Games and Economic Behavior 83 (2014) 207–223

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Imperfect competition in two-sided matching markets ☆

Eduardo M. Azevedo

Wharton, Business Economics and Public Policy Department, United States

ARTICLE INFO

Article history:
Received 20 June 2011
Available online 28 November 2013

JEL classification:
C78
D43
L13

Keywords:
Matching markets
Imperfect competition

ABSTRACT

This paper considers a simple equilibrium model of an imperfectly competitive two-sided matching market. Firms and workers may have heterogeneous preferences over matches on the other side, and the model allows for both uniform and personalized wages or contracts. To make the model tractable, I use the Azevedo and Leshno (2013) framework, in which a finite number of firms is matched to a continuum of workers.

In equilibrium, even if wages are exogenous and fixed, firms have incentives to strategically reduce their capacity, to increase the quality of their worker pool. The intensity of incentives to reduce capacity is given by a simple formula, analogous to the classic Cournot model, but depends on different moments of the distribution of preferences. I compare markets with uniform and personalized wages. For fixed quantities, markets with personalized wages always yield higher efficiency than markets with uniform wages, but may be less efficient if firms reduce capacity to avoid bidding too much for star workers.

1. Introduction

Two-sided matching markets are markets where participants on either side have preferences over who they interact with on the other side. Examples include matching CEOs to companies, students to colleges, advertisers to content providers, and many entry-level labor markets.[1] A stable matching is an allocation in which agents do not have incentives to break away from their matches and seek new ones. A well-known result by Roth (1985) shows that no mechanism that always produces a stable matching is strategy-proof for the firms. However, even though these markets have been studied by a large literature, most contributions ignore strategic behavior by firms, assuming them to command insignificant market share, or to act naïvely. This is in contrast to the standard approach in industrial organization, which typically focuses on Nash equilibrium in imperfectly competitive markets. This paper considers how the standard questions in imperfect competition models play out in matching markets. First, I investigate strategic quantity choices by firms, such as colleges or hospitals in centralized clearinghouses. Second, I consider the consequences of strategic behavior to equilibrium outcomes, and third, discuss a modest set of implications for the regulation and design of matching markets.

I consider a model analogous to Cournot oligopoly, but in a matching market. In the model, a number of firms compete to be matched to a set of workers. Both workers and firms have potentially heterogeneous preferences over match partners on the other side. I follow the literature on capacity manipulation games (Konishi and Ünver, 2006; Kojima, 2006; Mumcu and Saglam, 2009; Ehlers, 2010), assuming that firms strategically set capacity. Workers are then assigned according to a

☆ I would like to thank Nikhil Agarwal, Itai Ashlagi, Susan Athey, Eric Budish, Carlos E. da Costa, Mustafa Dogan, Drew Fudenberg, Felipe Iachan, Scott Kominers, Jacob Leshno, Alex Nichifor, Alvin Roth, Rodrigo Wagner, Yosuke Yasuda, two anonymous referees, the advisory editor and seminar participants at Harvard and USP for helpful comments and discussions. The paper benefited greatly from insightful suggestions from the editor and anonymous referees. Finally, I would like to thank Jonathan Kluberg, Ruben Lobel, Alexander Rikun, and Gareth Williams for hosting me at the MIT Operations Research Center where most of this paper was written.

E-mail address: eazevedo@wharton.upenn.edu.

[1] See Gabaix and Landier (2008), Becker (1973), Gale and Shapley (1962), Roth and Peranson (1999), Ginsburg and Wolf (2003).

0899-8256/$ – see front matter
http://dx.doi.org/10.1016/j.geb.2013.11.009

stable matching. This is analogous to the Cournot model, in which firms choose capacities, and prices are given by market clearing. The model departs from the literature in two key ways. First, instead of the standard Becker (1973) or Gale and Shapley (1962) frameworks, I use the Azevedo and Leshno (2013) model, in which a finite number of firms is matched to a continuum of workers.[2] The continuum model considerably simplifies the analysis, and reveals novel insights akin to standard price-theoretic analyses. Second, I consider the cases of both uniform and of personalized wages.

I first consider the case in which wages are exogenous, and uniform across workers. A surprising result from the matching literature on capacity manipulation is that firms may want to reduce quantities even if wages are fixed (Sönmez, 1997; Kesten, 2008). The intuition is that, by reducing capacity, firms may create rejection chains, in which a rejected worker causes further rejections, and eventually causes a better worker to apply to the firm. This can increase the quality of the worker pool available to the firm, at the expense of reducing the quantity of workers hired. Unsurprisingly, this result is still true in my model. However, unlike the previous literature, with the Azevedo and Leshno (2013) model, there is a simple first-order condition that quantifies the incentives to reduce capacity. The model shows that, if a firm is negligible compared to the rest of the market, it has no incentives to reduce capacity. However, if a firm has some market power, its marginal revenue from increasing capacity is lower than the productivity of a marginal worker. The reason for this is that, when a large firm hires more workers, it poaches employees from the competition, which leads competitors to be less selective. The first-order conditions show that the wedge in marginal revenue is proportional to the effect of a firm on the selectiveness of its competitors. This result clarifies the link between the matching literature on capacity manipulation games and the Cournot model.

Although the basic logic of the Cournot model extends to matching markets, some of the results have to be modified. For example, in the undifferentiated Cournot model the incentives for each firm to reduce capacity, as measured by the Lerner index, only depend on the inverse elasticity of demand and market share. This is not true in matching markets, where the incentives to shade depend on very different moments of the distribution of preferences. Therefore, the present model has a wealth of predictions that differ from the standard findings in homogeneous good markets. In particular, we highlight that, while the model assumptions are analogous to Cournot, the results often are not. Therefore, it is not appropriate to view every implication of the present model as analogous to implications of the standard Cournot model.

After considering uniform fixed wages, the paper considers matching markets with personalized wages. To my knowledge, this type of analysis has not been pursued in the literature for markets in which agents have heterogeneous preferences. It is shown that firms still have an incentive to reduce capacity, and a simple first-order condition quantifies by how much. Interestingly, the reasons for reducing capacity are quite different when wages are personalized. The gain is no longer caused by rejection chains, but by the fact that rejecting workers leads other firms to bid less aggressively for the best workers.

The paper then compares markets that have personalized wages and markets that have uniform wages. For example, in the market for junior associates in elite New York law firms, most firms pay every incoming lawyer the same wage. In contrast, senior lawyers are often paid personalized wages (Ginsburg and Wolf, 2003). A series of papers have debated the desirability of using personalized wages, which is a key market design variable. Notably, Bulow and Levin (2006) have shown that uniform wages may reduce matching efficiency, and compress wages. I show, however, that if firms choose capacities this conclusion may be reversed. In the imperfect competition model, there is a tradeoff. Personalized wages always generate higher matching efficiency for a given level of capacity, but they may increase firms' incentives to shade capacity. If firms are very similar, personalized wages have little impact on matching efficiency, but may induce firms to drastically reduce capacity to avoid entering into a bidding war for the best employees. In that case, markets organized around uniform wages generate higher welfare. However, if firms are more heterogeneous, the loss from matching inefficiency dominates the loss from capacity reduction, and personalized wages generate higher efficiency.

The paper proceeds as follows. Section 2 discusses the basic model, in which workers are paid uniform, fixed wages. Section 3 introduces personalized wages. Section 4 compares these two different institutional settings, and Section 5 concludes. Omitted proofs are in Appendix A.

2. Matching with uniform wages

2.1. Firms, workers, and stable matchings

A finite number I of firms compete for a continuum mass of workers. We use I to denote both the number of firms, and the set of firms $I = \{1, \ldots, I\}$. A particular firm is denoted by $i \in I$. Worker of type θ has productivity e_i^θ in $[0, 1]$ at firm i. Note that a worker's productivity may differ in different firms. We denote by e^θ the I-dimensional vector of worker productivity. Each worker has a complete strict preference ordering $\succ^\theta$ over the set of all firms, and over being unmatched. Formally, $\succ^\theta$ is defined over $I \cup \{\theta\}$, with θ representing being unmatched. Let $\mathcal{P}$ be the set of all such strict preference relations. The set of worker types is $\Theta = [0, 1]^I \times \mathcal{P}$. The distribution of workers is given by a finite measure η in Θ, defined over the σ-algebra containing all open sets.

[2] Azevedo and Leshno (2013) precede the present study. That paper introduces a matching model that allows for multidimensional heterogeneity in preferences and tractable derivation of comparative statics, two key ingredients in the present model of strategic firm behavior.

We impose a weak condition on the distribution of types, guaranteeing that stable matchings are unique. Assume henceforth that the measure

$$\eta\left(\left\{\theta\colon e^\theta \not< P,\ e^\theta < P'\right\}\right) > 0$$

for any two I-dimensional vectors $P \leqslant P'$ with $P \neq P'$. This is true, for example, if the distribution of productivity has full support.

The model builds upon the continuum matching model of Azevedo and Leshno (2013). A matching is a function $\mu : \Theta \cup I \to I \cup 2^\Theta$ such that

- Each worker is assigned either to a firm or to itself and each firm is assigned to a set of workers. That is, for all $\theta \in \Theta$, $\mu(\theta) \in I \cup \{\theta\}$, and for all $i \in I$, $\mu(i) \in 2^\Theta$.
- If a worker is matched to a firm, the firm is matched to the worker, and vice versa. That is, $i = \mu(\theta)$ iff $\theta \in \mu(i)$.
- μ is measurable with respect to the σ-algebra generated by the open sets of Θ.

Note that a worker being matched to itself is interpreted as being unmatched. Consider now a vector of capacities q. The standard market clearing concept in the literature, stability, is defined as follows.

Definition 1. A matching μ is stable with respect to $[\eta, q]$ if

1. For all i we have $\eta(\mu(i)) \leqslant q_i$.
2. If $\mu(\theta) = i$, then $i \succ^\theta \theta$.
3. If $i \succ^\theta \mu(\theta)$, then $\eta(\mu(i)) = q_i$, and for all $\theta' \in \mu(i)$ we have $e_i^{\theta'} \geqslant e_i^\theta$.
4. For all sequences θ^k, in which $\theta = \lim_{k\to\infty} \theta^k$, and all $e^{\theta^k} \geqslant e^\theta$, we have $\mu(\theta) = \lim_{k\to\infty} \mu(\theta^k)$.

Condition 1 asks that firms' capacity constraints are respected, condition 2 that no worker or firm receives an unacceptable match, condition 3 that no firm–worker pair could be made better off by matching with each other. Conditions 1–3 are analogous to those in the definition of a stable matching in the discrete college admissions model of Gale and Shapley (1962). Condition 4, which Azevedo and Leshno (2013) term right continuity, implies that, whenever a firm can hire extra workers without violating stability, it does so. It is a technical condition that eliminates multiplicities of stable matchings up to a measure 0 set of workers.

A stable matching is a resting point for the market, in which agents cannot gain by breaking off their matches and seeking new ones. Stability is a prominent solution concept in the literature, and is also the outcome of centralized clearinghouses that use stable matching mechanisms.

Under the above assumptions, Azevedo and Leshno's (2013) Theorem 1 guarantees that, for any vector q, a stable matching exists, and is unique. The intuition is that, in the continuum, the notion of stability is enough to uniquely clear the market, even though prices are not allowed to adjust. Later, when we consider personalized endogenous wages, we will see that stability still uniquely determines which workers are matched to which firms. However, the wages of workers that are not on the margin between two different firms are not uniquely determined.

2.2. The game

We now lay out the oligopoly game considered, in which the players are the firms I. The primitives are I, Θ, η, $c(\cdot) = (c_1(.), \ldots, c_I(.))$, a vector of wages $w = (w_1, \ldots, w_I)$, and $Q = (Q_1, \ldots, Q_I)$.

1. Firms simultaneously choose capacities q_i in compact intervals Q_i.
2. After capacity choices q, workers are hired according to the unique matching μ_q that is stable with respect to $[\eta, q]$.
3. Each firm's payoff is given by

$$\Pi_i(q) = \int\limits_{\mu_q(i)} \left[e_i^\theta - w_i\right] d\eta(\theta) - c_i(q_i), \tag{1}$$

   where w_i is firm i's wage, and the continuously differentiable function $c_i(\cdot)$ is the cost of investing in capacity. That is, profits are the integral of the productivity of the workers hired, net of the wage, minus the costs of investing in capacity.

Note that firm profits depend on wages w_i and on the cost of investing in capacity $c_i(\cdot)$. Moreover, firms are restricted to choose quantities in the compact intervals Q_i.

The game corresponds to a situation in which firms simultaneously commit to capacity choices. Workers are then assigned according to a stable matching, holding capacities fixed. The model is the matching analogue of the Cournot (1838) model in homogeneous good markets. Momentarily, we will relax the assumption that wages are uniform and fixed.

Because stable matchings are used to model centralized and decentralized outcomes, this model can correspond to a variety of situations, which need not be labor markets. For concreteness, we will maintain the firm/worker terminology, but there are other interesting examples. First, several labor market clearinghouses around the world use stable matching mechanisms. For example, in the United States, every year over 25 000 new medical school graduates are matched to residency positions in hospitals through a clearinghouse using a stable mechanism, the National Resident Matching Program (NRMP).[3] Second, seats in public schools and universities around the world are allocated through stable matching mechanisms.[4]

Third, entry-level labor markets for several types of professionals involve fixed wages and differentiated firms and workers. For example, in the United States, many of the graduates of top law schools join large law firms (Ginsburg and Wolf, 2003). The entry level position in this market is referred to as associate. An interesting feature of this market is that inside each firm the vast majority of associates are paid the same wage. Moreover, across firms, wages are mostly the same, with even the end of year bonuses being equal. Despite compensation being uniform across firms, candidates have strong preferences over firms, and pay close attention to prestige rankings in the industry.

Fourth, in markets for highly differentiated services, buyers and sellers may have (possibly heterogeneous) preferences over trade partners. Examples include entrepreneurs and venture capitalists (Sorensen, 2007), and advertisers and content providers (newspapers, content websites, search engines, or television channels).

We view the assumption of quantity competition as a stylized simplification for understanding firm competition, much like in industrial organization. Moreover, although this assumption is often strenuous, there are markets where quantity competition is a compelling benchmark. For example, in some school choice clearinghouses, school preferences are based on neighborhood priorities, lottery numbers, or scores in a single centralized exam, and as such cannot be manipulated by schools.[5]

The solution concept we adopt for most of the analysis is pure-strategy Nash equilibrium. Let $Q = \bigtimes_i Q_i$. A profile $q^* \in Q$ is an equilibrium if

$$\Pi_i(q^*) \geqslant \Pi_i(q_i, q^*_{-i}),$$

for all i and $q_i \in Q_i$. Henceforth, we assume that a pure strategy equilibrium exists.[6]

2.3. Cutoffs

This section makes a crucial observation for solving the model. Azevedo and Leshno (2013) show that stable matchings can be described in terms of cutoffs, which represent the productivity of a marginal worker at each firm. They show that stable matchings correspond exactly to the cutoffs that balance supply and demand in the market. This allows us to analyze the matching market much like a market with flexible prices and homogeneous goods, with the selectivity of each firm playing a role similar to prices.

A cutoff is simply a threshold p_i in $[0, 1]$, such that firm i accepts workers with a productivity higher than p_i, and rejects workers with lower productivity. A cutoff vector is a vector p specifying cutoffs for each firm. Given cutoffs, we may define a worker's demand as her favorite firm that would accept her. That is,

$$D^\theta(p) = \arg \max_{\{i \mid p_i \leqslant e^\theta_i\}} \succ^\theta.$$

To simplify notation, we define $D^\theta(p)$ as an I-dimensional vector, with value 1 in the coordinate corresponding to the chosen firm, and 0 in the other coordinates. We can then define aggregate demand as the I-dimensional vector

$$D(p) = \int D^\theta(p)\, d\eta(\theta).$$

We now define a market clearing cutoff.

Definition 2. A cutoff vector p is a market clearing cutoff with respect to $[\eta, q]$ if for all i

$$D_i(p) \leqslant q_i$$

with equality if $p_i > 0$.

---

[3] See Roth and Peranson (1999). One aspect in which this application does not fit the model well, is that the number of doctors matched to each program is small, and the continuum model represents a large number of doctors. In addition, the mechanism used is not exactly equal to the deferred acceptance mechanism, as it includes special provisions for couples, for example.

[4] For example, in Hungary and Turkey, college admissions are coordinated by centralized clearinghouses. Similarly, public schools in some US cities use centralized clearinghouses. See Balinski and Sönmez (1999), Biró (2007), Abdulkadiroglu et al. (2009).

[5] This point depends on schools' true preferences being equal to the preferences used by the clearinghouse, which is reasonable in some settings but not in others.

[6] As in the standard Cournot model, a pure strategy equilibrium can only be guaranteed to exist under restrictive conditions on the distribution of preferences in the population (Roberts and Sonnenschein, 1976). I follow the industrial organization literature in simply assuming existence. Note that, since Theorem 2 in Azevedo and Leshno (2013) implies that the set of stable matchings varies continuously with q, it can be shown that a mixed strategy equilibrium always exists.

The first observation we make is that there is a natural bijection between market clearing cutoffs and stable matchings. Given a stable matching μ, consider the operator $p = \mathcal{P}\mu$, where

$$p_i = \begin{cases} \inf_{\mu(i)} e_i^\theta & \text{if } \eta(\mu(i)) = q_i, \quad \text{and} \\ 0 & \text{otherwise.} \end{cases}$$

Given a market clearing cutoff p, let $\mu = \mathcal{M}p$ where

$$\mu(\theta) = D^\theta(p).$$

We then have

Lemma 1 *(Cutoff Lemma). (See Azevedo and Leshno, 2013.) Given $[\eta, q]$, there exists a unique market clearing cutoff p. Let μ be the unique stable matching. Then $p = \mathcal{P}\mu$ and $\mu = \mathcal{M}p$.*

Azevedo and Leshno (2013) give a more general statement of this lemma, for the case in which there may be multiple stable matchings. The intuition is that there is a parametrization of stable matchings by the admission thresholds in each firm.

In addition to being an admission threshold, a cutoff p_i is the productivity of a marginal hired (or rejected) worker at firm i. Therefore, we may think of cutoffs as being the marginal value of capacity for firm i, holding the pool of applicants fixed. As the shadow price of capacity, cutoffs share many properties with prices. Indeed, this analogy and basic price theory will be a central part of our argument, and we will analyze distortions caused by market power as wedges between marginal revenues and cutoffs faced by firms with nontrivial market share.

2.4. Equilibrium

Each strategy profile q induces a unique stable matching μ_q. Henceforth, we will denote by $P(q)$ the vector of market clearing cutoffs associated with μ_q. We begin by noting that when a firm raises its capacity it weakly lowers the cutoffs of all firms.

Lemma 2. *If $q' \geqslant q$, then $P(q') \leqslant P(q)$.*

This is the continuum analogue of the comparative statics results of Gale and Sotomayor (1985a, 1985b). To clarify the definitions and develop intuition, we now consider a simple example.

Example 1. There are two firms, which have a maximum capacity of 1, so that $Q_i = [0, 1]$. For simplicity, assume that wages and the cost of capacity are $w_i \equiv c_i(\cdot) \equiv 0$. There is a mass 1 of agents with preference list $\succ^\theta = 1, 2, \theta$ and a mass 1 with preferences $2, 1, \theta$. Productivity vectors e^θ are uniformly distributed in $[0, 1]^2$, and are independent of preferences. Fig. 1 depicts the relevant portion of the set of agent types. Note that, since all workers are productive, all agents should be employed in any Pareto efficient allocation. If both firms set $q_i = 1$, all workers are hired by their favorite firm, implying cutoffs $p_i = 0$.

Consider now the market clearing equations for this economy. If capacities are given by q, market clearing equations are

$$q_i = (1 + p_{-i})(1 - p_i).$$

That is, for each firm i, a measure $1 - p_i$ of the agents with preferences i, $-i$, θ are accepted. In addition, a measure $p_{-i}(1 - p_i)$ of the agents with preferences $-i$, i, θ, which were rejected by firm $-i$ are hired by firm i. Fig. 1 illustrates the allocation given cutoffs. These equations have a unique solution, which defines the unique market clearing cutoff $P(q)$ as a function of capacities. For example, assume firm 1 reduces capacity, setting $q_1 = 1/2$, while firm 2 sets $q_2 = 1$. Solving the system then yields $P_1 = (\sqrt{17} + 1)/8 \approx 0.64$, and $P_2 = (\sqrt{17} - 1)/8 \approx 0.39$. Therefore, when firm 1 reduces its capacity to $1/2$, it becomes more selective, and raises its cutoff, from 0 to 0.64. Even though firm 2 is still supplying full capacity $q_2 = 1$, its cutoff also goes up, albeit only to 0.39.

Note that, even though we computed the stable matching using the market clearing equations, Azevedo and Leshno (2013) show that it could also be computed as the outcome of the continuum version of the deferred acceptance mechanism, as is traditionally done. Because there is a unique stable matching, it does not matter whether the worker-proposing or firm-proposing version of the algorithm is used, as the outcome is the same.

The market clearing equations can also be used to calculate optimal strategies of each firm. Due to the uniform distribution, profits are given by $\Pi_i = q_i \cdot (1 + p_i)/2$. To find the optimal q, we use the implicit function theorem to calculate the marginal revenue of firm i increasing its capacity. Straightforward algebra shows that

$$MR_i = P_i - (1 - P_i)\frac{1 - P_i}{2}\left(-\frac{dP_{-i}}{dq_i}\right).$$

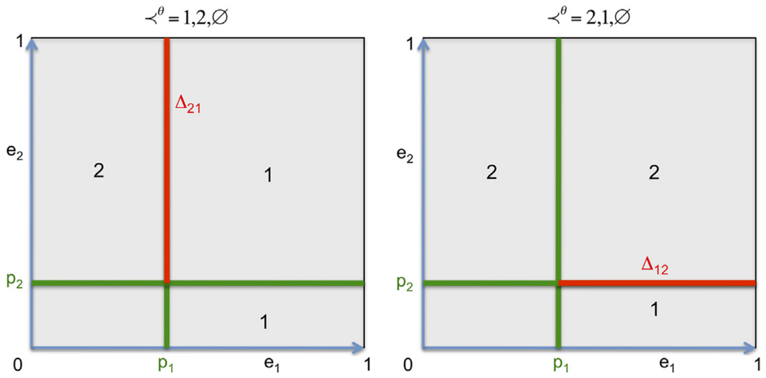

Fig. 1. The relevant portion of the set of agent types in Example 1. The left square represents agents who prefer firm 1, and the right square agents who prefer firm 2. Coordinates correspond to productivities. The numbers denote the firm to which agents in each region are matched, for a cutoff vector (p_1, p_2). The red lines represent the regions Δ_{12} and Δ_{21} of agents which are on the margin between firms 1 and 2, which are used to define the quantities M_{ij} and $\bar{P}_{ij}$ in Proposition 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In a symmetric equilibrium, we must have $MR_i(q^*, q^*) = 0$. Solving this equation we get $q^* = 4\sqrt{5} - 8 \approx 0.94$. Therefore, in equilibrium, some workers remain unemployed, even though firms cannot affect wages $w_i = 0$, and there are no costs of providing more capacity.

The puzzling feature of the example is that firms do not hire some workers with positive net productivity, even though they have enough capacity to do so. The reason why firms choose to reduce capacity in equilibrium is the possibility of rejection chains, which has been well understood since the work of Sönmez (1997). That is, by rejecting a worker, the firm sends him back to the worker pool. The rejected worker may then be hired by a competing firm, which will in turn reject another worker. Possibly, this newly rejected worker will then apply to the original firm, and be more productive than the original rejected worker. By reducing capacity, firms are shedding marginal workers, but they may gain workers who are marginal to the other firms. If preferences are not perfectly correlated, these workers may be better than the rejected workers.

The advantage of using the Azevedo and Leshno (2013) continuum framework is not to show that capacity reduction happens, which has been established in previous work. Instead, the continuum model allows us to consider the incentives facing an individual firm, and derive a first-order condition quantifying the incentives to reduce capacity. We now consider an expression for the marginal revenue, before wages and investment costs. Let

$$R_i(q) = \int_{\mu_q(i)} e_i^\theta \, d\eta(\theta) \quad \text{and} \quad MR_i(q) = \partial_{q_i} R_i(q),$$

in the case where this derivative exists.

We will say that a vector of strategies $\tilde{q}$ is interior if it is in the interior of the set $\{q \in \bigtimes_i Q_i \mid \eta(\mu_q(i)) = q_i\}$. In any interior equilibrium q^*, if profits are differentiable, then firm i's quantity choice satisfies the following first-order condition.

Proposition 1. *In any equilibrium q^* where firm i's profits $\Pi_i(q)$ are differentiable in q_i, and q^* is an interior point, we have*

$$MR_i(q^*) = w_i + c'(q_i).$$

This expression is analogous to the first-order condition for a firm in the Cournot oligopoly model (Fig. 3). The formula follows from differentiating the profit function defined in Eq. (1). The optimality condition says that the marginal productivity gain from increasing quantity must equal the marginal cost in wages and investment in capacity. If firm i does not affect the equilibrium cutoffs of its competitors, then increasing the quantity of hires by a small amount dq would add dq marginal workers of productivity P_i to its worker pool. Therefore, marginal revenue would be $MR_i = P_i$. However, when firm i's actions affect the cutoffs of the competitors, rejection chains induce a wedge between marginal revenue MR_i and cutoffs P_i. In the case where η admits a continuous density f, there is a simple intuitive expression for this wedge. Denote the set of workers who would be accepted by firm i, but are marginally accepted by a firm j that the worker prefers as

$$\Delta_{ij}(q) = \{\theta\colon e_i^\theta \geqslant p_i,\ e_j^\theta = p_j,\ j \succ^\theta i \succ^\theta \theta,\ k \succ^\theta i \Rightarrow e_k^\theta < p_k\},$$

where $p = P(q)$. Roughly speaking, these are the workers which firms i and j compete for, and which firm i may hope to poach from firm j (see Fig. 1). Let

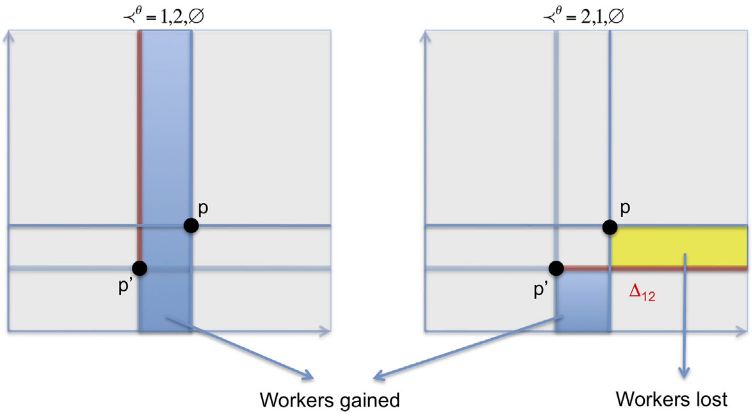

Fig. 2. The effects of a small increase in quantity by firm 1. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

$$M_{ij} = \int_{\Delta_{ij}} f(\theta)\,d\theta,$$

$$\bar{P}_{ij} = \int_{\Delta_{ij}} e_i^\theta f(\theta)\,d\theta / M_{ij} \quad \text{if } M_{ij} \neq 0$$

$$= P_i \quad \text{if } M_{ij} = 0.$$

Note that both integrals are defined on the $(I-1)$-dimensional set Δ_{ij}, not on the I-dimensional set Θ.[7] That is, M_{ij} is the $(I-1)$-dimensional mass of the set Δ_{ij} of disputed agents, and $\bar{P}_{ij}$ is their average productivity. If firm i has some market power, its quantity decisions affect the cutoff P_j. Therefore, by reducing quantity, firm i increases the cutoffs of firm j, and gains some of the agents in the disputed set Δ_{ij}. That is, by rejecting a small mass of agents dq, firm i can cause firm j to reject some of the agents in Δ_{ij}, which will then apply to firm i via a rejection chain. We have the following expression for the wedge between marginal revenue and productivity of a marginal worker.

Proposition 2. *If η has a positive continuous density f, then $P(q)$ is continuously differentiable at almost every interior point q, and*

$$MR_i(q) = P_i(q) - \sum_{j \neq i} \left[\bar{P}_{ij}(q) - P_i(q)\right] \cdot M_{ij}(q) \cdot \left(-\frac{dP_j(q)}{dq_i}\right).$$

Consequently, $MR_i(q) \leqslant P_i(q)$.

The intuition for this formula is as follows. When firm i reduces capacity by dq, it loses a measure dq of workers. If firm i had no market power, those workers would have productivity equal to the cutoff, P_i, and hence the first term. The second term measures the distortions caused by market power. If firm i has some degree of market power, its actions affect the cutoffs of other firms. By hiring more workers, firm i loses some workers in the set Δ_{ij}, which are marginal to firm j, but may be better than marginal for firm i. The difference $\bar{P}_{ij} - P_i$ measures this difference in productivity, and the term $M_{ij} \cdot (-\frac{dP_j}{dq_i})$ measures the mass of workers in this marginal set that are displaced. The intuition can be further clarified by considering the particular case of Example 1. Fig. 2 displays the effect of a small increase in quantity dq for firm 1. This increase in q_1 leads both p_1 and p_2 to decrease, so that cutoffs move from p to p'. The set of workers gained by firm 1 is highlighted as the two blue rectangles, while the set of workers that are lost is highlighted as the yellow rectangle. Notice that the workers gained have productivities close to p_1. Hence, if firm 1 has negligible market power, and its increase in quantity has a small effect in the cutoff of firm 2, the mass of workers lost would be small, and marginal revenue would be close to p_1. However, if that is not the case, firm 1 also has to take into account that it loses the mass of workers in the

[7] The integral M_{ij} is formally defined as

$$\sum_{\succ \in \mathcal{P}} \int_{((e_1,\ldots,e_{j-1},p_j,e_{j+1},\ldots,e_I),\succ) \in \Delta_{ij}} f\big((e_1,\ldots,e_{j-1},p_j,e_{j+1},\ldots,e_I),\succ\big)\,de_1\cdots de_{j-1}\,de_{j+1}\cdots de_I.$$

Other integrals over $(I-1)$-dimensional sets, such as $\bar{P}_{ij}$, are similarly defined.

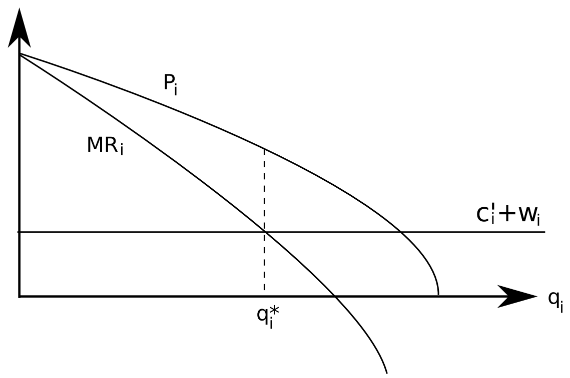

Fig. 3. Marginal revenue $MR_i(q)$, and the cutoff $P_i(q)$, as a function of q_i keeping q_{-i} fixed. The best response is denoted q_i^*.

yellow rectangle. Those workers have average quality $\bar{P}_{12}$. In addition, the total mass of workers lost must be approximately $M_{12} \cdot (-dP_2/dq_1) \cdot dq$. Therefore, the change in revenue is approximately $\{p_1 - (\bar{P}_{12} - p_1) \cdot M_{12} \cdot (-dP_2/dq_1)\}dq$.

The wedge between the productivity of a marginal worker and marginal revenue is analogous to the wedge between prices and marginal revenues in the Cournot model. Fig. 3 plots marginal revenue and cutoffs, as a function of the quantity chosen by firm i. Note that the marginal revenue curve MR_i is lower than the cutoff curve P_i. Moreover, when $q_i = 0$, we have that $\bar{P}_{ij} = P_i = 1$, so that $MR_i = P_i$. Note that the first-order condition for firm i is to provide quantity up to the point where the marginal revenue curve crosses the marginal cost plus wages curve, $w_i + c_i'$. Therefore, the equilibrium quantity is given by the point $q^* < 1$. At this point, we have $P_i > MR_i = w_i + c_i'$. Therefore, in equilibrium the firms do not hire some workers with strictly positive net productivity.

Equilibrium does not rely on firms reasoning through the rejection chains they set off. All that is necessary is that firms set their quantities optimally given the strategies of other firms. In several markets, firms (which may represent colleges, hospitals, or television networks) do seem to spend a lot of time deciding the quantities to be supplied. A college, for example, faces a quantity versus quality tradeoff when deciding on the size of each year's entering class. The equilibrium assumption is that the college gets this decision right, by any mix of trial and error, experience, or abstraction. However, it does not depend on each college fully understanding its impact on the rest of the market.

Notice that, if all n firms were acting in unison, as a monopolist maximizing profits, they would have even greater incentives for reducing capacity. Under suitable differentiability assumptions, at an interior point q we may write the first-order condition with respect to q_i as

$$w_i + c'(q_i) = MR_i(q) + \sum_{j \neq i} \frac{d\Pi_j}{dq_i} \leqslant MR_i(q),$$

where we used the fact that all $d\Pi_j/dq_i \leqslant 0$.[8] Therefore, a cartel has more incentives for quantity reduction than oligopolistic firms. For example, if 2 out of N firms in a market merge, they will have incentives to reduce their capacities *vis a vis* their pre-merger choices.

An immediate consequence of the previous discussion is that quantity distortions in markets with fixed wages are driven by preference heterogeneity. We say that workers have homogeneous ordinal preferences if all workers in the support of η have the same preference ordering over firms. We say that firms have homogeneous ordinal preferences if, for any two worker types $\theta \neq \theta'$ in the support of η, we have that either $e_i^\theta > e_i^{\theta'}$ for all i or $e_i^\theta < e_i^{\theta'}$ for all i. We record this as the following proposition:

Proposition 3. *If either workers or firms have homogeneous ordinal preferences, then, at any interior q, the marginal revenue of each firm equals the cutoff, $MR_i(q) = P_i(q)$. That is, firms have no incentives to reduce capacity.*

The reason why firm i can profit from rejection chains is that it may reject a worker θ who is accepted by firm j, and that leads firm j to reject a better worker θ'. However, if firms i and j have the same preferences, this is not possible. The same argument holds when all workers have the same preferences.

[8] This follows from the monotonicity result in Lemma 2.

3. Personalized wages

3.1. Firms, workers, and stable matchings with contracts

So far we have assumed that, within each firm, all workers are paid the same wage. However, this is not a realistic assumption in some markets. In this section, we consider quantity manipulation games when wages are personalized. We will use the model of matching with contracts introduced by Azevedo and Leshno (2013). This framework is similar to models of matching with contracts introduced by Kelso and Crawford (1982) and Hatfield and Milgrom (2005). We focus on the case where preferences are quasilinear, and contracts only specify wages.

A worker's type θ must now specify a utility u_i^θ of being matched to firm i, and profits e_i^θ that this match generates for the firm. The set of types is $\Theta_X \subseteq \Re^{2I}$. Denote the surplus of a matching by $s_i^\theta = u_i^\theta + e_i^\theta$. We assume that Θ_X is the set of all types with $s_i^\theta \in [0, 1]$. Let η be the distribution of worker types, which can be represented by a positive continuous density. The set of contracts is denoted by

$$X = (I \times \Theta \times \Re) \cup \{\emptyset\}.$$

A nonempty contract $x = (i, \theta, w)$ specifies a worker, a firm, and a wage.

A matching, $\mu : \Theta \cup I \to X \times 2^X$ assigns each worker to either the empty contract or to a contract that contains her, and each firm to a set of contracts that contain it (possibly the empty set). Much like in the version of the model without transfers, it turns out that stable matchings can be decentralized using cutoffs. Let p be a vector of cutoffs, representing the shadow price of capacity for each firm. Let the individual demand function assign each worker to the firm where he generates the greatest surplus net of the shadow cost of capacity:

$$D^\theta(p) = \arg\max_{i \in I \cup \{\emptyset\}} s_i^\theta - p_i,$$

where we denote $u_\emptyset^\theta = p_\emptyset = 0$. If two firms are tied (which only happens for a measure 0 set of workers), we let the worker be matched to the lowest numbered firm. Define aggregate demand for a firm as the mass of workers demanding the firm at cutoffs p

$$D_i(p) = \eta\big(\theta\colon\ D^\theta(p) = i\big).$$

Given a vector of capacities q, define a market clearing cutoff as in Section 2.3, mutatis mutandis. Azevedo and Leshno (2013) show that a unique market clearing cutoff $P(q)$ exists. Moreover, they define stable matchings, similarly to what is done in Section 2.1, and show that in any stable matching agent θ is matched to firm $D^\theta(P(q))$. However, even though market clearing cutoffs are unique, and all agents are matched to the same firms in any stable matching, wages are not uniquely determined. Intuitively, a firm may pay an agent a wage that is anything between the firm's willingness to pay, and the willingness to pay of the firm making the second best offer. That is, if j is the firm making the next to best offer, the wage may be any value such that

$$s_j^\theta - p_j \leqslant u_i^\theta + \text{wage} \leqslant s_i^\theta - p_i.$$

For concreteness, we will focus on the firm-optimal stable matching, where firms have all the bargaining power, and pay workers just enough for them not to switch to another firm. For simplicity, we define the firm-optimal stable matching using cutoffs. For the interested reader, Azevedo and Leshno (2013) derive this firm-optimal stable matching from the standard definition of stability, and the working paper version of the present article contains a fuller discussion.

Definition 3. Given $[\eta_X, q]$, the firm-optimal stable matching with contracts is defined as follows. Let $p = P(q)$. Worker θ is matched to firm $D^\theta(p)$. If $j \in I \cup \{\emptyset\}$ is the firm with the second highest value of $s_j^\theta - p_j$, the worker's wage is set to leave him at his reservation utility

$$\bar{u}^\theta = w + u_i^\theta = s_j^\theta - p_j.$$

3.2. The game

We define the oligopoly game with flexible wages as follows. The primitives are X, Θ_X, η_X, $c(\cdot)$, Q and the set of players I.

1. Firms simultaneously choose quantities q_i in the compact intervals Q_i.
2. After capacity choices q, workers are hired according to the firm-optimal stable matching with contracts[9] with respect to $[\eta_X, q]$. Let p denote the unique vector of market clearing cutoffs, and μ the firm-optimal stable matching.

[9] Note that, since a worker's wage depends on her outside option, in principle workers would have incentives to misreport their preferences, to inflate this value. However, we are assuming complete information regarding preferences, and firms simply manipulate their capacity investments.

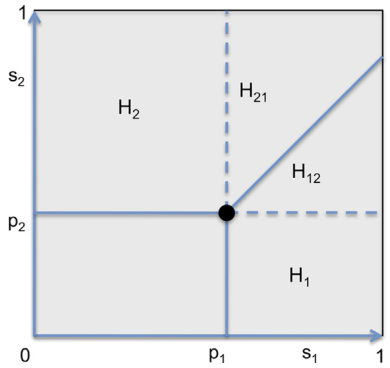

Fig. 4. A stable matching, plotted over the set of possible surplus vectors $s^\theta = (s_1^\theta, s_2^\theta) = u^\theta + e^\theta$.

3. Each firm's payoff is given by

$$\Pi_i = \int\limits_{\mu(i)} s_i^\theta - \bar{u}^\theta \, d\eta(\theta) - c_i(q_i),$$

where the continuously differentiable function $c_i(\cdot)$ is the cost of investing in capacity. That is, profits are the integral of the surplus of all contracts signed, minus the utility left to workers, minus the cost of capacity.

The game corresponds to a situation where firms first invest in capacities. The matching and wages are then determined by the firm-optimal stable allocation given those capacities, analogously to the Cournot model.

3.3. Equilibrium

Let $P(q)$ be the unique vector of market clearing cutoffs with respect to $[\eta_X, q]$. We will first consider a simple example of the model.

Example 2. There are two firms, with no cost of investing in capacity, $c_i \equiv 0$, and $Q_i = [0, 1]$. There is a mass 2 of workers, with surplus vectors $(s_1^\theta, s_2^\theta) = s^\theta = e^\theta + u^\theta$ uniformly distributed in $[0, 1]^2$. Fig. 4 illustrates a typical stable matching, with cutoffs (p_1, p_2). Workers are always assigned to the firm where $s_i^\theta - p_i$ is the highest, provided it is positive. Therefore, all workers with surplus vectors in region $H_1 \cup H_{12}$ are assigned to firm 1, and the workers with surplus vector in region $H_2 \cup H_{21}$ are assigned to firm 2 (Fig. 4). To illustrate how cutoffs are determined, consider the case where firm 1 sets $q_1 = 1/2$, while firm 2 supplies maximum capacity $q_2 = 1$. Since the mass of unemployed workers is $1/2$, we must have

$$2 \cdot p_1 p_2 = 1/2.$$

Moreover, the market clearing equation for firm 1 yields

$$q_1 = 1/2 = 2(1 - p_1)p_2 + 2(1 - p_1)^2/2.$$

Solving these equations yields $p_1 \approx 0.60$ and $p_2 \approx 0.42$. Note that not all workers are assigned to the firm where surplus s_i^θ is the highest. Since firms have limited capacity, workers are instead assigned to where $s_i^\theta - p_i$ is the highest. In the example, because firm 1 has smaller capacity, it has a higher cutoff, and therefore some workers with higher surplus at firm 1 ($s_1^\theta > s_2^\theta$) are assigned to firm 2.

Consider now a symmetric pure strategy equilibrium, where $q_1 = q_2 = q^*$. By market clearing, cutoffs are given by $q^* = 1 - p^{*2}$, where $p^* = P_i(q^*, q^*)$. Some algebra shows that the derivative of Π_i with respect to q_i is given by

$$MR_i\big(q^*, q^*\big) = p^* - \big(1 - p^*\big)^2 \cdot \left(-\frac{dP_j}{dq_i}\right).$$

Solving the first-order condition then yields that in equilibrium $p^* \approx 0.36$ and $q^* \approx 0.87$.

The example shows that, unsurprisingly, firms have incentives to reduce quantity when wages are personalized. The interesting point is that, although firms may reduce quantity with either fixed or flexible wages, the reasons to do so are very different. When wages are uniform, firms choose to reduce quantity when their preferences are heterogeneous, so that firing a worker may set off a profitable rejection chain. With personalized wages, we will see that firms choose to reduce quantity exactly when their preferences are aligned. Given quantities q, let the set of agents that are hired by firm i and that no other firm wants to hire as

$$H_i(q) = \{\theta \mid s_i^\theta \geqslant p_i,\ s_j^\theta < p_j \text{ for all } j \neq i\}.$$

Moreover, define as $H_{ij}(q)$ the set of agents who are hired by firm i, and whose second best offer is from firm j as

$$H_{ij}(q) = \{\theta \mid s_i^\theta - p_i \geqslant s_j^\theta - p_j \geqslant s_k^\theta - p_k \text{ for all } k \neq i, j \text{ and } s_j^\theta - p_j \geqslant 0\}.$$

Consider now the expression for the marginal revenue of firm i. Denote by the revenue of firm i its profits before deducting investment costs. That is, $R_i = \Pi_i(q) + c_i(q_i)$. We may write revenue as

$$R_i(q) = \int_{H_i(q)} s_i^\theta \, d\eta(\theta) + \sum_{j \neq i} \int_{H_{ij}(q)} s_i^\theta - s_j^\theta + P_j(q) \, d\eta(\theta). \tag{2}$$

The first term is the revenue from hiring workers for whom firm i is the sole bidder. Since firm i captures all the surplus of the relationship, revenues in this region are simply the integral of the surplus of each employment relationship. The other terms are the sum of the profits from hiring workers for whom firm i competes with firm j. In these regions, revenue is the integral of surplus, but net of the cost of outbidding firm j, which must be $s_j^\theta - P_j(q)$ (that is, the surplus firm j could obtain, minus firm j's shadow price of capacity).

Let MR_i be the derivative of revenue with respect to q_i. The following proposition gives an expression for the marginal revenue of firm i.

Proposition 4. *$P(q)$ is continuously differentiable at almost every interior point q, and*

$$MR_i(q) = P_i(q) - \sum_{j \neq i} \eta(H_{ij}(q)) \cdot \left(-\frac{dP_j(q)}{dq_i}\right).$$

Consequently, $MR_i(q) \leqslant P_i(q)$.

The intuition behind the proposition is as follows. If firm i raises its quantity by a small amount dq, it hires some additional workers. These come from small changes in the sets H_i, H_{ij} of hires. Since all these new workers were on the margin for being hired, the average profit from hiring them must be P_i. This equals a gain of $P_i \, dq$ in productivity. On the other hand, by raising its quantity, firm i decreases the cutoffs P_j of other firms. This means that firm j will bid more aggressively for workers. Therefore firm i must pay an extra $-\frac{dP_j}{dq_i} \cdot dq$ for the mass $\eta(H_{ij})$ of workers in the set H_{ij}.

As in the case of fixed wages, the marginal revenue curve MR_i is lower than the cutoff curve P_i. Therefore, in equilibrium, the profitability of a marginal hired worker is higher than the cost of investing in capacity (as in Fig. 3, replacing the $c_i' + w_i$ curve for c_i'). Consequently, firms with market power have incentives to reduce their capacities. The reason is that a firm i with nontrivial market share may affect the cutoffs of other firms, and therefore how much other firms are willing to bid for workers. Note that the rejection chain intuition from the model with fixed wages no longer applies, as capacity reduction is now driven by the effect of capacity on other firms' bids.

Note that firm i has more incentives to reduce capacity the larger the mass of contested workers on the sets H_{ij}. To see this formally, hold fixed preferences, and a choice of quantities for each firm, which determine cutoffs $P(q)$. Consider now a change in preferences, that moves some mass of workers from the sets $H_i(q)$ to the sets $H_{ij}(q)$ (Fig. 4). This change does not affect $P(q)$, as each firm still has the same number of workers matched to it. However, by Proposition 4, it increases the incentives of each firm to reduce capacity. Since this is a qualitatively important observation, we note it as a corollary.

Corollary 1. *In the model with personalized wages, at any quantity where $P(q)$ is continuously differentiable, holding fixed the inverse demand function $P(\cdot)$, the wedge between the marginal revenue of firm i and its cutoff is increasing in the mass of workers in the region H_{ij}, for any $j \neq i$.*

That is, the greater the number of workers that both firms are interested in, the greater the incentives to reduce capacity. This is the opposite of the conclusion from the analysis of markets with fixed wages. With fixed wages, firms further distort their capacity choices the more they disagree about who the best workers are. With flexible wages, distortions are greater the more firms' preferences agree.

4. Comparison between uniform and personalized wages

While in some markets workers are paid uniform wages in each firm, in others wages are personalized. For example, most top American law firms pay all graduating law students joining the firm as an associate the same wage (Ginsburg and Wolf, 2003). In contrast, senior lawyers receive personalized wages. When a firm hires a partner from a rival firm, what practitioners call a lateral transaction, the offer is often personalized, and lawyers of comparable seniority receive very different wages. Moreover, in markets organized around a centralized clearinghouse, a key design variable is whether the matching mechanism should use uniform wages, or allow for flexible wages, as in the mechanism proposed by Crawford (2008).

Previous work has considered which form of pricing is more efficient. In most models in this literature, personalized prices always yield higher efficiency.[10] In contrast, the quantity competition model implies that personalized prices have benefits (through higher matching efficiency) and costs (by possibly exacerbating quantity distortions). We have the following result.

Proposition 5. *For a given vector of capacities q, matching with personalized wages is always more efficient. However, if firms are allowed to set quantities, uniform wages may dominate personalized wages, as the latter may induce more capacity manipulation.*

To see how uniform wages may yield higher efficiency than personalized wages, we consider a simple example derived from Bulow and Levin (2006). We will show that, by changing the productivity dispersion across firms, we can make either uniform or personalized wages more efficient. There are two firms $i = 1, 2$, and a mass 1 of workers. Worker types are indexed by a productivity parameter θ uniformly distributed in $[0, 1]$. At firm 1, a worker of type θ generates $e_1^\theta = \theta$. However, firm 2 is more productive, and multiplies the output of a worker $e_2^\theta = A\theta = (1 + \epsilon)\theta$. Workers are assumed to have utility 0 from matching with either firm, and in case firms offer equal wages, workers choose randomly between them. Capacity at each firm is constrained to be in the interval $Q_i = [0, 1]$, and there are no costs of investing in capacity, $c_i \equiv 0$. Appendix A contains details on the computations below.

Consider first the case of uniform wages. For simplicity, assume both firms set $w_1 = w_2 = 0$. Since firm preferences are vertical, neither firm has incentives to reduce capacity. Therefore, both set $q_i = 1$. Even though there is no reduction in capacity, the matching is quite inefficient, as workers of all productivities are matched to either firm with probability $1/2$. Consequently, total welfare equals $1 + \epsilon/2$.

We now consider personalized wages. Given quantities, matching is efficient. However, aligned preferences and flexible wages give both firms strong incentives to reduce capacity. Applying our formula for marginal revenue, we obtain $MR_1 = p_1 - q_1/A$, while $MR_2 = p_2 - q_2$. Equilibrium quantities are therefore highly depressed, as depicted in Fig. 3. We have

$$q_1^* = \frac{1+\epsilon}{3+3\epsilon+\epsilon^2}, \qquad q_2^* = \frac{1+2\epsilon+\epsilon^2}{3+3\epsilon+\epsilon^2}. \tag{3}$$

Note that this outcome is particularly inefficient if the productivity differences are small. If $\epsilon \approx 0$, we have $q_i \approx 1/3$, so that $q_1 + q_2 \approx 2/3$. That is, despite the small productivity differences between the firms, quantities are very depressed. This makes personalized wages much more inefficient than uniform wages, as the gain in matching efficiency is negligible, while there is considerable loss due to capacity reduction.

Calculating surplus, we find that uniform wages are more efficient for a broad range of parameters, as long as $\epsilon \leqslant 1.21$. Personalized wages are more efficient for $\epsilon > 1.21$. This means that firm 2 has to be more than twice as productive as firm 1 for personalized wages to be more efficient. The intuition is that, when both firms have similar productivities, the cost of allocative inefficiency is small. Therefore, the quantity distortion dominates the allocative inefficiency. However, when one firm is much more productive than the other, the allocative inefficiency of uniform wages dominates, and personalized wages yield higher efficiency. Indeed, by Eq. (3), as the productivity difference ϵ grows, the most productive firm grows. As ϵ approaches infinity, q_1 converges to 0 and q_2 converges to 1. The more productive firm takes over the entire market, and the loss from capacity reduction becomes negligible.

Finally, note that the results do not depend on wages being set exogenously in the uniform case. To see this, consider the game where firms choose both a quantity q_i and a uniform wage w_i. Assume, for simplicity, that $Q_i = [0, 1/2]$ and $\epsilon < (\sqrt{5} - 1)/2$. In this case, equilibrium coincides with the solution in Bulow and Levin (2006). Firms offer random wages, since a deterministic wage offer could be undercut by a rival. However, firm 2, which is the most efficient firm, pays on average higher wages, and $\Pr(w_2 > w_1) = (1 + 2\epsilon)/2(1 + \epsilon) > 1/2$. Firms always supply the maximum quantity $q_i = 1/2$. This mixed equilibrium is consistent with Proposition 3, as the incentives to reduce capacity are small when wages are uniform and preferences are homogeneous. Therefore, endogenous uniform wages imply allocative inefficiency with positive probability, but no quantity reduction, as in the case of exogenous wages. Calculating welfare, one finds that uniform wages are more efficient for $\epsilon \leqslant 0.41$, and personalized wages are more efficient for $\epsilon > 0.41$.

[10] For example, Mailath et al. (2013) have proposed a continuum model of one-to-one assortative matching in which they compare flexible and uniform wages. In their model, due to the simple preferences, and since each firm hires a single worker, matching is always efficient. However, they show that, when prices are not personalized, workers may not have the right ex-ante incentives to invest in skills, as the market does not fully compensate them for it, which causes uniform wages to be less efficient. More closely related to our model, Bulow and Levin (2006) propose a discrete model of one-to-one assortative matching where firms must make wage offers simultaneously. Because firms must make offers simultaneously, if wages are uniform they employ mixed strategies. Therefore, the uniform wages generate a small degree of inefficiency *vis a vis* personalized wages. The model also implies that wages are lower, and more compressed than in the flexible case.

5. Conclusion

This paper considers an equilibrium model of imperfect competition in many-to-one matching markets. This is a first step towards understanding firm behavior, and its implications for the design and regulation of matching markets.

The main contribution is to extend standard price-theoretic insights of the Cournot model to matching markets, and understand to what extent these insights have to be modified. Market power induces a wedge between the marginal revenue of a firm, and the net productivity of a marginal hired worker. Interestingly, the determinants of the size of this wedge are very different when wages are uniform or personalized. With uniform wages, the wedge exists due to heterogeneous preferences between firms, which means that rejecting a worker may create a beneficial rejection chain. In contrast, when wages are personalized, the wedge exists because of aligned preferences. Firms that reduce capacity increase the pool of available workers, which induces competitors to bid less aggressively for workers that both firms covet.

As an application, these insights contribute to the debate over the desirability of uniform versus flexible wages. We have seen that taking strategic capacity setting into account qualifies the Bulow and Levin (2006) conclusion that flexible wages always generate more efficiency. In the Bulow and Levin model, if firms are allowed to choose capacity, flexible wages do produce higher matching efficiency given quantities, but they also give more incentives for firms to reduce capacities. Flexible wages are still more efficient if firms are sufficiently heterogeneous. However, if firms are very similar, so that the loss from matching inefficiency is small, uniform wages produce higher welfare, as they induce less capacity reduction. We note that, although this conclusion is intuitive, it relies on our assumption of quantity competition. It would be interesting to investigate to which extent it holds under other types of firm competition.

An important limitation of the analysis is that we only consider quantity competition. While in industrial organization quantity competition models figure prominently, they are by no means the only models available. As such, the assumption of quantity competition should be seen as a first step towards understanding strategic behavior in matching markets. It would be interesting to explore matching markets in which firms have a different set of strategic variables at their disposal. For example, models where firms can differentiate themselves, tailor products to specific market segments, or misrepresent their preferences.

An interesting application of analyzing behavior under different strategic variables is to compare different market design choices for centralized clearinghouses. For example, in the model of quantity competition we have shown that firms have incentives to reduce capacity. It can be shown that, if firms were to choose cutoffs, and quantities were assigned by market clearing, then in equilibrium firms would not have incentives to shade. This suggests that, in a market with perfect information and no aggregate randomness, in which firms cannot misreport their preference rankings, a clearinghouse in which firms choose cutoffs would perform better than one in which they report capacities. Naturally, this conclusion is subject to the caveat that, if there is some aggregate randomness in the market, having firms report cutoffs subjects them to the risk of getting too many or too few workers. Therefore, the comparison between these two rules will depend on the cost of firms matching to a number of workers that is different than ideal, on the level of aggregate randomness, and on whether firms can misrepresent preference orderings. Nevertheless, comparing different matching mechanisms by their equilibrium properties is an interesting direction for research, and is complementary to the axiomatic approach that most of the literature takes.

Appendix A. Proofs

A.1. Uniform wages

A.1.1. Stable matchings

First, note that by Azevedo and Leshno's (2013) Theorem 1, a unique stable matching exists. This, coupled with their Cutoff Lemma implies the version of the Cutoff Lemma stated in the text. Moreover, Azevedo and Leshno's (2013) Theorem 2 guarantees that market clearing cutoffs $P(q)$ vary continuously with q.

We now provide a proof of the monotonicity result in Lemma 2.

Proof of Lemma 2. Let $p = P(q)$, $p' = P(q')$, and $\hat{p}$ be the pointwise max of the vectors p and p'. For every i it is always the case that either $\hat{p}_i = p_i$ or p'_i. If $\hat{p}_i = p_i$, then because for all $j \neq i$ we have $\hat{p}_j \geqslant p_j$ we must have $D_i(\hat{p}) \geqslant D_i(p)$. If $\hat{p}_i = p'_i > p_i$ then by the same logic $D_i(\hat{p}) \geqslant D_i(p') \geqslant D_i(p)$. The second inequality follows because $p'_i > 0$, so that we have $D_i(p') = q'_i$. Moreover from the definition we have $q_i \geqslant D_i(p)$, hence we get $D_i(p') \geqslant D_i(p)$. Therefore $\hat{p} \geqslant p$ and $D(\hat{p}) \geqslant D(p)$. Since $\sum_i D_i(\hat{p}) \leqslant \sum_i D_i(p)$, we must have $D(\hat{p}) = D(p)$. If $p_i > 0$, then $D_i(\hat{p}) = D_i(p) = q_i$. If $p_i = 0$ and $\hat{p}_i = 0$ then $D_i(\hat{p}_i) \leqslant q_i$. Finally, if $p_i = 0$ and $\hat{p}_i > 0$ then $D_i(\hat{p}_i) \geqslant D_i(p'_i) = q'_i \geqslant q_i$, so that $D_i(\hat{p}_i) = q_i$. Therefore, $\hat{p}$ is a market clearing cutoff given q. By uniqueness of market clearing cutoffs, $\hat{p} = p$, and therefore $p' \leqslant p$. □

A.1.2. The oligopoly game with exogenous wages

We may now prove Proposition 2, which guarantees that the profit functions are continuously differentiable almost everywhere, and provides an expression for marginal revenues.

Proof of Proposition 2. First, note that since η admits a continuous density, the demand function $D(p)$ may be written as

$$D(p) = \int_{(\mathcal{M}p)(i)} f(\theta)\, d\theta,$$

where

$$\mathcal{M}p(i) = \{\theta \in \Theta \mid e_i^\theta \geqslant p_i,\ e_j^\theta < p_j \text{ for all } j \text{ such that } j \succ^\theta i,\ i \succ^\theta \theta\}. \tag{4}$$

Therefore, by Leibniz's rule for differentiation under the integral sign, $D(p)$ is continuously differentiable.

Let Q^* be the set of interior points of Q. That is, the interior of the set of points q such that all $\eta(\mu_q(i)) = q_i$. Note that, in Q^*, market clearing cutoffs $P(q)$ are the single root of the equation $D(P(q)) = q$. By Sard's Theorem,[11] for almost every point $q \in Q^*$, $D(\cdot)$ is continuously differentiable at $P(q)$, and its derivative is nonsingular. Therefore, by the inverse function theorem, $P(q)$ is continuously differentiable in a neighborhood of q.

Given a quantity vector q, and cutoffs $P(q)$, the revenue of firm i may be written as

$$R_i(q) = \int_{(\mathcal{M}P(q))(i)} e_i^\theta \cdot f(\theta)\, d\theta. \tag{5}$$

If P is continuously differentiable at q, the formula for the marginal revenue in the proposition follows directly from an application of Leibniz's rule. □

We now prove Proposition 3, which guarantees that when either side of the market has homogeneous cardinal preferences, then firms have no incentives to reduce capacity.

Proof of Proposition 3. *Part* 1: *Homogeneous ordinal worker preferences.*

Without loss of generality, assume that all workers have preference ordering $1, 2, \ldots, I$.[12] Note that, by the market clearing equations, cutoffs $P_i(q)$ do not depend on q_j for $j > i$. In addition, by the formula for the set of matched students $\mathcal{M}p(i)$ in Eq. (4) in the proof of Proposition 2, we have that revenue $R_i(q)$ may be written as a function

$$\tilde{R}_i\big(P_1(q), P_2(q), \ldots, P_{i-1}(q), P_i(q)\big),$$

which does not depend on $P_k(q)$ for $k > i$. If we consider an interior point q and q' with $q'_i = q_i + \epsilon$, $\epsilon > 0$ and $q'_j = q_j$ for all other coordinates, we have

$$\begin{aligned} R_i(q') &= \tilde{R}_i\big(P_1(q'), P_2(q'), \ldots, P_{i-1}(q'), P_i(q)\big) \\ &= \tilde{R}_i\big(P_1(q), P_2(q), \ldots, P_{i-1}(q), P_i(q')\big). \end{aligned}$$

Therefore, using again Eq. (4), $R_i(q') - R_i(q)$ may be written as

$$\int_A e_i^\theta\, d\eta(\theta),$$

where

$$A = \{e_j^\theta < P_j(q) \text{ for all } j < i,\ P_i(q') \leqslant e_i^\theta < P_i(q)\}.$$

Because $P(\cdot)$ is continuous, the productivity e_i^θ of all workers in the set A is approximately $P_i(q)$. Moreover, as the measure of A is ϵ, we have that

$$R_i(q') - R_i(q) = P_i(q) \cdot \epsilon + o(\epsilon).$$

Therefore, $R_i(\cdot)$ is differentiable at q with derivative $P_i(q)$, completing the proof.

Part 2: *Homogeneous ordinal firm preferences.*

First, note that there must exist increasing continuous functions $f_2, f_3, \ldots, f_I$ such that the support of η equals the set

$$\{(e_1^\theta, f_2(e_1^\theta), \ldots, f_I(e_1^\theta)) \mid e_1^\theta \in [0, 1]\}.$$

[11] See Milnor (1997).

[12] We are assuming that all firms are considered acceptable by the workers, as firms which no worker finds acceptable play no role in the proof.

To see this, note first that the support must include points with all possible values of $e_1^\theta \in [0,1]$, due to the assumption on the support of η made in the main text. Second, for a given value of e_1^θ, by the homogeneous ordinal preferences assumption, the support may only contain one point. We denote this point as $(e_1^\theta, f_2(e_1^\theta), \ldots, f_I(e_1^\theta))$, which defines the functions f_i. Again, by the homogeneous ordinal preferences assumption, the f_i are strictly increasing. Moreover, they must be continuous, as otherwise the support of θ would not include points with some value of $e_i^\theta \in [0,1]$, which would violate the assumption made on the support of η.

With this observation in hand, the rest of the proof is simple, and similar to the first part. Let f_1^θ be the identity map. Note that, for all i, $e_i^\theta \leqslant P_i$ iff $e_1^\theta \leqslant f_i^{-1}(P_i)$. Therefore, given an interior point q, we may denote the firms which are more selective than firm 1 as

$$I_+(q) = \{i\colon\ f_i^{-1}(P_i(q)) > P_1(q)\}.$$

Note that, by the market clearing equations, a small change in q_1 does not affect $P_i(q)$ for $i \in I_+$. Now take an interior point q and q' with $q'_1 = q_1 + \epsilon$, $\epsilon > 0$ and $q'_j = q_j$ for all coordinates $j \neq 1$. By the definition of the demand function, all of the elements θ in the symmetric difference $\mu_q(1) \triangle \mu_{q'}(1)$ must satisfy

$$P_i(q') \leqslant e_i^\theta < P_i(q)$$

for some $i \in I$. For small ϵ, we have $P_i(q') = P_i(q)$ for $i \in I_+(q)$, and therefore the equation has to hold for some $i \in I \backslash I_+(q)$. Therefore we must have $e_1^\theta \leqslant P_1(q)$. Since, by the definition of the demand function, every point in the symmetric difference must also satisfy $P_1(q') \leqslant e_1^\theta$, we must have

$$P_1(q') \leqslant e_1^\theta < P_1(q).$$

Therefore, using the same argument as in Part 1 on the continuity of P, we must have that the revenue of firm 1 is differentiable at q, and $MR_1(q) = P_1(q)$. □

A.1.3. Matching with contracts

The matching framework used in Section 3 is a particular case of the model in Appendix D of Azevedo and Leshno (2013). It follows from their results that our definition of the firm-optimal stable matching coincides with a definition based on the notion of stability, and also that this firm-optimal stable matching is unique and varies continuously with q. We now prove Proposition 4, which characterizes marginal revenue.

Proof of Proposition 4. The proof that $P(q)$ is differentiable for almost every interior point q is exactly the same as in the case with exogenous wages given in the proof of Proposition 2. We now take an interior point q where $P(\cdot)$ is differentiable, and derive the formula for the marginal revenue. The formula for marginal revenue then follows directly from the formula for the revenue $R_i(q)$ of firm i in Eq. (2), and a direct application of Leibniz's formula for differentiation under the integral sign. □

A.2. Applications

A.2.1. Comparison between uniform and personalized wages

Proof of Proposition 5. The proof that, with endogenous capacities, uniform wages may dominate follows from the example given in the text. Therefore it only remains to prove that, for a fixed capacity vector q, personalized wages generate at least as much welfare as matching with uniform wages. We will demonstrate this by showing a stronger efficiency result, that matching with personalized wages is efficient in an even broader class of allocations.

Consider a set of workers Θ_X satisfying the requirements of both the models with uniform wages and with flexible wages. Let a generalized allocation be a measurable map

$$x : \Theta_X \to [0,1]^{I+1}$$

designating a distribution $x(\theta)$ of each worker type over firms, with $I+1$ representing being unemployed. Therefore both a stable matching with uniform wages and firm-optimal stable matching with contracts induce an allocation, and one that only takes values in the extreme points of the simplex. Given a generalized allocation, we define social welfare as

$$\int_{\Theta_X} s^\theta \cdot x(\theta)\, d\eta(\theta).$$

Consider now the problem of finding a generalized allocation that maximizes social welfare subject to feasibility constraints

$$\max \int_{\Theta_X} s^\theta \cdot x(\theta)\, d\eta(\theta)$$
$$\text{s.t.} \int_{\Theta_X} x_i(\theta)\, d\eta(\theta) \leqslant q_i \quad \text{for } i = 1, \ldots, I.$$

A standard compactness argument implies that such a maximum value is attained by at least one generalized allocation x^*. Moreover, because the problem has allocations x where all constraints are strictly at slack, $\int_{\Theta_X} x_i(\theta)\, d\eta(\theta) < q_i$ for all i, strong Lagrange duality holds. By Theorem 1, p. 217 from Luenberger (1969), there exist numbers $\lambda_i \geqslant 0$ such that x^* maximizes

$$\int_{\Theta_X} s^\theta \cdot x(\theta)\, d\eta(\theta) + \sum_{i=1}^{I} \lambda_i \cdot \left[q_i - \int_{\Theta_X} x_i(\theta)\, d\eta(\theta) \right]$$

over all generalized allocations. Moreover, if $\lambda_i > 0$, then $\int_{\Theta_X} x_i(\theta)\, d\eta(\theta) = q_i$. Note that we can rewrite the expression above as

$$\sum_{i=1}^{I} \int_{\Theta_X} \left(s_i^\theta - \lambda_i\right) \cdot x_i(\theta)\, d\eta(\theta).$$

Therefore, any maximizer x^* satisfies that almost every type θ is matched with probability 1 to a firm that maximizes $s_i^\theta - \lambda_i$. So the measure of workers x^* allocated to each firm equals the demand for each firm when market clearing cutoffs are equal to λ, in the matching with contracts model. Therefore λ is a vector of market clearing cutoffs in the matching with contracts model. Since the market clearing cutoffs are unique, we have $\lambda = P(q)$. Therefore, the generalized allocation x^* coincides almost everywhere with the generalized allocation induced by the firm-optimal stable matching with contracts. Consequently, any stable matching with personalized wages maximizes social welfare. In particular, any stable matching with uniform wages generates weakly lower welfare, completing the proof. □

Details of the calculation in the example comparing uniform and personalized wages

For the case of uniform exogenous wages, the calculations are trivial. We now solve the case of uniform endogenous wages. The reasoning follows Bulow and Levin (2006). Assume for now that both firms set $q_i = 1/2$. The firm that offers higher wages attracts workers with an average quality of 3/4, while the firm with lower wages attracts workers with an average quality of 1/4. Following their algorithm to characterize equilibrium, firms must offer random wages, with distributions $G_i(\cdot)$ with the same interval as support. To find this support, we consider each firm's first-order condition with respect to wage in an interior point of the interval, as each firm must be indifferent between offering any wage in the support. We have

$$g_1(w) \cdot A \cdot 1/2 = 1,$$
$$g_2(w) \cdot 1/2 = 1.$$

Therefore the density of firm 2's offer is $g_2(w) = 2$. As argued in Bulow and Levin (2006, p. 659), the lowest wage offered must be zero. Therefore the support of the distributions is $[0, 1/2]$. Firm 2 offers a wage uniformly in this interval. Firm 1 has a density of only $2/A$. According to the Bulow and Levin (2006) algorithm, with probability $1/A$ it offers a wage uniformly at random in the interval $[0, 1/2]$, and offers 0 otherwise. Consequently, the probability that firm 1 offers a higher wage is $1/2A$.

We now show that it is in the interest of both firms to set $q_i = 1/2$, the maximum quantity. We will do the calculation for firm 1, as firm 2's case is analogous. If firm 2 plays $q_2 = 1/2$ and w_2 is uniformly distributed in $[0, 1/2]$, then firm 1's profits from offering q_1, w_1 with $w_1 \leqslant 1/2$ are

$$2w_1 \cdot \left(\frac{2 - q_1}{2} - w_1\right) \cdot q_1 + (1 - 2w_1) \cdot \left(\frac{1 - q_1}{2} - w_1\right) \cdot q_1.$$

We can see that the maximum of this expression in $q_1 \in [0, 1/2]$ is attained with $q_1 = 1/2$.

Consider now the case of personalized wages. Workers have surplus vectors $s^\theta = (s_1^\theta, s_2^\theta)$ uniformly distributed in the segment $[(0, 0), (1, A)]$ in $\Re^2$. Therefore firm 2 always hires the q_2 best workers, and firm 1 hires the next q_1 best ones. The market clearing equations imply $p_1 = 1 - q_1 - q_2$, and $p_2 = \epsilon(1 - q_2) + p_1$. We have $\eta(H_{21}) = q_2$, whereas $\eta(H_{12}) = 1 - q_2 - p_2/A = q_1/A$. Therefore, the marginal revenue formula yields

$$MR_2 = p_2 - q_2,$$
$$MR_1 = p_1 - q_1/A.$$

To solve for equilibrium we only have to set $MR_1 = MR_2 = 0$, and the formula in the text for q_1^* and q_2^* obtains.

References

Abdulkadiroglu, A., Pathak, P.A., Roth, A.E., 2009. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC High School match. Amer. Econ. Rev. 99 (5), 1954–1978.
Azevedo, E.M., Leshno, J., 2013. A supply and demand framework for two-sided matching markets. Mimeo, University of Pennsylvania.
Balinski, M., Sönmez, T., 1999. A tale of two mechanisms: Student placement. J. Econ. Theory 84 (1), 73–94.
Becker, G.S., 1973. A theory of marriage: Part I. J. Polit. Economy 81 (4), 813.
Biró, P., 2007. Higher education admission in Hungary by a score-limit algorithm. In: The 18th International Conference on Game Theory at Stony Brook.
Bulow, J., Levin, J., 2006. Matching and price competition. Amer. Econ. Rev. 96 (3), 652–668.
Cournot, A., 1838. Recherches sur les Principes Mathématiques de la Théorie des Richesses. Hachette, Paris.
Crawford, V.P., 2008. The flexible-salary match: A proposal to increase the salary flexibility of the national resident matching program. J. Econ. Behav. Organ. 66 (2), 149–160.
Ehlers, L., 2010. Manipulation via capacities revisited. Games Econ. Behav. 69 (2), 302–311.
Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much?. Quart. J. Econ. 123 (1), 49–100.
Gale, D., Shapley, L., 1962. College admissions and the stability of marriage. Amer. Math. Mon. 69, 9015.
Gale, D., Sotomayor, M., 1985a. Ms. Machiavelli and the stable matching problem. Amer. Math. Mon., 261–268.
Gale, D., Sotomayor, M., 1985b. Some remarks on the stable marriage problem. Discrete Appl. Math. 11, 223–232.
Ginsburg, T., Wolf, J.A., 2003. The market for elite law firm associates. Fla. State Univ. Law Rev. 31, 909.
Hatfield, J.W., Milgrom, P.R., 2005. Matching with contracts. Amer. Econ. Rev. 95 (4), 913–935.
Kelso, A.S., Crawford, V.P., 1982. Job matching, coalition formation, and gross substitutes. Econometrica 50 (6), 1483–1504.
Kesten, O., 2008. On two kinds of manipulation for school choice problems. Tepper School of Business. Paper 224.
Kojima, F., 2006. Mixed strategies in games of capacity manipulation in hospital-intern markets. Soc. Choice Welfare 27 (1), 25–28.
Konishi, H., Ünver, M.U., 2006. Games of capacity manipulation in hospital-intern markets. Soc. Choice Welfare 27 (1), 3–24.
Luenberger, D.G., 1969. Optimization by Vector Space Methods. Wiley–IEEE.
Mailath, George J., Postlewaite, Andrew, Samuelson, Larry, 2013. Pricing and investments in matching markets. Theoretical Econ. 8 (2), 535–590.
Milnor, J.W., 1997. Topology from the Differentiable Viewpoint. Princeton Univ. Press, ISBN 0691048339.
Mumcu, A., Saglam, I., 2009. Games of capacity allocation in many-to-one matching with an aftermarket. Soc. Choice Welfare 33 (3), 383–403.
Roberts, J., Sonnenschein, H.F., 1976. On the existence of Cournot equilibrium without concave profit functions. J. Econ. Theory 13 (1), 112–117.
Roth, A.E., 1985. The college admissions problem is not equivalent to the marriage problem. J. Econ. Theory 36 (2), 277–288.
Roth, A.E., Peranson, E., 1999. The redesign of the matching market for American physicians: Some engineering aspects of economic design. Amer. Econ. Rev. 89 (4), 748–780.
Sönmez, T., 1997. Manipulation via capacities in two-sided matching markets. J. Econ. Theory 77 (1), 197–204.
Sorensen, M., 2007. How smart is smart money? A two-sided matching model of venture capital. J. Finance, 2725.