3 have the two highest expected payoffs.

This is best seen as a justification for why students pursue “stretch schools.” For assume a world with just college i and many identical lower-ranked colleges j . Assume that even though any such college j has a lower payoff $v_j < v_i$, it has a higher expected value $\alpha_j v_j > \alpha_i v_i$. As a result, 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 more college j copies vanishes geometrically in their number, and so eventually falls below $\alpha_i(v_i - v_j) - c > 0$, for any application cost $c > 0$. By contrast, the marginal benefit of adding college i to a portfolio of n colleges j is

$$\begin{aligned} \alpha_i v_i - \alpha_i v_j (1 - (1 - \alpha_j)^n) - c \\ > \alpha_i (v_i - v_j) - c. \end{aligned}$$

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 acceptances 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 pursue a safety-school strategy. This remains a challenging but important 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 (2001) and Albrecht, Gautier, and Vroman (2006); as well as search problems with multiple applications and heterogeneous options such as Chade, Lewis, and Smith (2014) for college admissions, and Kircher (2009) and Galenianos and Kircher (2009) for labor markets.

3.3 Sequential Search

Amusingly, while Stigler's paper introduced price search—deploring 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. His classic model has become a fundamental building block for macroeconomic models of the labor market (see the discussion in section 3 in Rogerson, Shimer, and Wright 2005), and we will also use it extensively in section 4. The worker's optimal strategy is fully summarized by a reservation wage \bar{w} above which the worker stops searching, and below which he continues.²⁶ Since he is now indifferent when faced with his

reservation wage, and will in the future be, his new optimality condition requires:

$$(15) \quad \bar{w} = \int_0^\infty \max(\bar{w}, w) dF(w) - c \\ \Rightarrow c = \int_{\bar{w}}^\infty [1 - F(w)] dw.$$

Since the problem is stationary, a wage once rejected is forever rejected. As a result, an option to return to a previously declined option is worthless. This expression admits some immediate predictions. For instance, the hazard rate of finding jobs is constant through time. Also, 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, and hence it encourages risk-taking behavior—for instance, acquiring information about F .

The impact on search duration is a priori ambiguous—for the searcher is more ambitious with a mean-preserving spread, but there is also more probability weight in the upper tail. Choi and Smith (2016) resolve this ambiguity, showing that if every pair of percentiles of the wage distribution shift apart—namely, there is a right shift in the *dispersion order*—then the hazard rate of stopping $1 - F(\bar{w})$ falls, and consequently the search duration rises.

We pursue a richer model than McCall (1970) that allows for ex ante heterogeneous options: the Pandora's box of Weitzman (1979). Assume 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. Only one prize may ultimately be accepted. Payoffs are independent, and the decision maker must sample them sequentially. At any point in time, she can

²⁶Morgan and Manning (1985) endogenized the sample size at each stage of the search process, thereby blending sequential and simultaneous search.

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 infinite horizon multi-armed bandit problem in Gittins and Jones (1974) and later Gittins (1979)—the so-called *Gittins index*.²⁷ When arm payoffs are independent, the index for each arm *solely* reflects the uncertainty of that arm. Capturing the contingent decision making, it is the fixed prize \bar{w}_k that leaves the decision maker indifferent about choosing a prize, and paying to open box k , knowing that the prize awaits him if he wants it. Namely:

$$(16) \bar{w}_k = \delta_i \int_0^\infty \max(\bar{w}_k, w) dF_k(w) - c_i.$$

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 samples the option with the largest index and stops when the reward observed exceeds the reservation values of all the remaining options.²⁸ Notice that the reservation wage equation (15) emerges from (16) with homogeneous options and costs and no discounting—namely, $F_i = F$, $c_i = c$, and $\delta_i = 1$. McCall (1965) intuitively feels like a special case of Weitzman (1979) when there is a vast number of identical options. This logic shows that the reservation wage coincides with Weitzman's index with only finitely many options when there is recall of

past options. Any such recall freedom has no value in the stationary job search setting.

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 are statically dominated, with a lower mean and a higher cost. They are valuable provided they have a high enough variance.

Let's revisit the college application problem of section 3.2. Assuming that one could apply in sequence to colleges, one would optimally employ Weitzman's rule. This is true even with college-specific application costs. With our binary payoff distribution, the index equation (16) 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 individuals act more aggressively with sequential decision making than simultaneous choices. One should pursue less likely options first. For instance, in the example of section 3.2, college $i = 2$ is the first applied to; however, the indexes are

$$\bar{w}_1 = u_1 - c/\alpha_1 = 1 - 0.15/0.5$$

$$= 0.7 > 0.6125 = 0.8 - 0.15/0.8$$

$$= u_2 - c/\alpha_2 = \bar{w}_2.$$

A general lesson is that sequential decision making pushes toward more risk-taking behavior. For example, early admission might be unwise for the most elite schools because it encourages more

²⁷A multi-armed bandit is a finite action, infinite horizon Bayesian experimentation problem. When the payoff of each "arm" is independent of all others, the optimal strategy is given by Gittins indices.

²⁸Ólszewski and Weber (2015) explore the limits of this class of index rules for different payoff functions.

aggressive “stretch” applications by weaker students who otherwise would not apply.

These optimal stopping problems have long been explored in operations research, where it is assumed that a decision maker chooses a sequentially optimal time 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 (2016); 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., Rogerson, Shimer, 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 section 4).

One might view search theory as optimal stopping when one knows the payoff distribution, but not the realizations. In a key extension of the sequential-search problem, Rothschild (1974) explored the implications of learning the distribution while searching. This has been revisited in many guises, recently by Adam (2001), and Gershkov and Moldovanu (2012).

Finally, in equilibrium applications that we will study in section 4, 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 (16), and think of $\delta_i = 1 - r dt$, and c_i as a flow search cost. Then (16) becomes

$$(17) \quad r \bar{w}_k \\ = \rho \int_0^\infty \max(w - \bar{w}_k, 0) dF_k(w) - c_i.$$

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 Sequential Search with Hidden and Known Components

Choi and Smith (2016) explore another specialization of Weitzman (1979) in which they assume that the prize distributions F_k reflect the sum of hidden and known components, each with a common distribution. For simplicity, we touch on their application to web search, since it is a major new way that sorting and matching are proceeding now.²⁹

In this case, assume that all payoffs W hail from a Gaussian distribution. Since the Gaussian distribution is stable, it can be parsed as:

$$(18) \quad W = \alpha X + \sqrt{1 - \alpha^2} Z,$$

in which X and Z are each standard normal random variables. For instance, after entering the keyword, the search engine ranks the search outcomes by the realized *known components* X . To learn the *idiosyncratic component* Z , the user must click on the website and read it.

The search engine accuracy α represents how effective the search engine is in reducing the idiosyncratic noise, rendering more predictable web searches. When $\alpha = 0$, the problem reduces to 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

²⁹Varian (1999) mused on practical advice for offering books to a rushed consumer in an airport bookstore. He identified Weitzman (1979) as a parable for web search engines.

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 nonstationary 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: $x_1 > x_2 > \dots$. In this model, a user clicks on the k th web site if and only if 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. Given (18), the threshold $\zeta(\alpha)$ obeys a reservation equation akin to (15), 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 (2016) 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 if and only if $u < \zeta(\alpha)$. In other words, there may be a conflict of interest between online shopping sites (for, say, Amazon) and consumers. In particular, when the outside options are high or the price is low, a shopping website secures higher sales with a noisier search engine.

Choi and Smith (2016) derive an array of results for this nonstationary search

environment. For instance, search intensifies, with recall rates and quitting rates rising over time.

3.5 Sequential 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 former (AAV), since it shows how search costs skew the partner search process. Consider a department seeking to hire job candidates that arrive sequentially, one per period. The search cost is impatience: the payoff is discounted by $0 < \delta < 1$. Each member i of hiring committee of N members observes a random private value $W \in [0, 1]$, independently drawn from the common cdf F . After seeing 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 flavor of the theory, consider first a committee of one. It votes to hire a candidate when $W \geq \bar{w}$, given the reservation value $\bar{w} = \delta V(\bar{w})$, where V is a fixed point $V = TV$ of the Bellman operator $TV(w) = (1 - F(w))E[W|W \geq w] + \delta F(w)V(w)$. Next, assume a size $N = 2$ committee. In the symmetric equilibrium, each player still employs a reservation value w , and secures Bellman value $V_M(w)$. For $M = 1$, the value solves the Bellman recursion $T_1 V_1 = V_1$, where

$$\begin{aligned} T_1 V(w) &= (1 - F(w))E[W|W \geq w] \\ &+ F(w)(1 - F(w))E[W|W \leq w] \\ &+ F(w)^2 \delta V(w). \end{aligned}$$

For any player j 's payoff exceeds the threshold w with chance $1 - F(w)$, whereupon search ends with that payoff $E[W|W \geq w]$. Next, j 's payoff is below w with chance $F(w)$. In this case, if the other player k 's payoff exceeds w (chance $1 - F(w)$), then j earns his low payoff $E[W|W \leq w]$. Otherwise, j earns the discounted continuation value.

Since the operators are ordered $T_1 V < TV$ for all v , their fixed points are ranked $V_1 < V$, and their reservation values likewise so: $\bar{w}_1 < \bar{w}$. In other words, committee members are individually less picky: the single agent i rejects any candidate that committee member i rejects. The value reflects a *stopping externality*, 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 then a fixed point of:

$$T_2 V(w) \equiv (1 - F(w))^2 E[W|W \geq w] \\ + F(w)(2 - F(w)) \delta V(w)$$

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 trade-off—a candidate must independently pass several lower thresholds. AAV show that when $M < N$, the committee concludes search faster than 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. For a mean-preserving spread can increase either (a) the stopping externality, by making low draws more costly, or (b) the continuation externality, by increasing the chance that another member of the committee blocks you, or both.

This literature adding strategic elements to the search problem is an inviting future direction, in light of the important of collective decision making in resolving search frictions.

4. Search and Matching

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. This twist is important, since it eliminates discontinuous matching sets and wage profiles (in section 3.1) predicted by the frictionless model, and explains important phenomena like equilibrium unemployment and imperfect sorting (mismatch), and rationalizes price and wage dispersion.

We distinguish between time intensive random search, and directed or 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,³⁰ and this remains a major direction of future research.

But the quest to enrich Becker's framework and account for search frictions has received an enormous amount of attention in recent years. Among the earliest such models of heterogeneous agent search were Bergstrom and Bagnoli (1993), which we explore in section 5.2, and Smith (1992). These papers assumed NTU, and were followed up by Burdett and Coles (1997) and Smith (2006). But the proper extension of Becker's model required perfect 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 perfect TU or NTU. First, it is standard to assume continuous time with an exponential arrivals of matching opportunities. Next, for models of partner search, it is common to model search cost as impatience, rather than an explicit search cost. Also, with a few salient exceptions, the unisex model is assumed for simplicity, but a similar analysis can be done with two distinct sides, as in the frictionless case.

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 proportionality constant? In a *linear search technology*, potential partners arrive with constant rate $\rho > 0$. To wit, the density mass of new matches is linearly proportional to the mass

of unmatched agents. On the other hand, in the *quadratic-search technology*, unmatched individuals face a constant arrival rate of potential partners ρ times the mass of unmatched agents; here, the mass of new matches is proportional to the squared mass of unmatched agents. We will refer to ρ as a *meeting rate*, or possibly, the *rendezvous rate*. For intuition, this arises if invitations to meetings arrive at fixed rate ρ to everyone, but when either party is already matched, he misses the meeting.

The quadratic-search technology embeds a crucial analytic advantage: players are unaffected by the matching decisions of those unwilling to match with them. This *strategic independence* greatly simplifies equilibrium analysis. By contrast, in order to hold the matching rate constant with a linear-search technology, new individuals that enter one's matching set crowd out previous individuals. This complicates and sometimes renders impossible equilibrium analysis, especially in a nonstationary environment. We now illustrate this with Smith (1992), the first heterogeneous agent search model that properly tracked the demographics. This leads us into the sorting results.

4.2 *Sorting with Random Search and Nontransferable Utility*

4.2.1 *Block Assortative Matching*

When individuals sort into matches by anonymous random search, an intriguing equilibrium matching pattern emerges. To see this, assume that everyone is summarized by a scalar type in $[0, 1]$, and posit a uniform density on types. Assume no match complementarities,³¹ and posit that anyone matching with type x earns payoff x . A pairwise matching market is newly opening, with everyone initially unmatched, and meeting potential partners according to a quadratic-search technology

³⁰This assumption has already been explored in a search and trading model in Smith (1995).

³¹Some introspection reveals that this analysis also works when the payoff to the match of types x and y is xy .

with meeting rate ρ . We assume that matching is irreversible, but this will be optimal ex post anyhow. Everyone wishes to enter a permanent match with the highest discounted expected present value. In the equilibrium that transpires, intuitively everyone will start off wanting to match with the highest types, but then these types gradually vanish, and the cutoff monotonically falls. By the logic of section 3.3, the marginal type that leaves one indifferent between matching and continuing is the unmatched value \bar{v} . If this marginal type monotonically vanishes, then the equilibrium assumes the form: any types in $[\bar{v}, 1]$ agree to a match when they meet, and everyone declines to match with anyone in $[0, \bar{v}]$.

To verify this equilibrium, we adapt the reservation wage equation (17) for a nonstationary world, in which there is a further capital loss reflecting the falling value $\bar{v}' < 0$ (suppressing the time subscripts). For any interest rate $r \geq 0$, the return is the sum of the “capital gains” and “dividends,” namely:

$$(19) \quad r\bar{v} = \bar{v}' + \rho \int_{\bar{v}}^1 (x - \bar{v})u(x) dx,$$

where $u(x)$ is the unemployment mass density of type $x \in [0, 1]$. Tracking the unmatched mass density $u(x)$ is essential in heterogeneous agent search models. But we can greatly simplify the problem, and capture the evolution of this threshold simply using two state variables—the total unmatched measure mass $\bar{u} = \int_0^1 u(x) dx$, and the first moment $\chi = \int_0^1 xu(x) dx$ of unmatched agents. So the average type of an unmatched agent is χ/\bar{u} . Rewrite the law of motion (19) for the unmatched value as:³²

$$(20) \quad \bar{v}' = r\bar{v} + \rho(\bar{v}\bar{u} - \chi + \bar{v}^2/2).$$

³²Write (19) as $\bar{v}' = r\bar{v} - \rho \int_{\bar{v}}^1 (x - \bar{v}) dx$. Since types below \bar{v} have not matched, they still have a uniform density, so that $u(x) = 1$ for all $x \leq \bar{v}$. Hence, $\int_{\bar{v}}^1 (x - \bar{v})u(x) dx = \int_0^1 xu(x) dx - \int_0^{\bar{v}} x dx - \bar{v}(\bar{u} - \bar{v}) = \chi - \bar{v}\bar{u} + \bar{v}^2/2$.

Since types in $[\bar{v}, 1]$ match just among themselves and they have mass $\bar{u} - \bar{v}$, the quadratic-search technology implies that their unmatched density $u(x)$ falls at rate $\rho(\bar{u} - \bar{v})$. Because the mass and first moment of agents below \bar{v} is respectively \bar{v} and $\bar{v}^2/2$, the laws of motion for \bar{u} and χ are thus:

$$\bar{u}' = -\rho(\bar{u} - \bar{v})^2$$

and

$$\chi' = -\rho(\bar{u} - \bar{v})(\chi - \bar{v}^2/2).$$

All told, this nonstationary “rush” equilibrium is captured by this three-dimensional state (\bar{v}, \bar{u}, χ) . Naturally, the threshold \bar{v} vanishes; Smith (1992) argues that it does so at rate $O(1/t)$.³³

Assume now a flow entry with uniform density $e > 0$ on $[0, 1]$, the threshold \bar{v} no longer vanishes. For a constant inflow of high types, there is now a strictly positive lower bound on the option value of waiting. This equilibrium offers a unique approach to thinking about steady-state analysis. In the long-run limit equilibrium, we should approach steady state, in which the threshold \bar{v}_1 is intuitively constant and strictly positive. For the matching threshold need not vanish in order to satisfy dynamic optimality. The logic of steady state requires that entry of new unmatched agents balance the flow of agents from the unmatched pool into matches. This yields $e = \rho u_1(1 - \bar{v}_1)$. Meanwhile, the steady-state condition $\bar{v}'_1 = 0$ in (20) intuitively yields an optimality condition:

$$\begin{aligned} r\bar{v}_1 &= \rho u_1 \int_{\bar{v}_1}^1 (x - \bar{v}_1) dx \\ &= \rho u_1(1 - \bar{v}_1)^2/2. \end{aligned}$$

³³In pursuant work, Damiano, Li, and Suen (2005) develop a general theory of how matching unravels.

Jointly, we can solve for the pair (u_1, \bar{v}_1) . But then to understand how types $x < \bar{v}_1$ match, the logic starts anew. Recursively, there is a sequence of thresholds $1 > \bar{v}_1 > \bar{v}_2 > \dots$ and associated unmatched rates u_2, u_3, \dots computed inductively.³⁴ In other words, whenever two types in the same interval $[\bar{v}_k, \bar{v}_{k-1})$ meet, for some $k = 1, 2, \dots$, they agree to match; otherwise, the higher type declines the match. Smith (2006) suggestively called this equilibrium *block segregation*. This *balanced-flow* approach was formally developed for any matching sets in (the 1997 working paper version of) Smith (2006), with a general existence proof.

The earliest heterogeneous agent sorting paper, McNamara and Collins (1990) assumed a stationary equilibrium, without clarification. Later, Bloch and Ryder (2000) made an endogenous “cloning assumption,” positing that agents who leave the matching market are somehow magically replaced by clones. This simply requires solving a set of difference equations, as does the later approach in Burdett and Coles (1997), who assumed a fixed flow entry of all types and no match dissolution. Finally, if match dissolution is indeed economically central, then one can assume a constant stock of agents, and just ask that severances balance new matches for each type; Shimer and Smith (2000) take this tack, as their primary application is labor.

We see how the first taste of sorting models with search frictions entailed jumps: Individuals match in ranked-equivalence classes. Smith (2006) highlights the counterfactual discontinuous aspects of this block segregation—extremely close individual

types do not typically have entirely disjoint sets of match partners. Burdett and Coles (1997) instead took it as a parable of “marriage and class” in Britain. It is a very stark form of PAM.³⁵ The question then arose as to when matching sets were continuous in types. Relatedly, 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 NAM that simultaneously applies to singleton matching sets 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.

By drawing a suitably small rectangle at the edge of the matching set, this definition implies that each type x matches with types in an interval $[a(x), b(x)]$, with both $a(x)$ and $b(x)$ weakly increasing in x . 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

³⁴Further, by strategic independence of the quadratic-search technology (see section 4.1), these can be computed in isolation: for we have $r\bar{v}_{k+1} = \rho u_{k+1} \int_{\bar{v}_{k+1}}^{\bar{v}_k} (x - \bar{v}_{k+1}) dx = \rho u_{k+1} [\bar{v}_k - \bar{v}_{k+1}]^2/2$, as $k = 1, 2, \dots$

³⁵This early literature on sorting with search frictions has many other participants, some with overlapping results, with varying insights into search costs, intermediation, and the block segregation logic. See McNamara and Collins (1990), Morgan (1996), Bloch and Ryder (2000), Chade (2001), and Eeckhout (1999).

NAM. This is a testable implication for the data.

4.2.2 Strict Assortative Matching

Under 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 this case with search frictions.

In Smith (2006), time is continuous on $[0, \infty)$ and search is a time cost: unmatched agents discount the future at rate $r > 0$. There is a continuum of types $x \in [0, 1]$ with cdf G and a positive density g . Unmatched agents earn zero flow payoffs, while a match with a type x agent yields flow payoff $f(x, y) > 0$. Here, f is strictly increasing in partner's type, so that $f_y > 0$ 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 non-negative 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 $A(x) \cap \Omega(x)$. Observe how NTU captures the classic double coincidence of wants that money solves. 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 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:

$$\begin{aligned} (21) \quad rv(x) &= \rho \int_{\Omega(x)} \max(v(x|y) - v(x), 0) u(y) dy \\ &= \rho \int_{A(x) \cap \Omega(x)} (v(x|y) - v(x)) u(y) dy. \end{aligned}$$

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 $A(x) = [a(x), 1]$. 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 $A(x) = [a(x), 1]$ into (21) yields the explicit recursion equation:

$$\begin{aligned} (22) \quad rv(x) &= \frac{\rho}{r + \kappa} \int_{A(x) \cap \Omega(x)} [f(x, y) - rv(x)] u(y) dy. \end{aligned}$$

This expression reveals how the return on the unmatched value reflects how matches dissolve at rate κ . Finally, an equilibrium is a triple (v, a, u) , such that $v(x)$ obeys (22) for the acceptance set $A(x) = [a(x), 1]$, $a(x)$ solves $f(x, a(x)) = rv(x)$ given $u(\cdot)$, and $u(x)$

obeys the balanced flow condition (23) at every type x .³⁶

$$(23) \quad \kappa(g(x) - u(x)) \\ = \rho u(x) \int_{A(x) \cap \Omega(x)} u(y) dy.$$

Smith (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. If $a(x)$ is weakly increasing in x , then so too is its inverse $b(x)$.³⁷ Graphically, matching then engulfs all types between two increasing bands $[a(x), b(x)]$, whereupon PAM obtains.

Assume for now that $\Omega(x) = [0, b(x)]$ for all x , and with $b(x)$ weakly increasing. For the highest types, this is true since $b(x) = 1$. Since $a(x)$ is an optimal lower threshold, (22) becomes

$$(24) \quad rv(x) = \frac{\int_{a(x)}^{b(x)} f(x, y) u(y) dy}{\psi + \int_{a(x)}^{b(x)} u(y) dy} \\ = \max_a \frac{\int_a^{b(x)} f(x, y) u(y) dy}{\psi + \int_a^{b(x)} u(y) dy},$$

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

$$(25) \quad \psi = \int_a^{b(x)} \left(\frac{f(x, y)}{f(x, a)} - 1 \right) u(y) dy.$$

³⁶Smith (1997, 2006) analyzes this equation. Smith (2011) shows that one essentially applies the implicit function theorem to deduce a continuous map from $a(\cdot)$ to $u(\cdot)$.

³⁷The logic for PAM holds with two-sided matching, with two populations, like men and women. In that case, the upper bound function $b(x)$ is derived from the acceptance threshold of the other population.

Since $y > a$, the integrand on the right side increases in x if $f(x, y)/f(x, a)$ is strictly increasing in x —namely, $f(x, y)$ is strictly log supermodular (that is, $\log f$ is 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 . To follow the key observation that led to this general log-supermodularity condition for PAM, note that block segregation arises in Smith (1992) for any multiplicative payoffs $f(x, y) = f_1(x)f_2(y)$, where each f_i is positive and increasing. Next observe that such multiplicative functions are obviously log modular. Not surprisingly, $a(x)$ is constant if $f(x, y)/f(x, a)$ is always constant. This also offers another way of understanding why block segregation emerges for the assumed payoffs when one cares about one's partner's type, namely, $f(x, y) = y$.

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

$$(26) \quad c = \int_a^{b(x)} (f(x, y) - f(x, a)) u(y) dy.$$

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. But in this fixed-search case, block-segregation PAM arises with a modular production function $f(x, y) = f_1(x) + f_2(y)$ (Chade 2001). Morgan (1996) also studies NTU matching with fixed-search costs.

All told, the productive conditions for PAM are harder to satisfy in the presence of time cost search frictions. In the frictionless setting, supermodularity gives higher types greater gains from matching up. But since the search costs rise in proportion to the value, a stronger assumption is required—log supermodularity rather than just supermodularity. This ensures that higher types have

proportionately greater gains from matching up. While log-supermodular payoffs are *sufficient* for PAM, it is also *necessary* to ensure PAM for all unmatched distributions—otherwise, one could put a large mass on the failure type set, and violate PAM.

The outlined argument yields a unique equilibrium, *given* the unmatched density function $u(x)$. But an equilibrium is really a triple (v, a, g) . 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 *nonselective* 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.

4.3 Sorting with Random Search and Perfectly 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 nontransferable 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 nonmonetary (such as division of child care or household chores).

The present value to any two matched types $x, y \in [0, 1]$ is no longer exogenously fixed, since the surplus split is endogenous.

Easily, *match surplus* equals match output less the sum of the returns on the unmatched values: $s(x, y) \equiv f(x, y) - rv(x) - rv(y)$. In a unisex model, it is natural to split surplus equally.³⁸ Since match surplus is nonnegative, we need no longer keep track of an acceptance set and opportunity set, but a single *matching set* $M(x) = \{y \mid s(x, y) \geq 0\}$. Analogizing (22):

$$(27) \quad rv(x) = \frac{1}{2} \frac{\rho}{r + \kappa} \times \int_{M(x)} [f(x, y) - rv(x) - rv(y)] u(y) dy.$$

The sufficient condition for PAM or NAM is much less obvious than it was with NTU, for the integral includes endogenous value functions. Shimer and Smith (2000) simplified matters and restricted focus to increasing payoffs $f_x, f_y > 0$. As in Becker (1973), they first argued that the locus of zero-surplus matches is increasing in one's type provided $f_{xy} > 0$. PAM then obtains if matching sets are convex and $a(0) = 0$, as $a(x)$ and $b(x)$ are then weakly increasing.

Now, if we assume $f_x(x, 0) \equiv 0$, then $a(0) = 0$. Next, to deduce convex matching sets, differentiate the Bellman equation (27). Since match surplus $s(x, y)$ vanishes along the edge of the matching set $M(x)$, an intuitive application of the fundamental theorem of calculus yields:

$$(28) \quad rv'(x) = \frac{1}{2} \frac{\rho}{r + \kappa} \times \int_{M(x)} [f_x(x, y) - rv'(x)] u(y) dy = \frac{\int_{M(x)} f_x(x, y) u(y) dy}{2\psi + \int_{M(x)} u(y) dy}.$$

³⁸Called the Nash bargaining solution, this uses none of that concept's richness, as the surplus frontier is linear.

This expression obviously parallels (22) and (24), and so the proof takes inspiration from Smith (1997, 2006). Now, a simple sufficient condition for the matching set of type x to be convex is that his surplus $s(x, y)$ be quasi-concave in his partner y . To argue this, Shimer and Smith (2000) use $f_{xy} > 0$ and the right $v'(x)$ formula in (28) to argue that $s_x(x, y) > 0$ for all large enough $y < 1$. In this case, $s(x, y)$ is quasi concave in y . Next, using the same $v'(x)$ expression, and f_x log supermodular, Shimer and Smith (2000) deduce that $rv'(x)/f_x(x, y)$ is increasing for all small $y > 0$. So in this case, $s_x(x, y) \equiv f_x(x, y) - rv'(x)$ is reverse single-crossing in y , i.e., positive and then negative as y rises. This implies that $s(x, y)$ is quasi concave in y . Finally, a third single crossing property argues that y is large enough for the first case, or small enough for the second, provided f_{xy} is log supermodular. Smith (2011) carefully distills this argument.

Besides providing sufficient conditions for PAM, Shimer and Smith (2000) also prove existence of equilibrium. Their proof assumes the quadratic-meeting technology; later, Noldeke and Troger (2009) showed existence under the technically harder linear-meeting technology. Manea (2017) recently dispensed with the quadratic assumption, deducing existence in a larger class of stationary search and matching models when there is a finite number of types.

Instead of time discounting, one could assume a fixed search cost. 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 now many papers bringing 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 is an important agenda.

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 line of research beginning with 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 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.³⁹

It has long been recognized that labor markets can generate multiplicity, and therefore cyclical outcomes. Diamond (1982) shows that multiple steady states can arise in a simple exchange economy, but his logic exploits the increasing return-to-scale property of the quadratic-search technology. Pissarides and Petrongolo (2001) find evidence of constant returns in the observed matching technology in the labor market. Burdett and Coles (1997) deduce that heterogeneous types and an endogenous searching pool creates local increasing returns for subsets of types. In a related two-type model with transferable utility, Shimer and Smith (2001) analyze optimal policy and characterize the planner's assignment constrained by the search technology. They find that even absent any intrinsic uncertainty, optimal allocations may be nonstationary. This draws into question 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 explored models with nonvanishing 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 limits of equilibria of the decentralized market with search frictions. Without cloning, Lauermaun and Noldeke (2014) show that 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.

4.4 Directed Search and Sorting

4.4.1 Background

Anonymous random search takes the dynamic process of partner quest in labor and marriage markets seriously. Yet, when agents are heterogeneous, the purely random meetings process performs as if it assumes that high types meet low types, even though they know they will never form a match. This is particularly costly and, not surprisingly, quasi-market structures or coordination games have emerged that seek to mitigate these losses.

In a competitive economy, prices play an informative role, since they signal willingness to buy and sell. With anonymous random search, prices simply determine the output split between buyers and sellers, but in no way influence the actual meeting process. Directed search instead assumes that prices influence the meeting process rather than solely the surplus split. Whereas trading partners meet and then determine the price through bargaining with random search, the order is reversed under directed search. There, sellers commit to a price and post it, and after observing the price, buyers choose with whom to trade. This allows buyers to direct their search towards sellers that offer better prices.⁴⁰ But it also allows coordination

³⁹See also the related work by Teulings and Gautier (2004), De Melo (2009), Bagger and Lentz (2014), Lamadon et al. (2013), and Bartolucci and Devicienti (2013).

⁴⁰The 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

failures, in which either multiple workers turn up for one job and some workers remain unemployed in equilibrium, or in which no workers show up, and some vacancies are thus left unfilled. This yields wholly different predictions than the Walrasian one.

There are two formulations of directed search in the literature. In one, the frictions are captured by *queues*, which determine a trading probability. This is often referred to as competitive search, following Moen (1997), who analyzes stationary equilibria in a continuous-time setting. In another formulation, the friction is embodied by the chance of a *stockout*. Here, if several workers turn up for a job, it is probabilistically rationed. Both approaches generate a trading probability as a function of the ratio of applicants to jobs, and a wage (transfer).⁴¹

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

4.4.2 Sorting and Directed Search

There is a continuum mass of buyers each with a characteristic $x \in [0, 1]$ and unit demands, and likewise of sellers with characteristic $y \in [0, 1]$, each holding a unit. For instance, this might be the housing-exchange model of Shapley and Shubik (1972). If the

buyer x pays price p for a good bought from seller y , then her payoff is $f(x, y) - p$ and the seller's is p . The seller's characteristics are observable. The characteristic distributions $G(x)$ and $H(y)$ are continuous, with positive densities $g(x)$ and $h(y)$.

With anonymous random matching, the expected number of newly forming matches depends on the mass of buyers and sellers in the market. Under a standard assumption that the "matching function" exhibits constant returns to scale—e.g., twice as many buyers and sellers leads to twice as many matches—the matching function is linear in the ratio of buyers to sellers θ . This same matching-function logic applies to the directed-search environment, where θ now coincides with the *expected queue length* in each submarket at each seller. The interaction here takes two stages. First, each seller y posts a price p at which she is willing to sell the good. Second, buyers choose which seller (y, p) to visit. We illustrate this idea in the model of Eeckhout and Kircher (2010a).⁴²

With queue length θ , buyers meet sellers with chance $q(\theta)$ and sellers meet buyers with chance $m(\theta) = \theta q(\theta)$. Buyer x thus has expected payoff $q(\theta)[f(x, y) - p]$ in pursuing seller y with price p and queue θ , and that seller has expected payoff $m(\theta)p$. Here, m is assumed twice continuously differentiable, strictly increasing, and strictly concave. Seller y could choose different pairs (p, θ) , but buyer x would have to enjoy the same reservation utility, say $v(x)$. All told, seller y solves

$$\max_{x, \theta, p} m(\theta)p$$

subject to

$$q(\theta)(f(x, y) - p) = v(x)$$

(1997), Acemoglu and Shimer (1999), and Burdett, Shi, and Wright (2001).

⁴¹The difference between the two interpretations becomes real for more general mechanisms than mere price posting (Eeckhout and Kircher 2010b).

⁴²Versions of this model have been analyzed by Shi, (2001), Mortensen and Wright (2002), Eeckhout and Kircher (2010a), and Jerez (2012).

where $v(x)$ is the reservation utility of x . We will show that this optimization is of the imperfect TU form in section 2.3 and, as a result, we derive the sorting conditions by applying the differential inequality (9).

Substituting for p from the constraint yields the optimization $\max_{x,\theta} m(\theta)f(x,y) - \theta v(x)$. Equivalently, each type y must maximize $\psi(y,x,v(x)) \equiv m(\theta(x,y,v(x)))f(x,y) - \theta(x,y,v(x))v(x)$ in x , where the optimal queue length $\theta(x,y,v(x))$ solves the FOC $m'(\theta)f(x,y) = v(x)$. Despite having linear preferences over money, the objective function is nonlinear in v , reflecting the imperfect TU structure of matching. PAM or NAM arises in equilibrium depending on whether $\psi(y,x,v) = m(\theta(x,y,v))f(x,y) - \theta(x,y,v)v$ satisfies $\psi_{xy} - (\psi_x/\psi_v)\psi_{vy} \leq 0$. This allows us a quick derivation of the sorting condition (14) of Eeckhout and Kircher (2010a). For one can verify that $\theta_y = -m'f_y/m''f$, as well as $\psi_x = mf_x$ and $\psi_v = -\theta$ by the envelope theorem. Also, differentiating $\psi_x = mf_x$ and $\psi_v = -\theta$ respectively yield $\psi_{xy} = mf_{xy} + m'\theta_y f_x$ and $\psi_{vy} = -\theta_y$. Hence:

$$\begin{aligned} (29) \quad \psi_{xy} - \frac{\psi_x}{\psi_v} \psi_{vy} &= [mf_{xy} + m'\theta_y f_x] - [mf_x] \theta_y / \theta \\ &= mf_x f_y \left[\frac{f_{xy} f}{f_x f_y} - \xi(\theta) \right], \end{aligned}$$

where $\xi(\theta) = m'(\theta m' - m)/(m''m\theta)$. Then PAM or NAM arises as $f_{xy} f / (f_x f_y) \geq \xi(\theta)$ for all θ .

To gain some intuition about this condition, assume first no match complementarities $f_{xy} = 0$, as in the labor market model of Mortensen and Wright (2002). If the elasticity $\xi(\theta)$ is always positive, then NAM arises, by this criterion. Since output is purely additive, high-type firms have the most to lose with a

vacancy; therefore, they prefer to trade with higher probability, and so opt for a shorter queue θ . As a result, many high-type workers match with fewer low-type firms, and vice versa. PAM only emerges if there are sufficiently strong match complementarities.

If the elasticity $\xi(\theta)$ is identically zero, then supermodular production f suffices for PAM, by the criterion. For in this case, $m(\theta) = \theta$, and so $q(\theta) = 1$, which is to say that the arrival rate of sellers for each buyer is constant. At the opposite extreme, if the elasticity $\xi(\theta)$ is identically one, then the criterion asserts that log-supermodular production f gives PAM. So unless production f is “sufficiently supermodular,” the search frictions lead to NAM.

The required condition on production is stronger than supermodularity and weaker than log supermodularity for intermediate elasticities $0 < \xi(\theta) < 1$. For instance, if $0 < \xi(\theta) < 1 - \Upsilon < 1$, then PAM obtains if

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

This suggests a simple sufficient condition for PAM is supermodularity of the power function f^Υ , and in the limit as $\Upsilon \rightarrow 0$, we recover log supermodularity (to see this, take the limit of $(f^\Upsilon - 1)/\Upsilon$). Altogether, Becker’s supermodularity assumption on production no longer suffices for sorting in the frictional world. With anonymous random search and ex post equal surplus splitting, log supermodular production is required. When search is directed, the reduced frictions entail generally an intermediate level of supermodularity.

4.4.3 Market Segmentation

An alternative to directed search assumes random matching, but allows heterogeneous agents to set up separate trading posts. Jacquet and Tan (2007) find that even with

random matching, there are gains for individuals to set up segmented markets. For example, starting from the block assortative matching equilibrium in section 4.2, all agents in the upper class prefer to meet only among themselves in a segregated market.

Since the distribution of singles is truncated, the acceptance threshold of each agent rises. This gives them incentives to continue segregating, further refining their type partition. One might expect that this leads to perfect segregation, as in the frictionless model, with each agent type matching with a unique type. Jacquet and Tan (2007) refute this intuition. For no agent can commit to rejecting a partner slightly below his ideal. Because of search frictions, it is costly to wait for the ideal partner, and so if the current candidate is marginally lower, the agent will accept. So equilibrium segmentation consists of a sequence of nondegenerate intervals of types.

4.4.4 *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. Assume that buyers have heterogeneous valuations for the good sellers offer, and sellers have heterogeneous costs for the good they sell. Sellers simultaneously choose and commit to a mechanism that maximizes their profits within a broad class, including auctions, price posting, and bargaining. Observing the announced mechanism, buyers then visit a seller, equally randomizing over all sellers who announce the same mechanism.

McAfee (1993) finds an equilibrium in which firms choose a second-price auction with a reserve price equal to the firm's cost for the good. Buyers visit each seller with a probability falling in the reserve price. The reserve price strictly exceeds the cost in the

optimal second-price auction by a monopolist, but here it coincides with the cost. This reflects the competition among sellers for buyers, as with Bertrand price competition. Since the number of buyers each seller attracts is inversely related to the seller's reserve price, this is a force toward lower reserve prices.

In an endorsement of second-price auctions, McAfee (1993) also shows that this mechanism is always a best response to any arbitrary set of mechanisms by the other sellers. Peters and Severinov (1997) and Peters (1997) later justified the large market assumption in McAfee (1993) by considering the limit of equilibria with finite markets as the number of agents grows large. McAfee's results suggest that auctions are superior to posted prices when there is competition between auctioneers. But Eeckhout and Kircher (2010b) argue that this result depends on the ability of auctioneers to extract rents *ex post* from buyers by having them simultaneously participate in the auction and then screening them. To see this, notice that in the queueing interpretation of directed search (Moen 1997), a posted price is optimal because a firm only faces one bidder at a time. Thus, the use of posted prices or auctions reflects the nature of the search frictions. If the auctioneer cannot round up sufficient applicants to bid in the mechanism, she is better off posting a price.⁴³ Auctions are optimal when meeting probabilities are unaffected by the decision of another buyer to visit a submarket.⁴⁴ With posted prices,

⁴³Pinheiro (2012) provides a microfoundation in a model where the firm's mechanism design problem takes place in continuous time with random arrival of buyers. Now firms face a trade-off between trading fast and having few bidders or waiting longer to round up more bidders. He shows there is an optimal interior solution with finite time before trading and applies it to initial public offerings.

⁴⁴Lester, Visschers, and Wolthoff (2015) strengthen this "non-rivalry" condition—namely, the quadratic-search technology defined in section 2.4—by deriving a sufficient condition that is also necessary and which they call

buyers sort across posted price-trading probability pairs, revealing their types *ex ante*. But with competing auctions, buyers visit the sellers randomly and are screened *ex post*.

In a labor market setting, Shi (2002) and Shimer (2005) consider different forms of competition with the flavor of an auction, where the price paid depends on the composition of the *ex post* demand—specifically, the number and characteristics of agents that show up. With observable worker characteristics, in Shimer’s equilibrium, high-type workers obtain a job with the highest probability, while low types only succeed if no high types show up. This allocation is similar to that of a second-price auction. As Shi (2002) argues, this is not only realistic in many market settings, but is also important for the efficient allocation of resources.

4.4.5 Directed Search and Large Firms

The directed-search model can be used to analyze large firms in the presence of frictions. Under random search, Smith (1999) focuses on the role of hiring in large firms that have decreasing returns to employment. To understand how these firms set wages, Smith (1999) assumes a reduced-form bargaining process (in the spirit of an earlier paper by Stole and Zwiebel 1996) where each worker is treated as the marginal worker and wages depend on marginal productivity. He shows that the outcome is inefficient and leads to overemployment, with firm size larger than optimal. Kaas and Kircher (2015) and Schaal (forthcoming) propose a directed-search model with large firms, finding that price posting with coordination frictions yields a constrained efficient surplus division. These search models can handle realistic environments where firms hire multiple workers and technology is nonadditive.

This is useful for applied analyses of matching workers to firms, and where firm size is endogenous.⁴⁵ Finally, Eeckhout and Kircher (2012) (in section 4.1) contains an extension with directed search and derives a sorting result with heterogeneous workers and large firms.

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 expand 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 Sorting in Static Models

5.1.1 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 because in 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 et al. (2014), who analyze matching with one sided incomplete information and TU.⁴⁶ Their

⁴⁵For example, Sepahsafari (2016) analyzes cyclical variations in the presence of credit frictions.

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

“invariance,” i.e., the action of one buyer does not affect the distribution of buyers in the submarket.

incomplete information stability notion is in the spirit of rationalizability in game theory, rather than mechanism design. An interesting result they show is that a mild strengthening of supermodularity yields PAM under incomplete information.

5.1.2 *Sorting with Signaling 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 signaling 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, everyone 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 signaling 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 $2E[x]E[y] = 2\left(\int_0^1 xg(x) dx\right)^2$. Under perfect PAM, the total expected output is $2\int_0^1 x^2g(x) dx$. If both sides use the same signaling quantity

function β , 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. Since every type x must optimally bid as if it had type x , the envelope theorem yields $\pi'(x) = x$, and so $\pi(x) = \int_0^x s ds$. Altogether,

$$\begin{aligned} \beta(x) &= x^2 - \pi(x) \\ &= x^2 - \int_0^x s ds = \int_0^x s ds, \end{aligned}$$

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

$$\begin{aligned} &2\int_0^1 x^2g(x) dx - 2\int_0^1 \left(\int_0^x s ds\right)g(x) dx \\ &= \int_0^1 x^2g(x) dx. \end{aligned}$$

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

$$\begin{aligned} (30) \quad &\int_0^1 x^2g(x) dx - 2\left(\int_0^1 xg(x) dx\right)^2 \\ &= \text{var}(x) - E[x]^2 \\ &= E[x]^2(CV(x)^2 - 1), \end{aligned}$$

where $CV(x) = \sqrt{\text{var}(x)}/E[x]$ is the coefficient of variation of x . Thus, if $CV(x) < 1$ random matching outperforms PAM, and the opposite holds 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

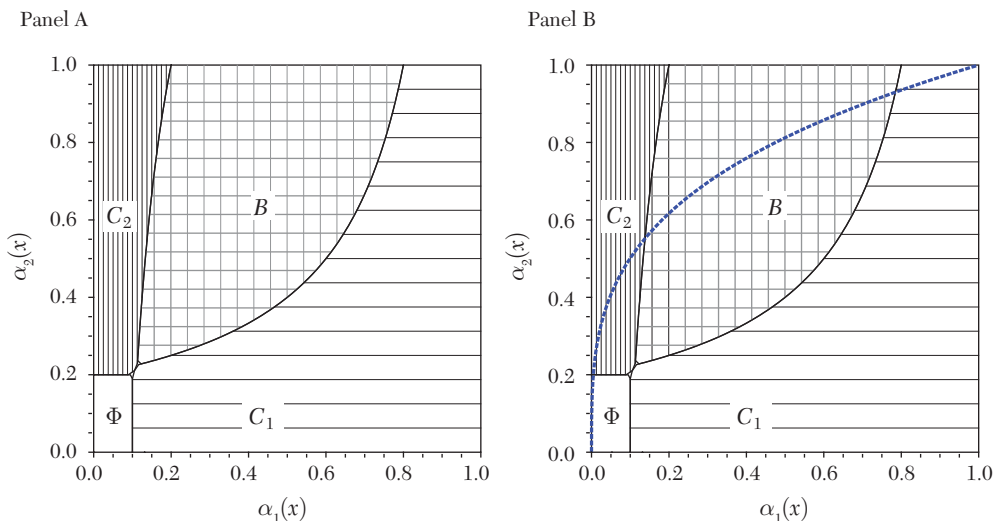


Figure 2. Students' Portfolio Problem

Notes: 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 $\psi(\alpha_1) = \alpha_1^{\sigma/\sigma_1}$, which arises with exponential signals $m(\sigma|x) = (1/x)e^{-\sigma/x}$. 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.

Source: Chade, Lewis, and Smith (2014).

monotone hazard rate functions yields strong predictions regarding the welfare under random matching compared to PAM.

5.1.3 College Student Matching

Gale and Shapley (1962) ignore the wealth of search and information frictions that afflict the sorting of students into colleges. Chade, Lewis, and Smith (2014) explore how students and colleges react to these frictions and what happens to sorting.

In their model, there are 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.⁴⁷ 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.

The paper largely focuses on the case when students know their type, but colleges only observe a noisy conditionally independent signal of each applicant. Signal outcomes σ

⁴⁷ Enrollment here is obviously deterministic. Che and Koh (2015) explore a different model of college admissions with stochastic enrollments, instead.

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 $M(\sigma|x) \rightarrow 0$ as $x \rightarrow \infty$ and $M(\sigma|x) \rightarrow 1$ as $x \rightarrow 0$ for any interior σ .

Students choose a portfolio of college applications $S(x) \in \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$ for each x , while colleges set admissions standards $\underline{\sigma}_i$, such that college i admits students with signal realizations above $\underline{\sigma}_i$. 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 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 (2014). Consider a student with respective admission chances $0 \leq \alpha_1, \alpha_2 \leq 1$. Using the simultaneous-search solution in section 3.2, we obtain the student's optimal portfolio choice. 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 given by $MB_{21} = (1 - \alpha_1)\alpha_2 u$ and $MB_{12} = \alpha_1(1 - \alpha_2)u$. The plot of these two curves looks like figure 2 when $c < u(1 - u)$ and $c < u/4$, i.e., with applications not too costly.

This optimal-decision rule neatly partitions the unit square into four application regions, corresponding to the four portfolio choices, denoted Φ, C_2, B, C_1 , shaded in the right panel of figure 2. 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 acceptance chances for each type x given colleges thresholds:

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

Although the acceptance function need not in general be concave as in figure 2, it does have a falling secant: α_2/α_1 is a decreasing function. The acceptance function and the application strategy respectively capture opportunities and preferences for student applications. Superimposing them, figure 2 (right panel) depicts a monotone application strategy, in which higher types apply more aggressively to college. And since $\underline{\sigma}_1 > \underline{\sigma}_2$

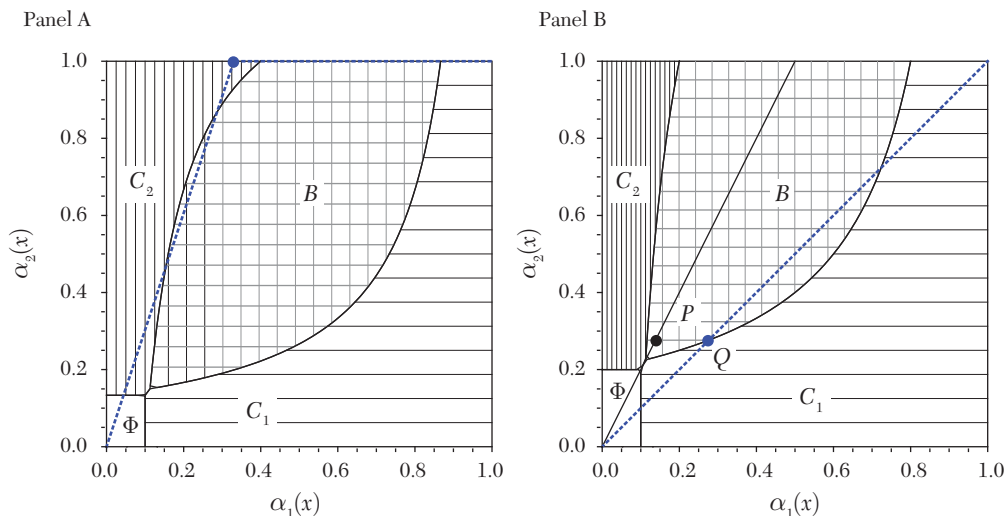


Figure 3. Non-Monotone Behavior

Notes: 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.

Source: Chade, Lewis, and Smith (2014).

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 3. 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 3, 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.

The second violation occurs when the lesser college imposes a higher admissions

threshold. To see this, assume that both colleges set the same thresholds. As seen in the right panel of figure 3, 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 small enough capacity of college 2. For the paper shows that college 2 imposes a higher standard than college 1 if its capacity is small enough—thus explaining how a poorly ranked small private college can nonetheless impose higher standards in equilibrium than a much larger public university.⁴⁸ The paper

⁴⁸The sorting failures can be drastic. For instance, consider the right panel of figure 3. 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.

also shows that all equilibria exhibit sorting if college 2 is sufficiently worse than college 1 (specifically, $u \leq 0.5$), and college 1 is small enough in capacity relative to college 2: Graphically, in this case the acceptance function traverses the unit square high enough as to preclude the case in the right panel of figure 3.

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.

5.2 *Sorting in Dynamic Models*

5.2.1 *Sorting with Evolving Reputations*

Anderson and Smith (2010) ask 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.⁴⁹ 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 agents’ attributes that may allow them to improve

⁴⁹Early 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 on complementarities and sorting patterns.

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 nor 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 exist, and prove the welfare theorems. We illustrate this finding in their simpler motivational two-period partnership model.

Anderson and Smith (2010) 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, \dots, 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

$$\begin{aligned} f(x, y) &= \sum_i q_i p_i(x, y) \\ &= xyH + [x(1 - y) + y(1 - x)]M \\ &\quad + (1 - x)(1 - y)L. \end{aligned}$$

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 (Becker 1973).

Consider 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 , Anderson and Smith (2010) notice that if he matches with y , then after observing output q_i in the first period x , he updates his belief that his type is high to $z_i(x, y) = p_i(1, y)x/p_i(x, y)$.⁵⁰ Since the expected continuation payoff for x is $\psi(x|y) = \sum_i p_i(x, y)w(z_i(x, y))$, the present value of a match between agents 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 $\delta < 1$.

Anderson and Smith (2010) then make a key preliminary observation. If $\delta = 1$, then $v(x, y) = \psi(x|y) + \psi(y|x)$ and only the continuation payoff matters for matching. They show that ψ is strictly convex in x and in y .⁵¹ We now ask whether the

matching exhibits PAM. To see this, consider three pairs $(0, 0)$, $(1, 1)$, and (x, x) , where $x \in (0, 1)$. Strict 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 Bayesian updating when either of these extreme types match with anyone. So either $\psi(x|0) + \psi(0|x) > \psi(x|x) + \psi(0|0)$ or $\psi(x|1) + \psi(1|x) > \psi(x|x) + \psi(1|1)$. Hence, PAM fails since rematching x agents with either 0 or 1 raises total payoffs. Since this holds for $\delta = 1$, by continuity PAM fails for a high enough discount factor δ . Intuitively, the learning value of matching outweighs the productive complementarities in this case. Since any nonproductive variability in a match with an extreme type (0 or 1) reflects uncertainty about the uncertain $x \in (0, 1)$, assortatively matching x is intuitively informationally dominated by cross-matching them with type 0 or 1.

So sufficiently forward-thinking behavior leads to a failure of PAM. Unfortunately, as the discount factor rises to one in an infinite-horizon model, the continuation value tends to linear. For intuitively, in the perfect patience limit, almost all production arises when one perfectly knows 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 turns on the asymptotic behavior of the continuation value function, that the second derivative explodes near 0 and 1.

⁵⁰This is just an application of Bayes's 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 .

⁵¹For any information about one's own or one's partner's type is intuitively productively valuable to the social

planner in assigning matches, and also induces mean zero noise in the posterior reputation. Since all zero-mean gambles have positive expected value, both strict convexity claims follow from Pratt (1964).

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.

5.2.2 *Sorting and Evolving Types*

Inspired by the changing reputational types in Anderson and Smith (2010), Anderson (2015) explores the dynamics that arise when individuals are changed by the association with their match partners. Assume an initial distribution over human capital G .⁵² 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.⁵³ Given the final wage $w(x) \equiv f(x, x)/2$, the period-one continuation value is:

$$(31) \quad V(\mu) \equiv \int w(x) dH(x|\mu).$$

For a high enough discount factor, PAM is initially optimal when it maximizes (31) across all feasible matchings. Since the wage $w(x)$ is increasing, PAM maximizes (31) 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 $\mathcal{T}(s|x, y) = \mathbf{1}_{s \geq \tau(x, y)}$. 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 results to the infinite-horizon model, Anderson (2015) first analyzes the planner’s preferences over human capital distributions. These analytical 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.

⁵²Jovanovic (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.

⁵³Anderson and Smith (2010) establish the welfare theorems for this dynamic matching model. In particular, PAM is optimal if and only if PAM is a market outcome.

5.2.3 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.⁵⁴ 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 positively assortatively matches men of age 2 and the best 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 men of age 1. Since the type of every age 2 man x is revealed, he will be assigned to marry woman $\mu(x)$, where $\mu'(x) > 0$ and $\mu(1) = 1$. The remaining lesser women are randomly assigned to age 1 males who opt to enter the marriage market when young, whose true types are as yet hidden. Any unmatched age 1 men or women remain 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 mass of each entering period 2, so that everyone eventually matches. We now explore which men choose to marry when young.

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 $C \subset [0, 1]$ seek to marry age 1 men. Then a type x man strictly prefers to delay marriage until age 2 when

$$(32) \quad -c_1 + f_1(x)f_2(\mu(x)) \geq \frac{\int_C f_1(x)f_2(s) dH(s)}{\int_C dH(s)}.$$

⁵⁴In 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).

Easily, if this inequality holds for any type x , then it also holds for any higher type, thereby confirming the cutoff value property. 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 (32) with equality, namely:

$$(33) \quad f_1(\bar{x}) \left(f_2(\mu(\bar{x})) - \frac{\int_0^{\mu(\bar{x})} f_2(s) dH(s)}{H(\mu(\bar{x}))} \right) = c_1.$$

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 (33) has a unique solution. First, the left side of (33) vanishes in the limit $\bar{x} \downarrow 0$ by l'Hopital's rule. Since f_2 is increasing, the left side of (33) exceeds c_1 at $\bar{x} = 1$, for small enough $c_1 > 0$. Existence follows by continuity. Uniqueness follows if the left side of (33) 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

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

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' strictly 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 log concave. If we assume a strictly 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.

5.2.4 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 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 for the information in the event that the partner agrees to match. And if agents on the other side of the market grow more choosy 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 women's types $x \in [0, 1]$ is $g(x)$, and of men'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 woman x meets a man y , he observes a signal $\sigma \in [0, 1]$ drawn from $m(\sigma|x)$, and she 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 man of type y or a woman 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 man of type y facing a population of women. 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 he is accepted by the current partner. Formally,

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

Consider a man y seeing a signal σ . That woman accepts with probability $a(y|\sigma)$. In this event, the man 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 woman does not accept, which occurs with probability $1 - a(y|\sigma)$, then the man continues to search. His Bellman equation is thus:

$$\begin{aligned} (34) \quad v(y|\sigma) &= a(y|\sigma) \max\{f(y|\sigma), \delta\Psi(y)\} \\ &\quad + (1 - a(y|\sigma))\delta\Psi(y), \end{aligned}$$

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

Similarly, the optimal strategy of a woman 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 section 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 curve emerges: For since $f(y|\sigma) \leq E[X/(1-\delta)|\sigma]$, the event of being accepted is a discouraging signal for a man 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 corporate finance to marriage markets. The many variations of the model allow for a simple relation between technology and the resulting sorting pattern.

We have also explored sorting with 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 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, and 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 with whom to trade should prove useful in all fields of economics.

REFERENCES

- Acemoglu, Daron, and Robert Shimer. 1999. "Holdups and Efficiency with Search Frictions." *International Economic Review* 40 (4): 827–49.
- Ackerberg, Daniel A., and Maristella Botticini. 2002. "Endogenous Matching and the Empirical Determinants of Contract Form." *Journal of Political Economy* 110 (3): 564–91.
- Adachi, Hiroyuki. 2003. "A Search Model of Two-Sided Matching under Nontransferable Utility." *Journal of Economic Theory* 113 (2): 182–98.
- Adam, Klaus. 2001. "Learning While Searching for the Best Alternative." *Journal of Economic Theory* 101 (1): 252–80.
- Albrecht, James, Axel Anderson, and Susan Vroman. 2010. "Search by Committee." *Journal of Economic Theory* 145 (4): 1386–407.
- Albrecht, James, Pieter A. Gautier, and Susan Vroman. 2006. "Equilibrium Directed Search with Multiple Applications." *Review of Economic Studies* 73 (4): 869–91.
- Anderson, Axel. 2015. "A Dynamic Generalization of Becker's Assortative Matching Result." *Journal of Economic Theory* 159 (Part A): 290–310.
- Anderson, Axel, and Lones Smith. 2010. "Dynamic Matching and Evolving Reputations." *Review of Economic Studies* 77 (1): 3–29.
- Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg. 2006. "Offshoring in a Knowledge Economy." *Quarterly Journal of Economics* 121 (1): 31–77.
- Ashlagi, Itai, Yash Kanoria, and Jacob D. Leshno. 2017. "Unbalanced Random Matching Markets: The Stark Effect of Competition." *Journal of Political Economy* 125 (1): 69–98.
- Atakan, Alp E. 2006. "Assortative Matching with Explicit Search Costs." *Econometrica* 74 (3): 667–80.
- Athey, Susan. 2001. "Single Crossing Properties and the Existence of Pure Strategy Equilibria in Games of Incomplete Information." *Econometrica* 69 (4): 861–89.
- Athey, Susan, Paul R. Milgrom, and John Roberts. 1998. "Robust Comparative Statics." Unpublished.
- Bagger, Jesper, and Rasmus Lentz. 2014. "An Empirical Model of Wage Dispersion with Sorting." National Bureau of Economic Research Working Paper 20031.
- Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-Concave Probability and Its Applications." *Economic Theory* 26 (2): 445–69.
- Barlow, Richard E., and Frank Proschan. 1996. *Mathematical Theory of Reliability*. Philadelphia: Society for Industrial and Applied Mathematics.
- Bartolucci, Cristian, and Francesco Devicienti. 2013. "Better Workers Move to Better Firms: A Simple Test to Identify Sorting." Institute for the Study of Labor Discussion Paper 7601.
- Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46.
- Bergstrom, Theodore C., and Mark Bagnoli. 1993.

- “Courtship as a Waiting Game.” *Journal of Political Economy* 101 (1): 185–202.
- Bhaskar, V., and Ed Hopkins. 2016. “Marriage as a Rat Race: Noisy Premarital Investments with Assortative Matching.” *Journal of Political Economy* 124 (4): 992–1045.
- Bloch, Francis, and Harl Ryder. 2000. “Two-Sided Search, Marriages, and Matchmakers.” *International Economic Review* 41 (1): 93–115.
- Browning, Martin, Pierre-André Chiappori, and Yoram Weiss. 2014. *Economics of the Family*. Cambridge and New York: Cambridge University Press.
- Burdett, Kenneth, and Melvyn G. Coles. 1997. “Marriage and Class.” *Quarterly Journal of Economics* 112 (1): 141–68.
- Burdett, Kenneth, and Kenneth L. Judd. 1983. “Equilibrium Price Dispersion.” *Econometrica* 51 (4): 955–69.
- Burdett, Kenneth, and Dale T. Mortensen. 1998. “Wage Differentials, Employer Size, and Unemployment.” *International Economic Review* 39 (2): 257–73.
- Burdett, Kenneth, Shouyong Shi, and Randall Wright. 2001. “Pricing and Matching with Frictions.” *Journal of Political Economy* 109 (5): 1060–85.
- Butters, Gerard R. 1977. “Equilibrium Distributions of Sales and Advertising Prices.” *Review of Economic Studies* 44 (3): 465–91.
- Chade, Hector. 2001. “Two-Sided Search and Perfect Segregation with Fixed Search Costs.” *Mathematical Social Sciences* 42 (1): 31–51.
- Chade, Hector. 2006. “Matching with Noise and the Acceptance Curse.” *Journal of Economic Theory* 129 (1): 81–113.
- Chade, Hector, and Jan Eeckhout. 2016. “Stochastic Sorting.” Unpublished.
- Chade, Hector, and Jan Eeckhout. 2015. “Competing Teams.” Unpublished.
- Chade, Hector, and Jan Eeckhout. Forthcoming. “Matching Information.” *Theoretical Economics*.
- Chade, Hector, Gregory Lewis, and Lones Smith. 2014. “Student Portfolios and the College Admissions Problem.” *Review of Economic Studies* 81 (3): 971–1002.
- Chade, Hector, and Isle Lindenlaub. 2015. “Risky Matching.” Unpublished.
- Chade, Hector, and Lones Smith. 2006. “Simultaneous Search.” *Econometrica* 74 (5): 1293–307.
- Chakraborty, Archisman, Alessandro Citanna, and Michael Ostrovsky. 2010. “Two-Sided Matching with Interdependent Values.” *Journal of Economic Theory* 145 (1): 85–105.
- Che, Yeon-Koo, and Youngwoo Koh. 2016. “Decentralized College Admissions.” *Journal of Political Economy* 124 (5): 1295–338.
- Chiappori, Pierre-André, Robert J. McCann, and Lars P. Nesheim. 2010. “Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness.” *Economic Theory* 42 (2): 317–54.
- Chiappori, Pierre-André, and Philip J. Reny. 2016. “Matching to Share Risk.” *Theoretical Economics* 11 (1): 227–51.
- Choi, Michael, and Lones Smith. 2016. “Optimal Sequential Search among Alternatives.” University of Wisconsin PhD. Thesis.
- Chow, Yuan Shih, Herbert Robbins, and David Siegmund. 1971. *Great Expectations: The Theory of Optimal Stopping*. Boston: Houghton Mifflin.
- Chvátal, Vašek. 1983. *Linear Programming*. New York: W. H. Freeman and Company.
- Cole, Harold L., George J. Mailath, and Andrew Postlewaite. 2001. “Efficient Non-contractible Investments in Large Economies.” *Journal of Economic Theory* 101 (2): 333–73.
- Compte, Olivier, and Philippe 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. “Rearrangements of Functions.” *Journal of Functional Analysis* 66 (3): 432–38.
- Damiano, Ettore, Hao Li, and Wing Suen. 2005. “Unraveling of Dynamic Sorting.” *Review of Economic Studies* 72 (4): 1057–76.
- DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. Hoboken, NJ: Wiley.
- Diamond, Peter A. 1982. “Aggregate Demand Management in Search Equilibrium.” *Journal of Political Economy* 90 (5): 881–94.
- Diamond, Peter A., and Eric Maskin. 1979. “An Equilibrium Analysis of Search and Breach of Contract, I: Steady States.” *Bell Journal of Economics* 10 (1): 282–316.
- Edlin, Aaron S., and Chris Shannon. 1998. “Strict Single Crossing and the Strict Spence–Mirrlees Condition: A Comment on Monotone Comparative Statics.” *Econometrica* 66 (6): 1417–25.
- Eeckhout, Jan. 1999. “Bilateral Search and Vertical Heterogeneity.” *International Economic Review* 40 (4): 869–87.
- Eeckhout, Jan, and Philipp Kircher. 2010a. “Sorting and Decentralized Price Competition.” *Econometrica* 78 (2): 539–74.
- Eeckhout, Jan, and Philipp Kircher. 2010b. “Sorting versus Screening: Search Frictions and Competing Mechanisms.” *Journal of Economic Theory* 145 (4): 1354–85.
- Eeckhout, Jan, and Philipp Kircher. 2011. “Identifying Sorting-In Theory.” *Review of Economic Studies* 78 (3): 872–906.
- Eeckhout, Jan, and Philipp Kircher. 2012. “Assortative Matching with Large Firms: Span of Control over More versus Better Workers.” Unpublished.
- Felli, Leonardo, and Kevin Roberts. 2016. “Does Competition Solve the Hold-Up Problem?” *Economica* 83 (329): 172–200.
- Ferguson, Thomas S. 2016. “Optimal Stopping and Applications.” <http://www.math.ucla.edu/~tom/Stopping/Contents.html>.
- Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, MA and London: MIT Press.
- Gabaix, Xavier, and Augustin Landier. 2008. “Why Has

- CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123 (1): 49–100.
- Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69 (1): 9–15.
- Galenianos, Manolis, and Philipp Kircher. 2009. "Directed Search with Multiple Job Applications." *Journal of Economic Theory* 144 (2): 445–71.
- Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904.
- Gershkov, Alex, and Benny Moldovanu. 2012. "Optimal Search, Learning and Implementation." *Journal of Economic Theory* 147 (3): 881–909.
- Gittins, J. C. 1979. "Bandit Processes and Dynamic Allocation Indices." *Journal of the Royal Statistical Society, Series B* 41 (2): 148–77.
- Gittins, J. C., and D. M. Jones. 1974. "A Dynamical Allocation Index for the Sequential Design of Experiments." In *Progress in Statistics*, edited by Joseph Mark Gani, et al. Amsterdam: North-Holland.
- Gretsky, Neil E., Joseph M. Ostroy, and William R. Zame. 1992. "The Nonatomic Assignment Model." *Economic Theory* 2 (1): 103–27.
- Gretsky, Neil E., Joseph M. Ostroy, and William R. Zame. 1999. "Perfect Competition in the Continuous Assignment Model." *Journal of Economic Theory* 88 (1): 60–118.
- Grossman, Gene M., Elhanan Helpman, and Philipp Kircher. 2013. "Matching and Sorting in a Global Economy." National Bureau of Economic Research Working Paper 19513.
- Grossman, Gene M., and Giovanni Maggi. 2000. "Diversity and Trade." *American Economic Review* 90 (5): 1255–75.
- Guadalupe, Maria, Veronica Rappoport, Bernard Salanie, and Catherine Thomas. 2014. "The Perfect Match: Assortative Matching in International Acquisitions." Unpublished.
- Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2012. "Identifying Equilibrium Models of Labor Market Sorting." Unpublished.
- Hardy, G. H., J. E. Littlewood, and G. Polya. 1952. *Inequalities*, Second edition. Cambridge and New York: Cambridge University Press.
- Hatfield, John William, and Fuhito Kojima. 2008. "Matching with Contracts: Comment." *American Economic Review* 98 (3): 1189–94.
- Hatfield, John William, and Paul R. Milgrom. 2005. "Matching with Contracts." *American Economic Review* 95 (4): 913–35.
- Heckman, James J., and Bo E. Honore. 1990. "The Empirical Content of the Roy Model." *Econometrica* 58 (5): 1121–49.
- Holmstrom, Bengt, and Paul R. Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55 (2): 303–28.
- Hoppe, Heidrun C., Benny Moldovanu, and Aner Sela. 2009. "The Theory of Assortative Matching Based on Costly Signals." *Review of Economic Studies* 76 (1): 253–81.
- Jacquet, Nicolas L., and Serene Tan. 2007. "On the Segmentation of Markets." *Journal of Political Economy* 115 (4): 639–64.
- Jerez, Belen. 2014. "Competitive Equilibrium with Search Frictions: A General Equilibrium Approach." *Journal of Economic Theory* 153: 252–86.
- Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (5 Part 1): 972–90.
- Jovanovic, Boyan. 1984. "Matching, Turnover, and Unemployment." *Journal of Political Economy* 92 (1): 108–22.
- Jovanovic, Boyan. 2014a. "Learning and Recombination." Unpublished.
- Jovanovic, Boyan. 2014b. "Misallocation and Growth." *American Economic Review* 104 (4): 1149–71.
- Kaas, Leo, and Philipp Kircher. 2015. "Efficient Firm Dynamics in a Frictional Labor Market." *American Economic Review* 105 (10): 3030–60.
- Kantorovich, L. V. 1942. "On the Translocation of Masses." *Dokl. Akad. Nauk SSSR* 37 (7–8): 227–29.
- Kelso, Alexander S, Jr., and Vincent P. Crawford. 1982. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica* 50 (6): 1483–504.
- Kircher, Philipp. 2009. "Efficiency of Simultaneous Search." *Journal of Political Economy* 117 (5): 861–913.
- Koopmans, Tjalling C., and Martin Beckmann. 1957. "Assignment Problems and the Location of Economic Activities." *Econometrica* 25 (1): 53–76.
- Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics* 108 (3): 551–75.
- Kremer, Michael, and Eric Maskin. 1996. "Wage Inequality and Segregation by Skill." Unpublished.
- Lamadon, Thibaut. 2014. "Productivity Shocks, Dynamic Contracts and Income Uncertainty." Unpublished.
- Lamadon, Thibaut, Jeremy Lise, Costas Meghir, and Jean-Marc Robin. 2013. "Matching, Sorting, and Wages." Unpublished.
- Lauermann, Stephan, and Georg Noldeke. 2014. "Stable Marriages and Search Frictions." *Journal of Economic Theory* 151: 163–95.
- Legros, Patrick, and Andrew F. Newman. 2007. "Beauty Is a Beast, Frog Is a Prince: Assortative Matching with Nontransferabilities." *Econometrica* 75 (4): 1073–102.
- Legros, Patrick, and Andrew F. Newman. 2010. "Co-ranking Mates: Assortative Matching in Marriage Markets." *Economics Letters* 106 (3): 177–79.
- Lester, Benjamin, Ludo Visschers, and Ronald Wolthoff. 2015. "Meeting Technologies and Optimal Trading Mechanisms in Competitive Search Markets." *Journal of Economic Theory* 155: 1–15.
- Lindenlaub, Ilse. 2017. "Sorting Multidimensional Types: Theory and Application." *Review of Economic Studies* 84 (2): 718–89.
- Lise, Jeremy, Costas Meghir, and Jean-Marc Robin. 2013. "Mismatch, Sorting and Wage Dynamics." National Bureau of Economic Research Working

- Paper 18719.
- Lise, Jeremy, and Jean-Marc Robin. 2017. "The Macrodynamics of Sorting between Workers and Firms." *American Economic Review* 107 (4): 1104–35.
- Liu, Qingmin, George J. Mailath, Andrew Postlewaite, and Larry Samuelson. 2014. "Stable Matching with Incomplete Information." *Econometrica* 82 (2): 541–87.
- Lopes de Melo, Rafael. 2009. "Sorting in the Labor Market: Theory and Measurement." Yale University Ph.D. Thesis.
- Lorentz, G. G. 1953. "An Inequality for Rearrangements." *American Mathematical Monthly* 60 (3): 176–79.
- Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (2): 508–23.
- Makowski, Louis, and Joseph M. Ostroy. 1995. "Appropriation and Efficiency: A Revision of the First Theorem of Welfare Economics." *American Economic Review* 85 (4): 808–27.
- Manea, Mihai. 2017. "Steady States in Matching and Bargaining." *Journal of Economic Theory* 167: 206–28.
- McAfee, R. Preston. 1993. "Mechanism Design By Competing Sellers." *Econometrica* 61 (6): 1281–312.
- McCall, John J. 1965. "The Economics of Information and Optimal Stopping Rules." *Journal of Business* 38 (3): 300–317.
- McCall, John J. 1970. "Economics of Information and Job Search." *Quarterly Journal of Economics* 84 (1): 113–26.
- McNamara, J. M., and E. J. Collins. 1990. "The Job Search Problem as an Employer–Candidate Game." *Journal of Applied Probability* 28: 815–27.
- Milgrom, Paul R., and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62 (1): 157–80.
- Milgrom, Paul R., and Robert J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50 (5): 1089–122.
- Moen, Espen R. 1997. "Competitive Search Equilibrium." *Journal of Political Economy* 105 (2): 385–411.
- Monge, Gaspard. 1781. "Memoire sur la Theorie des Deblais et des Remblais." *Histoire de l'Academie Royale des Sciences* 666–704.
- Morgan, Peter. 1996. "Two-Sided Search and Matching." Unpublished.
- Morgan, Peter, and Richard Manning. 1985. "Optimal Search." *Econometrica* 53 (4): 923–44.
- Mortensen, Dale T., and Randall Wright. 2002. "Competitive Pricing and Efficiency in Search Equilibrium." *International Economic Review* 43 (1): 1–20.
- Noldeke, Georg, and Larry Samuelson. 2015. "Investment and Competitive Matching." *Econometrica* 83 (3): 835–96.
- Noldeke, Georg, and T. Troger. 2009. "Matching Heterogeneous Agents with a Linear Search Technology." Unpublished.
- Olszewski, Wojciech, and Richard Weber. 2015. "A More General Pandora Rule?" *Journal of Economic Theory* 160: 429–37.
- Osborne, Martin J., and Ariel Rubinstein. 1990. *Bargaining and Markets*. Bingley, UK: Emerald.
- Peskir, Goran, and Albert Shiryaev. 2006. *Optimal Stopping and Free-Boundary Problems*. Basel and Boston: Birkhauser Verlag.
- Peters, Michael. 1984. "Bertrand Equilibrium with Capacity Constraints and Restricted Mobility." *Econometrica* 52 (5): 1117–27.
- Peters, Michael. 1997. "A Competitive Distribution of Auctions." *Review of Economic Studies* 64 (1): 97–123.
- Peters, Michael, and Sergei Severinov. 1997. "Competition among Sellers Who Offer Auctions Instead of Prices." *Journal of Economic Theory* 75 (1): 141–79.
- Petrongolo, Barbara, and Christopher A. Pissarides. 2001. "Looking into the Black Box: A Survey of the Matching Function." *Journal of Economic Literature* 39 (2): 390–431.
- Pinheiro, Roberto. 2012. "Venture Capital and Underpricing: Capacity Constraints and Early Sales." Unpublished.
- Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70 (6): 2295–350.
- Pratt, John W. 1964. "Risk Aversion in the Small and in the Large." *Econometrica* 32 (1–2): 122–36.
- Pycia, Marek, and M. Bumin Yenmez. 2015. "Matching with Externalities." Unpublished.
- Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature* 43 (4): 959–88.
- Rosen, Sherwin. 1981. "The Economics of Superstars." *American Economic Review* 71 (5): 845–58.
- Ross, Sheldon M. 1983. *Introduction to Stochastic Dynamic Programming*. London: Academic Press.
- Roth, Alvin E. 1989. "Two-Sided Matching with Incomplete Information about Others' Preferences." *Games and Economic Behavior* 1 (2): 191–209.
- Roth, Alvin E., and Marilda A. Oliveira Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge and New York: Cambridge University Press.
- Rothschild, Michael. 1974. "Searching for the Lowest Price When the Distribution of Prices Is Unknown." *Journal of Political Economy* 82 (4): 689–711.
- Sasaki, Hiroo, and Manabu Toda. 1996. "Two-Sided Matching Problems with Externalities." *Journal of Economic Theory* 70 (1): 93–108.
- Sattinger, Michael. 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature* 31 (2): 831–80.
- Schaal, Edouard. Forthcoming. "Uncertainty and Unemployment." *Econometrica*.
- Schulhofer-Wohl, Sam. 2006. "Negative Assortative Matching of Risk-Averse Agents with Transferable Expected Utility." *Economics Letters* 92 (3): 383–88.
- Sepahsalar, Alireza. 2016. "Financial Market Imperfections and Labour Market Outcomes." Unpublished.

- Serfes, Konstantinos. 2005. "Risk Sharing vs. Incentives: Contract Design under Two-Sided Heterogeneity." *Economics Letters* 88 (3): 343–49.
- Shapley, L. S., and M. Shubik. 1971. "The Assignment Game I: The Core." *International Journal of Game Theory* 1 (1): 111–30.
- Shi, Shouyong. 2001. "Frictional Assignment I: Efficiency." *Journal of Economic Theory* 98 (2): 232–60.
- Shi, Shouyong. 2002. "A Directed Search Model of Inequality with Heterogeneous Skills and Skill-Biased Technology." *Review of Economic Studies* 69 (2): 467–91.
- Shimer, Robert. 2005. "The Assignment of Workers to Jobs in an Economy with Coordination Frictions." *Journal of Political Economy* 113 (5): 996–1025.
- Shimer, Robert, and Lones Smith. 2000. "Assortative Matching and Search." *Econometrica* 68 (2): 343–69.
- Shimer, Robert, and Lones Smith. 2001. "Nonstationary Search." Unpublished.
- Shiryaev, A. N. 1978. *Optimal Stopping Rules*. New York and Berlin: Springer.
- Smith, Eric. 1999. "Search, Concave Production, and Optimal Firm Size." *Review of Economic Dynamics* 2 (2): 456–71.
- Smith, Lones. 1992. "Cross-Sectional Dynamics in a Two-Sided Matching Model." Unpublished.
- Smith, Lones. 1995. "A Model of Exchange Where Beauty Is in the Eye of the Beholder." Unpublished.
- Smith, Lones. 1997. "The Marriage Model with Search Frictions." Unpublished.
- Smith, Lones. 2006. "The Marriage Model with Search Frictions." *Journal of Political Economy* 114 (6): 1124–44.
- Smith, Lones. 2011. "Frictional Matching Models." *Annual Review of Economics* 3: 319–38.
- Sørensen, Morten. 2007. "How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital." *Journal of Finance* 62 (6): 2725–62.
- Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 69 (3): 213–25.
- Stole, Lars A., and Jeffrey Zwiebel. 1996. "Intra-firm Bargaining under Non-binding Contacts." *Review of Economic Studies* 63 (3): 375–410.
- Tervio, Marko. 2008. "The Difference that CEOs Make: An Assignment Model Approach." *American Economic Review* 98 (3): 642–68.
- Teulings, Coen N., and Pieter A. Gautier. 2004. "The Right Man for the Job." *Review of Economic Studies* 71 (2): 553–80.
- Topkis, Donald M. 1998. *Supermodularity and Complementarity*. Princeton and Oxford: Princeton University Press.
- Varian, Hal R. 1999. "Economics and Search." Unpublished.
- Villani, Cédric. 2009. *Optimal Transport: Old and New*. Berlin: Springer.
- Vince, A. 1990. "A Rearrangement Inequality and the Permutahedron." *American Mathematical Monthly* 97 (4): 319–23.
- Weitzman, Martin L. 1979. "Optimal Search for the Best Alternative." *Econometrica* 47 (3): 641–54.