Matching and Price Competition*

Jeremy Bulow
Graduate School of Business
Stanford University

Jonathan Levin
Department of Economics
Stanford University

July 2003

Abstract

We develop a model in which firms set their salary levels before matching with workers. Wages fall relative to any competitive equilibrium while profits rise almost as much, implying little inefficiency. Furthermore, the best firms gain the most from the system while wages become compressed. We explore the performance of alternative institutions and discuss the recent antitrust case against the National Residency Matching Program in light of our results.

*Email: jbulow@stanford.edu and jdlevin@stanford.edu. We thank Simon Board and Peter Coles for their research assistance. We also particularly thank Susan Athey, Jeffrey Ely, Amy Finkelstein, Paul Klemperer, Paul Milgrom, Muriel Niederle, and Alvin Roth for helpful suggestions. Levin gratefully acknowledges support from the Hoover Institution.

1. Introduction

A recent antitrust suit charges that the National Resident Matching Program suppresses the wages of medical residents. The match, which uses a Gale-Shapley procedure to assign seniors in medical schools to residency programs in various medical specialties, was developed for efficiency reasons, and on that score it appears to do quite well.[1] That is, the right residents appear to get assigned to pretty much the right residency programs. At the same time, for young doctors who have just completed four years of medical school, salaries are both low, averaging under $40,000 per year, and compressed, and work hours are long, 80 hours a week in many programs.[2]

We develop a model that shows why a market like that for medical residents is likely to have the features described, namely good efficiency properties, salaries that are below those in any competitive allocation, and severe compression in compensation. The key elements are two: competition is likely to be somewhat localized, with hospitals basically competing against others like themselves, and hospitals cannot easily make offers that discriminate among candidates. The model does not argue against a centralized match, but rather explains how the absence of personalized prices can soften competition in a matching market.

We consider a model in which both "hospitals" and "residents" are easily ranked so there is no ambiguity about what constitutes an efficient match. This is the cleanest case for our analysis; in other cases deriving our results would be easier. Hospitals make offers, with the hospital that offers the highest wage getting the best resident and so on. An analogy would be to a condominium auction where buyers make sealed bids and pay their own bids, with bidders receiving priority in choosing units based on the rank order of their bids. The crucial feature is the "all-pay" element to

[1] Alvin Roth and collaborators have written a fascinating series of papers documenting the history of the match, the reasons for its success, and changes in its structure over time; key references include Roth (1984), Roth and Xing (1994), and Roth and Peranson (1999). The first theoretical study of matching algorithms is by Gale and Shapley (1962), who analyzed a "deferred acceptance" procedure that is similar both to the procedure then used by the NRMP and to the one currently in use. Their algorithm was extended to allow for endogenous price determination in two important papers by Crawford and Knoer (1981) and Kelso and Crawford (1982). Milgrom (2003) unifies and extends many of the central results in the literature.

[2] Recent discussions of the antitrust case include Chae (2003) and Miller and Greaney (2003).

competition. A hospital will pay its offer regardless of the resident it actually matches with; it cannot offer $5x$ for the obstetrical Barry Bonds, but only x for the obstetrical Mario Mendoza.[3]

Were all hospitals symmetric in their distributions for the value of obtaining the best residents this system would lead to the same average wages as a system with discrimination, although it would create some fairly mild compression in salaries. But the reality is that hospitals have a sense of where they stand, with more highly ranked hospitals effectively competing for more highly ranked candidates. This, combined with the all-pay feature, dampens competition.

In a competitive equilibrium the salaries of the residents will adjust so that each hospital prefers to hire the resident who is its efficient match. The difference in wages between two "adjacent" residents must be in the range between the amount extra that the lower and higher of the firms with which they will match are willing to pay for the superior worker. The surplus of the better hospital will exceed that of the lower hospital by at least the difference in the value of the output of the two firms with the lesser worker and at most the difference in the value of the output of the two firms with the superior worker.

With price setting followed by matching, the expected surplus of the better hospital will exceed that of the lesser hospital by more than the difference in output with the superior worker.[4] The reason is that the salary a hospital must offer to obtain in expectation its appropriately matched resident is less than what the hospital ranked just below it would have to offer to match in expectation with the same resident. This is because the higher ranked hospital will, on average, offer higher wages than the lower ranked hospital and therefore the higher ranked hospital faces less stiff competition than the lower ranked hospital.[5] Therefore, the difference in the expected surplus of two adjacent hospitals reflects the maximum differential in a competitive

[3] Most readers will recognize Bonds as baseball's greatest player over the past 15 years; Mendoza is best known for his struggles in keeping his batting average above his weight; a standard that has become known as the "Mendoza Line".

[4] Actually, for the top two hospitals the differential will equal this output differential but in all other comparisons the surplus comparison will be strict.

[5] To continue with our baseball analogy, the Yankees have an easier schedule than the Orioles because they face all the same opponents except that the Yankees get to play the Orioles and the Orioles must play the Yankees.

equilibrium, plus the savings from the lower wage that the higher firm must pay relative to its near competitor to achieve its expected efficient quality worker.

The incremental surplus differential between two adjacent workers may be small, but it cumulates: the surplus of a highly ranked hospital exceeds its competitive surplus by at least the sum of all the incremental differentials between the hospitals below it. So the highest ranked hospitals get the most extra surplus. Of course the entire increase in surplus must come from wages rather than increased productivity, since the competitive equilibrium is efficient. In fact, given that the pricing and matching structure allows some inefficiency to creep into the match, resident wages will decline in aggregate by more than the increase in hospital surplus. Because competition is localized, however — hospitals may offer wages that match them with residents a bit above or below their competitive match, but no one will offer a wage that leads to a massively inefficient allocation — the inefficiency in the market will be small. Finally, because the top hospitals gain the most, and they match with the best residents, the residents whose salaries are most reduced are the very best ones.[6] This accounts for the compression in the wage distribution.[7]

While we frame the paper in terms of the residency match, we caution readers not to draw strong inferences for the antitrust case against the NRMP. First, the match was designed for efficiency purposes; any wage effects were ancillary and unforeseen. Second, while we compare the match with an efficient competitive equilibrium, it is far from obvious that this is the alternative to the NRMP; other arguably similar labor markets that do not employ an NRMP-type algorithm perform quite poorly.[8] Third, we have chosen our assumptions for analytical simplicity and transparency,

[6] Such as Jon's wife, Amy, whose long hours inspired his work on this topic.

[7] One might expect the average wage problem to be mitigated by entry, though compression would still remain. In the resident market the accreditation process that also limits the size of residency programs may effectively limit entry. One might also expect the compression problem to be "relieved" in part by the exodus of high quality workers from the market. Mitigating this, a residency lasts for a relatively small fraction of most doctors' careers and part of the lower wages that residents receive may return to older doctors, meaning that part of the effect of the system may be a steepening of doctors' experience-income profile, particularly for the best doctors.

[8] See, for example, Avery, Jolls, Posner, and Roth (2001). It is also worth noting that in the context of medical fellowships, where a centralized matching process is used for some, but not all sub-specialties, salaries appear to be low across the board and do not seem to depend much on whether the market is centralized (Niederle and Roth, 2003).

not to be the most realistic possible model of the NRMP.

Our goal is really to make a set of points about markets like the residency market that share certain salient characteristics. In many professions (law, investment banking and academic economics being three important examples though not necessarily in that order) employers may reasonably conclude that it is in their interest to pay the same wage to all incoming employees. It is this non-discrimination feature that we regard as the focus of our study.[9]

The model also has some relevance for product markets, although one unusual feature is the assumption that firms who set price are also greatly concerned about the quality of their match. This seems reasonable in labor markets, but perhaps less so in many product markets. For example, a manufacturer's profits will typically be a function of the volume of its sales, not the identity of its customers. That said, there are businesses in which customer identity is crucially important. Life, health, and automobile insurance are the most obvious examples; credit markets are another. Other relevant cases are markets where service is bundled with the product and some customers are systematically more demanding. More generally, firms will care about customer identity whenever expected profits from any given transaction are affected by the buyer. The effect may be direct — a casino saving its most desirable rooms for customers who are known to gamble heavily , or indirect — a three star Parisian restaurant reserving half its tables for locals to avoid developing a touristic ambiance, or a nightclub striving to attract glamorous customers. Thus, while we believe our model has applications beyond labor markets, it cannot be universally applied to all product markets.[10]

[9] Within firms, the model is also broadly consistent with tournaments that allow the most talented workers to obtain the most responsible jobs but receive a disproportionate share of the firm's surplus. Within the context of our model one incentive for adopting such a pay mechanism, which raises average compensation, is that in equilibrium it might help the firm attract more of the most talented workers. But we have not pursued this analogy in any detail.

[10] Among the key features in our model are quality differentiation across sellers (for example, in credit cards American Express and the cards that credit customers with airline miles attract a vastly disproportionate fraction of high volume users; some other issuers have focused heavily on the subprime market), and capacity constraints. Our model would predict that the "best" customers would be disproportionately profitable, implying that free entry would lead to more suppliers focusing on the high end of the market.

A final example involves the U.K. 3G spectrum auction. Auction designers had to plan for the contingency that only four licenses, none quite identical to another, would be available in a market with four incumbents. A concern was that in a straight Vickrey (simultaneous multiple round) auction no entrants would think they could outbid an incumbent and so prices would end up very low.[11] On the other hand, if one of the licenses were set aside for an entrant then one of the incumbents would have to be excluded, which might be very inefficient. In order to encourage entry while still making it possible that the four incumbents would ultimately win Paul Klemperer, the principle auction theorist, proposed an "Anglo-Dutch" design, which would have involved a conventional English ascending auction until five bidders were left followed by a sealed bid final round. One possibility was that after the final round the highest bidder would get first choice of the licenses, the second bidder second choice, and so on.[12]

Had the U.K. gone this route, the problems facing the various bidders in a sealed bid round would have been somewhat different. The new entrant might have regarded it as unlikely, and probably undesirable, that it would pay enough to get one of the larger licenses, but would hope that its bid would enable it to defeat 1-2-1 or Orange, the two weaker incumbents. British Telecom and Vodafone, on the other hand, might have been confident that they would be in the top four and might be principally concerned with whether they ended up with one of the biggest licenses. So the bidding would have contained an all-pay element like that in our model, with the prediction that the expected premia for the biggest licenses would be lower than in a Vickrey auction. Of course the primary reason for adopting this procedure would have been to guarantee a market price for the least expensive license, which in this case was much more important than achieving market differentials in price among the various licenses.

[11] Ultimately it was possible to create five licenses out of the available spectrum, assuring that at least one new entrant would win a license and thereby guaranteeing competition. So ironically an increase in the supply of licenses gave the designers confidence that prices would rise to competitive levels (Binmore and Klemperer, 2002).

[12] Another possibility was that the top four bidders in the sealed bid round could have all been guaranteed a license, the four then competing in a simultaneous multiple round auction for the four specific licenses with a minimum bid on each license equal to the fourth highest bid in the sealed bid round.

The outline of the paper is as follows: In section 2 we present a numerical example that illustrates the basic results in our model and an approach for solving for equilibrium. Section 3 briefly describes the model itself and section 4 describes hospitals' equilibrium salary offers, which will be determined by mixed strategies.[13] Sections 5 and 6 describe the competitive equilibrium, the high profits, and the salary compression that the model produces. Section 7 explains why the market has very good performance in terms of efficiency. Section 8 argues that while some of the wage compression would occur in a symmetric model in which firms are limited to one wage offer, most of the compression and all of the reduction in average wages is due to the combination of the one wage restriction and the asymmetry of the firms.

Section 9 discusses extending the model to permit more generalized preferences. Section 10 describes the consequence of allowing personalized offers to residents, concluding that each firm being able to offer only a small number of different wages would make a big difference (assuming that equity considerations outside the model would not prevent firms from doing this). Section 11 considers dynamic competition, in which instead of all firms simultaneously offering a fixed wage they are able to adjust their one wage offer as they observe the offers of the competition. Section 12 concludes.

2. A "Multiplication Game" Example

Assume there are N firms (hospitals) $1, 2, ..., N$, each interested in matching with one worker (resident). The workers, also labeled $1, 2, ..., N$ are strictly interested in maximizing their wage.[14] Their reservation wage level is zero. Firms are interested in maximizing the value of their output less the amount they pay in wages. The value

[13] Our interpretation of mixed strategies in this context is that when candidates visit hospitals they learn about subtle changes in the program from the previous year that might not be known to competitors, such as any increase in the number of nights on call that residents might expect to work. These changes affect the attractiveness of a hospital's compensation package and of course make its desirability appear to competitors to have some randomness.

[14] It is a simple generalization to make the number of firms and workers different. Excess workers simply would not match, so they are irrelevant. Excess firms force the minimum wage up to the minimum competitive equilibrium wage for the bottom worker who matches, and therefore raise all wages by exactly that amount.

of the output of firm n if it matches with worker m is $i \cdot j$, so we refer to this example as the "multiplication game". All this information is common knowledge. Here we explicitly solve the example for $N = 3$.

The efficient match clearly puts worker 1 with firm 1, worker 2 with firm 2, and worker 3 with firm 3, creating a total output of 14. In a competitive equilibrium the salaries of the three workers must be such that they all want to work and, at the given salaries, each firm prefers to hire the worker with whom it is matched. Therefore the wage for worker 1 will be $p_1 \in [0, 1]$, so that firm and worker 1 will each get surplus from matching with one another; the wage for worker 2 will be $p_2 \in [p_1 + 1, p_1 + 2]$ so that firm 2 but not firm 1 will be willing to pay the increment needed to hire worker 2 instead of worker 1, and the wage for worker 3 will be $p_3 \in [p_2 + 2, p_2 + 3]$. The lowest competitive wages for the workers are therefore 0,1, and 3, which are the wages that they would earn if the market cleared using a Vickrey auction. Similarly, at these wages the hospitals receive their maximum competitive surpluses of 1,3, and 6.[15]

In our model of the match, each of the three firms simultaneously offers a wage and a ranking of the workers into a computerized match system. The workers observe the wages and then list a ranking of the firms. Each worker will rank the firms from highest wage to lowest. Each firm will rank the best worker, worker 3, first, and worker 2 second. The match will then assign worker 3 to the firm that has offered the highest wage, worker 2 to the firm with the second highest wage, and worker 1 to the firm with the lowest wage. What we will show in the context of this example is a set of results — sub-competitive average wages, wage compression, and a high level of efficiency — that we generalize later in the paper.

In solving this example, we know immediately that firms will use mixed strategies. With pure strategies the middle bidder would only want to offer a fraction of a penny above the low bidder and the high bidder a fraction of a penny above the medium bidder. But in that case both of the lower two bidders would benefit by raising their bids to attract the top worker. The same logic implies that there will be no "atoms" in firm strategies, except possibly at a salary of zero. We know that zero must be the lowest salary offered; because a firm making the lowest offer is sure to obtain the lowest worker, having the lowest offer be strictly positive is inconsistent with profit

[15] Because of the symmetry of the problem the highest possible competitive wages are 1,3, and 6 and the lowest hospital surpluses are 0,1, and 3.

maximization.[16] What is more, every salary between zero and the maximum must be potentially offered by at least two firms. If no firms ever offer some range of salaries, it would be better to make an offer at the bottom of this range than just above it. Similarly, if only one firm makes offers on some range, it would always benefit by making offers at the bottom of the range. Finally, we conjecture (and later prove) that each firm will randomize over an interval of prices.

Therefore, the only real issue in solving for equilibrium is whether at the top and bottom wages only two or all three firms are randomizing. To answer this question we begin by assuming that all three firms are randomizing within a given range, with densities or "quit rates" of q_1, q_2, and q_3. For firm 1 to be indifferent to quitting or not requires that $q_2 + q_3 = 1$. That is, the value of obtaining a worker who is one level up is 1 to firm 1, so for it to be worth paying an extra Δ in wages it must raise the expected quality of its match by Δ also, and the expected increase in the quality of its match is just $(q_2 + q_3) \cdot \Delta$. Similarly, for firm 2 to be randomizing $q_1 + q_3 = \frac{1}{2}$, since an expected increase in quality of $\frac{1}{2}\Delta$ is just worth an extra payment of Δ to firm 2, and for firm 3 $q_1 + q_2 = \frac{1}{3}$. Solving these equations simultaneously yields a negative value for q_1, which is not feasible.[17] So only 2 and 3 will be randomizing at the top wage, and only 1 and 2 at the bottom.

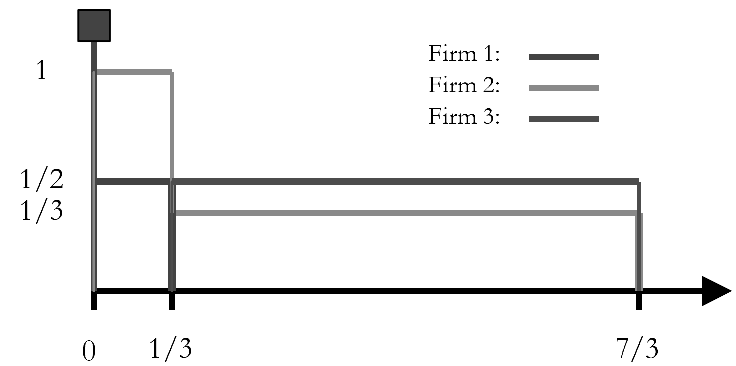
It is easiest to solve the equilibrium from the top. At the highest wages only firms 2 and 3 are randomizing, with densities $q_3 = \frac{1}{2}$ and $q_2 = \frac{1}{3}$ which are required to make the other indifferent to the randomization. Because $q_3 = \frac{1}{2} > q_2$ the range of wages over which 2 and 3 randomize must be from the maximum wage, call it $\overline{p}$, down to $\overline{p} - 2$. Within the range $[\,\overline{p} - 2, \overline{p}]$ firm 3 will exhaust its total bidding probability (a density of $\frac{1}{2}$ times a range of 2).Firm 2, however, will exhaust only $\frac{2}{3}$ of its probability (a density of $\frac{1}{3}$ over a range of 2). This leaves firm 2 with a probability mass of $\frac{1}{3}$ to employ over a range in which it competes with firm 1.

In the range where 1 and 2 compete $q_1 = \frac{1}{2}$ and $q_2 = 1$ to make the other indifferent to quitting. Given 2's quit rate and its available probability the length of the range must be $\frac{1}{3}$. Since we know that the minimum wage must be zero this means that 1

[16] To nail down this argument we must account for the possibility that some firms offer the lowest salary with discrete possibility. There cannot be two firms with "atoms" at the bottom, however, or one would want to bid a bit higher; and if there is just one, our argument applies to this firm.

[17] The solution is $q_3 = \frac{7}{12}$, $q_2 = \frac{5}{12}$, and $q_1 = \frac{-1}{12}$.

and 2 compete in a range of $(0, \frac{1}{3}]$ and therefore 2 and 3 compete in a range of $[\frac{1}{3}, \frac{7}{3}]$. Multiplying its quit rate times the range in which it competes only exhausts $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$ of firm 1's probability mass, which implies that it will bid 0 with probability $\frac{5}{6}$.



Equilibrium in the Multiplication Game

The equilibrium can be summarized as follows, with comparison to the competitive equilibrium that is most favorable to the firms in parentheses:

Table 1. The Multiplication Game

Wages		Profits	
Worker 1	0.02 (0)	Firm 1	1.00 (1)
Worker 2	0.73 (1)	Firm 2	3.67 (3)
Worker 3	1.56 (3)	Firm 3	6.67 (6)
Total	2.31 (4)	Total	11.33 (10)

The key results that will hold more generally are that wages are reduced and compressed but that the match is pretty efficient. In the example, almost four fifths of the reduction in expected wages goes to increased profits, while one fifth is deadweight loss. In the multiplication game with $N = 16$, 95 percent of the reduced wages goes to profits; this ratio approaches 100 percent as the number of matches rises. Relative to random assignment, the 3 person match captures 82 percent of the efficiency gain from competitive equilibrium; the 16 person match more than 97 percent. As N

increases, the expected wage of the bottom worker will always increase slightly, but more highly ranked workers lose more and more. In the three player game, virtually the entire wage reduction is borne by the top worker; in the 16 worker game the average wage declines by 11.5 but the expected wage of the best worker declines by 35.1.[18]

The profits results also generalize. The bottom firm always gets the same expected surplus as in the firm-optimal competitive equilibrium or Vickrey auction. Higher ranked firms get strictly more in expectation, with the very top two firms getting the same gain.

Finally, the algorithm for solving this game is basically the one we use to solve our general model. For example, in a five bidder "multiplication game" firms 3,4, and 5 compete over the top range. The bottom of 5's range is the top of 2's range. In the second range, then, 2,3, and 4 compete. When the bottom of 4's range is reached 1 will not find it profitable to compete with 2 and 3 (for exactly the same reason as in the three match game), so 2 and 3 will compete until the bottom of 3's range. Then 1 and 2 compete at prices down to 0, and 1 exits with an atom at a wage of 0. All firms exit at a constant rate within any given range.

3. The Model

Our general model has N firms and N workers. Each firm wants to hire a single worker. Firm n's surplus from hiring worker m is $v(n,m) = \Delta_n \cdot m$, where $\Delta_N \geq ... \geq \Delta_1 \geq 0$. Each firm that hires a worker pays a salary. If firm n hires worker m at a salary p, firm n's net utility is $v(n,m) - p$, while worker m's utility is p. A firm or worker that fails to match receives zero utility. These preferences, and the workers' abilities, are commonly known to the firms.

Several points deserve emphasis. First, the model allows some or all of the firms to have identical preferences. Nevertheless, the most notable results arise when there is some asymmetry between firms. Second, the multiplicative form of match surplus

[18] The percentage decline in wages in this example is slightly larger for higher ranked residents, and this effect would obviously greater if the reservation wage were above zero. But the claim we are making, and that we can show generally, is that the absolute decline in wages is largest for the top residents.

usefully simplifies the equilibrium, but is not essential. What is important is that v is increasing in m, so that workers are ranked in terms of their ability, and that v has increasing differences in n and m, so that firms are ranked in terms of how they value talent.

Finally, we emphasize that although we model salaries as prices, we think of a salary as broadly encompassing job features such as responsibility, hours and training, in addition to financial compensation. That being said, it is still limiting to assume that workers have homogenous preferences. We address this in Section 9.

The market unfolds in three stages. Each firm simultaneously makes a salary offer. These offers are observed by the workers. Matching follows. Workers rank firms by their offers, so the firm that makes the highest offer obtains the most able worker, and so on. To resolve ties, we assume that if several firms offer the same salary, matching is efficient — the firm with the highest value for talent gets the preferred worker. Once matching occurs, each firm pays its worker the salary it initially offered. In a *pricing equilibrium*, each firm chooses its offer to maximize its expected surplus, taking into account the matching process and the strategies of other firms. Note that each firm is implicitly required to rank all workers, so that every firm will match with a worker. While this assumption does not affect the equilibrium in the multiplication game, it can be relevant in some cases, particularly for lower ranked firms.[19]

4. Pricing and Matching

This section describes equilibrium salary offers. We start with the basic structure of the equilibrium, then proceed to the details. We defer a few technicalities to the Appendix.

The equilibrium, as in our example, involves mixed strategies. A mixed strategy for firm n is a distribution G_n, where $G_n(p)$ is the probability that n offers a salary less than or equal to p. We let g_n denote the density for firm n's offer distribution. As argued in the example, no firm can offer a price above zero with discrete probability. Also, standard arguments imply that in equilibrium each firm must randomize over an interval of prices.

[19] We discuss this further in section 11. We thank Jeffrey Ely and Amy Finkelstien for bringing this point to our attention.

We first establish the key qualitative feature of the equilibrium: higher ranked firms make (stochastically) higher offers.

Lemma 1 *If $\Delta_n \geq \Delta_m$, then in equilibrium firm n makes higher offers than firm m in the sense of first order stochastic dominance; for all p, $G_n(p) \leq G_m(p)$.*

Proof. Consider the returns to firm n to offering $p+dp$ rather than p. The benefit is the expected increase in worker quality, equal to $\Delta_n \cdot \sum_{k \neq n} g_k(p) \cdot dp$. The cost is the additional salary dp. Now compare this to the returns to firm $m < n$. Because $\Delta_m \leq \Delta_n$, the only way firm m could have a greater (or even equal) incentive to make the higher offer is if $g_n(p) \geq g_m(p)$. Now suppose that in equilibrium firm n makes offers over some interval $[p', p'']$. Since firm n prefers offering p'' to any higher price and $g_n(p) = 0$ above p'', firm m must also prefer p'' to any higher price. Between p' and p'', firm n is indifferent. This means that if firm m's offer interval overlaps with firm n's, then for any price p offered by both firms, $g_m(p) \leq g_n(p)$. Below p', firm n never makes offers, but firm m might. If follows that $1 - G_n(p) \geq 1 - G_m(p)$ for all p, establishing the claim. *Q.E.D.*

The logic behind Lemma 1 is that offering a higher salary attracts a more qualified worker (at least in expectation), but the higher salary must be paid regardless. Firms that care more about quality focus more on the benefit and make higher offers. If two firms are symmetric, so $\Delta_n = \Delta_m$, then they use the same equilibrium strategy. But if $\Delta_n > \Delta_m$, then n uses a strictly higher strategy: $G_n(p) < G_m(p)$ for all p between the lowest price offered by m and the highest offered by n.

The monotonicity property means that, in equilibrium, firms make offers over staggered price intervals. This basic structure is depicted in Figure 2.

Figure 2: Price Supports

Now consider the "head-to-head" competition that occurs at some given price p. If p is offered in equilibrium, it is offered by a consecutive set of firms $l, ..., m$. Each of these firms must be just indifferent to changing its offer slightly away from p. So for each $n = l, ..., m$,

$$\Delta_k \cdot \sum_{k \neq n} g_k(p) = 1.$$

By solving this system of equations, we obtain the firms' offer densities at p. For each firm $n = l, ..., m$:

$$g_n(p) = \frac{1}{m-l} \sum_{k=l}^{m} \frac{1}{\Delta_k} - \frac{1}{\Delta_n} \equiv q_n(l, m). \tag{1}$$

Conveniently, the offer densities depend on the set of firms competing, but not on p.[20] Taking advantage of this, we define $q_n(l, m)$ to be firm n's offer density given firms $l, ..., m$ are competing. Note that $q_n(l, m)$ is increasing in n; that is, higher firms "drop out" at a faster rate.

Our next Lemma resolves the question of which firms compete head-to-head.

[20] This is where the linear form of the surplus function comes into play: the incremental benefit to getting worker 3 rather than 2 is the same as from getting 2 rather than 1.

Lemma 2 *If firm m is the highest-ranked firm that offers p, then firm $l(m), ..., m$ all offer p, where:*

$$l(m) \equiv \min \{l : q_l(l, m) > 0\}. \tag{2}$$

So if firm m is the highest ranked firm to offer p, we can write each firm n's offer density at p as $q_n(m)$, where $q_n(m) \equiv q_n(l(m), m)$ if $l(m) \leq n \leq m$, and $q_n(m) \equiv 0$ otherwise.

With these preliminaries, we can provide an algorithm to describe equilibrium behavior. We will let p_{N+1} denote the highest salary offered, and p_n denote the *lowest* salary offered by firm n.

As in our earlier example, the algorithm starts at the top. On the interval below p_{N+1}, firms $l(N), ..., N$ compete head-to-head; and each firm's offer density is given by $q_n(N)$. Now, because $q_N(N) \geq q_n(N)$ for all n, firm N will "use up" its offer probability below p_{N+1} faster than other firms. So this top interval will have length $1/q_N(N)$. Or, letting p_N denote the lowest price offered by N,

$$q_N(N) \cdot (p_{N+1} - p_N) = 1.$$

What happens just below p_N? Provided that the firms are not all identical, lower-ranked firms will have residual offer probability that is not used up between p_N and p_{N+1}. Suppose for instance that $\Delta_{N-1} < \Delta_N$. Then below p_N, firms $l(N-1), ..., N-1$ compete head-to-head; and each firm's offer density is given by $q_n(N-1)$.

More generally, suppose firms $m+1, ..., N$ "use up" their offer probability above p_{m+1}, but firm m does not. Then between p_m and p_{m+1}, firms $l(m), ..., m$ compete head-to-head; and each firm's offer density is given by $q_n(m)$. Firm m will use up its offer probability at its lowest offer p_m. By recursion,

$$\sum_{n \geq m} q_m(n) \cdot (p_{n+1} - p_n) = 1.$$

Given a starting point p_{N+1}, this process continues until we have specified the behavior of firms $2, ..., N$. At this point, there are two possibilities. If $\Delta_1 = \Delta_2$, then firms 1 and 2 must use identical strategies, so we have also specified firm 1's behavior. If $\Delta_1 < \Delta_2$, then firm 1 has some residual probability, so it offers the lowest price with discrete probability equal to:

$$G_1(0) = 1 - \sum_{n=2}^{N} q_1(n) \cdot (p_{n+1} - p_n).$$

In either case, the lowest price offered by the two lowest firms must be zero, so $p_1 = p_2 = 0$. Given this, we complete the derivation by adding up the differences $p_{n+1} - p_n$ to obtain the highest price p_{N+1}.

Proposition 1 *There is a unique pricing equilibrium. Letting $q_n(\cdot)$ and $p_1, ..., p_{N+1}$ and $G_1(0)$ be defined as above, then for each firm n, and each non-empty interval $[p_m, p_{m+1}]$, $g_n(p) = q_n(m)$ for all $p \in (p_m, p_{m+1}]$.*

Figure 3 illustrates the equilibrium offer distributions with five firms (using multiplication game payoffs). Only two firms mix concurrently over the lowest range of prices, but more than two firms may mix over higher ranges of prices. Indeed the "pool size" is increasing over the price range.[21]

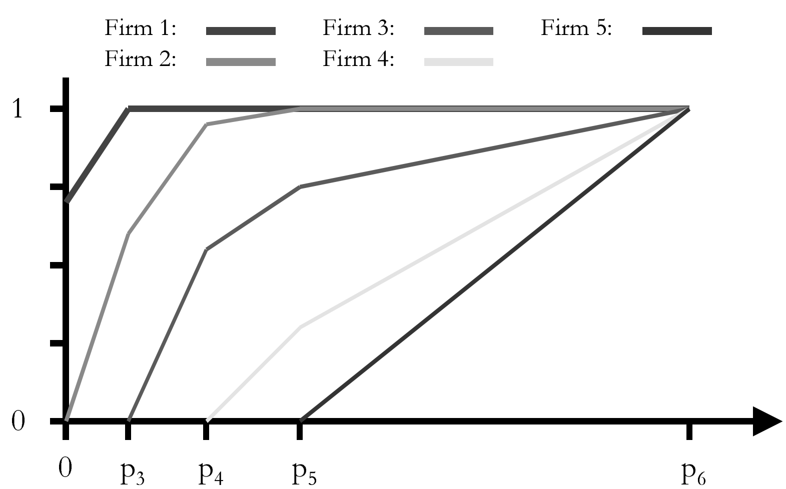


Figure 3: Equilibrium with Five Firms

Interpreting the Mixed Strategies

A concern with any mixed strategy equilibrium is how to interpret the predicted behavior. In reality, it seems unlikely that hospitals randomize their offers to residents. A more appealing view is that mixing captures strategic uncertainty inherent in the competition between the firms. For instance, suppose that hospitals alter their

[21] In general, the "pool size" increases over the price range provided that Δ_n is concave in n (or at least less convex than an exponential curve x^n).

programs each year for a wide variety of private reasons. Residents observe these changes when they interview, but competing programs do not. Under appropriate assumptions, behavior in such a world will be non-random, but observationally equivalent to the equilibrium we have described (Harsanyi, 1974).

5. Competitive Equilibria

In the model, as in the residency match, firms make salary offers prior to matching. An alternative would be to negotiate salaries in the process of matching. This section describes the competitive equilibria that might arise from an idealized form of negotiations and relates them to the Vickrey auction.

A *competitive equilibrium* is a matching of firms and workers, with corresponding salaries, that satisfies two conditions. First, it is individually rational; each firm and worker gets at least zero utility. Second, at the going worker salaries, each firm prefers the worker with which it is matched to any other worker.

There are a range of competitive equilibria. Each involves efficient matching, but salaries vary. To see why, suppose that firm 1 hires worker 1 at an individually rational salary $p_1 \in [0, \Delta_1]$. Firm 2 must pay worker 2 enough that firm 1 is not tempted to hire worker 2, but not so much that firm 2 wants to hire worker 1. Any salary $p_2 \in [p_1 + \Delta_1, p_2 + \Delta_2]$ serves this purpose. More generally, equilibrium requires that for each n, $p_n - p_{n-1} \in [\Delta_{n-1}, \Delta_n]$.

This puts a bound on competitive equilibrium salaries. Firm n must pay at least $p_n^F = \sum_{k<n} \Delta_k$, but not more than $p_n^W = \sum_{k \leq n} \Delta_k$. From the firms' perspective, the best competitive equilibrium has salaries $p_1^F, ..., p_N^F$; the worst has salaries $p_1^W, ..., p_N^W$.

In this regard, the Vickrey auction provides a useful benchmark. If firms bid for workers in a Vickrey format, the outcome is efficient matching with "firm-best" salaries $p_1^F, ..., p_N^F$. So each firm's Vickrey profit provides an upper bound on what it could expect in any competitive equilibrium.[22] Similarly, each worker's Vickrey salary provides a lower bound on her competitive equilibrium salary.

[22] Note that in the Vickrey outcome, each worker receives less that her marginal contribution to social surplus; p_n^F is less than worker n's marginal contribution p_n^W.

6. Profits and Salaries

This section compares firm profits and worker salaries in our model to competitive equilibrium profits and salaries. We obtain three main results. First, each firm's equilibrium profit is *at least as large* as its Vickrey profit. Second, worker salaries are lower in aggregate than their Vickrey salaries. Finally, worker salaries are compressed: relative to competitive equilibrium, the worst worker may benefit from the pricing and matching system, but salaries at the top are reduced.

We start with firm profits. Let $\Pi_n(p)$ denote firm n's expected profit if it offers p and other firms use their equilibrium strategies:

$$\Pi_n(p) \equiv \Delta_n \cdot \left[1 + \sum_{k \neq n} G_k(p)\right] - p.$$

If firm n offers p, it expects to attract a worker quality of $1 + \sum_{k \neq n} G_k(p)$ and to pay p.

Firm n's equilibrium profit Π_n is equal to $\Pi_n(p)$ for any p in the support of n's equilibrium strategy. In contrast, firm n's Vickrey profit is equal to:

$$V_n = \Delta_n \cdot n - \sum_{k<n} \Delta_k.$$

These two profits are exactly equal for the lowest ranked firm. In equilibrium, firm 1 is willing to offer zero and receive the lowest worker with certainty, so $\Pi_1 = \Pi_1(0) = \Delta_1$. Similarly $V_1 = \Delta_1$.

To compare profits more generally, we consider the profit differential between adjacent firms. Let $\hat{p}_n$ denote the price such that if firm n offers $\hat{p}_n$, its expected worker quality is n, its Vickrey quality. Such a price must exist in n's offer region: when n makes its highest offer, it expects to beat all firms $k < n$ with certainty and obtain at least worker n; on the other hand, when n makes its lowest offer, it expects to lose to all firms $k > n$ with certainty and obtain no better than worker n. So $\hat{p}$ must lie between these two extremes. Moreover, firm $n-1$ must also offer $\hat{p}_n$, or else n would expect quality strictly greater than n when it offered $\hat{p}_n$.

The difference in equilibrium profits between firms n and $n-1$ is $\Pi_n(\hat{p}_n) - \Pi_n(\hat{p}_n)$. Substituting and re-arranging:

$$\Pi_n - \Pi_{n-1} = (\Delta_n - \Delta_{n-1}) \cdot n + \Delta_{n-1} \cdot \left[G_{n-1}(\hat{p}_n) - G_n(\hat{p}_n)\right].$$

The first term is exactly $V_n - V_{n-1}$, the difference in the Vickrey profits of n and $n-1$. The second term, which is always non-negative, arises from the fact that firm $n-1$ makes lower offers in equilibrium, so its competition is tougher. The second term is *strictly* positive whenever $\Delta_n > \Delta_{n-1}$ and $\hat{p} < \overline{p}$. So $\Pi_n - \Pi_{n-1} \geq V_n - V_{n-1}$ and the inequality is typically strict.

So equilibrium profit and Vickrey profit are the same for the lowest firm, but if $\Delta_2 > \Delta_1$, firm two's equilibrium profit is strictly higher than its Vickrey profit, and the same is true for every firm $n > 2$. Moreover, if $\Delta_{n+1} > \Delta_n$, the gap increases for firms 3,4, and so on. The gap does not increase at the very top. Instead, the difference in equilibrium profits between the top two firms coincides with the Vickrey differential. This occurs because $\hat{p}_N = \overline{p}$, and $G_n(\overline{p}) = 1$ for all firms.

We summarize as follows.

Proposition 2 *All firms have expected equilibrium profits greater than their Vickrey profits. Moreover, the difference cumulates: the lowest firm has no profit change, while the highest firm sees the biggest increase.*

The key force is that low ranked firms are less aggressive in equilibrium than high ranked firms. So firm n not only derives greater value from a given worker that firm $n-1$, it also expects, conditional on offering a given salary, to receive a better worker. This generates a larger profit differential between firms than in a Vickrey auction, where n's extra profit relative to $n-1$ is just the difference in their values for a given worker.

Now consider the workers' perspective. Because there is mixing, firms and workers may not be efficiently matched. So the expected equilibrium surplus is less than the efficient competitive equilibrium surplus. Because firm profits are higher, worker salaries must be lower in the aggregate.

Not every worker is necessarily worse off. The worst worker expects a non-zero salary in equilibrium. This improves on her Vickrey salary of zero (though it may or may not be lower than her highest possible competitive salary). For the best worker, however, even her highest possible equilibrium salary falls below her Vickrey salary. We will provide some quantitative examples in the next section to show that the difference (equal to $\Pi_N - V_N$) is often very large.

Proposition 3 *The aggregate surplus that accrues to workers in equilibrium is strictly less than in any competitive allocation. Moreover, wages are compressed; the worst worker does better and the best worker does worse than under competition.*

Propositions 2 and 3 generalize easily to the case where match surplus is not simply multiplicative. Suppose that firm n's value for worker m is given by $v(n, m)$, where v is increasing in m and has increasing differences in (n, m). (Recall that v has *increasing differences* if for all $m' > m$, $v(n, m') - v(n, m)$ is increasing in n.) This specification includes the multiplicative case $v(n, m) = \Delta_n \cdot m$, as well as cases where firms have increasing or decreasing returns to worker quality. We show in the Appendix that the qualitative features of the equilibrium are preserved in this more general model and establish the following result.

Proposition 4 *Suppose firm values are given by $v(n, m)$, where v is increasing in m and has increasing differences in (n, m). Then equilibrium firm profits exceed Vickrey profits, while equilibrium worker salaries are less than Vickrey salaries on aggregate and more compressed.*

7. Local Competition and Efficiency

This section examines market efficiency. Low-ranked firms may outbid higher-ranked firms in equilibrium, but because firms compete "locally" against similar opponents, the inefficiency this creates is limited. Relative to competitive equilibrium, there is far more redistribution of surplus than loss of surplus. This is noteworthy from an antitrust standpoint because an institution that creates market power may be justified if it generates substantial efficiency gains.

To expand on this point, we approximate market efficiency and firm profits. Two simplifications prove useful for this purpose. First, we focus on the case where firms' valuations are uniformly distributed, with $\Delta_n = n/N^2$.[23] The normalization means that the surplus from matching the highest worker and firm is always 1, while the bottom match has a value approaching 0 as N increases. Second, we focus on markets with a "large" number of firms (though we provide some numerical results below for small N).

[23] Our approximations generalize, with similar conclusions, provided that $1/\Delta_n$ is convex in n.

We first address the extent to which competition is local. It turns out that the number of higher-ranked firms that a given firm n could conceivably outbid — the "pool size" of firm n — is roughly $\sqrt{2n}$.

Lemma 3 *Suppose $\Delta_n = n/N^2$. If $\rho(n) \equiv l^{-1}(n) - n$, then $\rho(n) \approx \sqrt{2n}$.*

Now consider the efficiency loss in equilibrium. The social cost from firm n displacing firm $m > n$ by one place is $\Delta_m - \Delta_n$. So relative to an efficient assignment, the loss from an equilibrium assignment is the cost generated by all such displacements, weighted by the probability that they occur. Thus,

$$I(N) = \sum_{n=1}^{N-1} \sum_{k=1}^{N-n} (\Delta_{n+k} - \Delta_n) \cdot \Pr\left[n \text{ beats } n+k\right].$$

Our pool size result bounds this expression. Because firm n cannot beat any firm greater than $n + \rho(n)$, and cannot beat *any* higher-ranked firm with more than 1/2 probability:

$$I < \frac{1}{2} \sum_{n=1}^{N-1} \sum_{k=1}^{\rho(n)} \frac{k}{N^2} \approx \frac{1}{2} \sum_{n=1}^{N-1} \Delta_n \approx \frac{1}{4}.$$

This bound is rough in that it assumes n beats every firm between $n+1$ and $n+\rho(n)$ with probability 1/2. Our numerical results suggest substantially less inefficiency, though of the same order of magnitude.

Now consider the magnitude of re-distribution. From the previous section, we know that

$$\Pi_n - V_n = \sum_{m=1}^{n-1} \Delta_m \cdot \left[G_m(\hat{p}_{m+1}) - G_{m+1}(\hat{p}_{m+1})\right].$$

Our pool size result suggests that, in a large market, the bracketed term is approximately $1/\rho(m)$. Thus

$$\Pi_n - V_n \approx \sum_{m=1}^{n-1} \frac{\Delta_m}{\rho(m)} \approx \frac{1}{3}\Delta_n \rho(n).$$

This already provides a rough sense of equilibrium wage compression. In a market with N firms and workers, with $\Delta_n = n/N^2$, the competitive equilibrium salary of worker N is at most $\frac{1}{2} + \frac{1}{2N}$ and at least the Vickrey salary of $\frac{1}{2} - \frac{1}{2N}$. In contrast, the highest salary that could *possibly* be offered to worker N in equilibrium is precisely $N - \Pi_n$. The difference between these salaries is $\Pi_N - V_N$. It follows from our approximation that N's equilibrium salary is depressed by *at least* $\sqrt{2}/(3\sqrt{N})$. Furthermore, the expected salary of the top worker will of course be less than the maximum possible, by approximately $\sqrt{N}/(\sqrt{N}-1)$.

To identify the aggregate gain to firms in equilibrium, we sum the excess profits for each firm. Then,

$$E(N) = \sum_{n=1}^{N}(\Pi_n - V_n) \approx \frac{1}{15}\rho(N)^3 \approx \frac{N^{1/2}}{5}.$$

Of course, the fall in aggregate salaries is just $E(N) + I(N)$.

Relative to competitive equilibrium, profits rise and salaries fall by an order of magnitude more than the change in total surplus. In this sense, equilibrium generates far more re-distribution than inefficiency. As a rough benchmark for calibrating the size of these effects, note that the total surplus at stake, i.e. the surplus difference between an efficient matching and random assignment, is:

$$S(N) = \sum_{n=1}^{N} \Delta_n \cdot \left(n - \frac{N+1}{2}\right) \approx \frac{N}{12}.$$

We summarize the discussion in the following Proposition.

Proposition 5 *Suppose that $\Delta_n = n/N^2$. Then when the market is sufficiently large, the inefficiency of the market is approximately of order* 1*, excess profits are of order $N^{1/2}$ and market surplus is of order N.*

As the market becomes very large, equilibrium becomes approximately efficient in the sense that virtually 100% of the total possible surplus is realized. Indeed, in the limiting case when firm values and worker qualities are continuously distributed, there is a pure strategy equilibrium with Vickrey prices.

It is perhaps useful to calibrate the magnitude of our results. Table 2 provides some calculations. To avoid decimals, we assume firm values are distributed between 0 and $100/N$ (i.e. $\Delta_n = 100n/N^2$), rather than between 0 and $1/N$.[24]

[24]This is equivalent to assuming that both firm and worker values are uniformly distributed between 0 and 10.

Table 2. Numerical Calculations

N	5	10	50	100	500	1000
Surplus at Stake	40	83	417	833	4167	8333
Excess Profits	25	48	144	218	527	759
Inefficiency	4	4	3	3	3	3

The first row shows the surplus at stake, or the difference between the surplus generated by efficient matching and random assignment. This grows at rate N. The second row is the aggregate difference between equilibrium profits and Vickrey profits. In small markets, this is a large fraction of the surplus at stake (about 25% with 100 firms), and is still significant in a large market (about 10% with 1000 firms). The last row shows the loss of efficiency from equilibrium matching. The efficiency loss is nearly 10% of the surplus at stake with five firms, but it drops off rapidly. By the time there are 100 firms, the equilibrium efficiency loss is roughly 0.5%.

As discussed above, the gains and losses are largest at the top. With 100 firms, the top firm expects a profit of 55.2 in the match and 50.5 in the Vickrey auction, a 9.3 percent gain. The top worker expects a wage of 49.5 in the Vickrey auction. In the match, her maximum possible wage is 44.8. Her expected wage is 43.9, an 11.2 percent loss relative to Vickrey. The top firm's gain is more than twice that of the average firm. The top worker's wage is depressed two and a half times the amount of the average worker.

8. The Role of Nondiscrimination

It is tempting to attribute our results entirely to the fact that firms do not target their salaries offers. On this account, firms are less aggressive because they must make offers without knowing precisely whom they are hiring. This argument is incomplete, however. While nondiscrimination does generate salary compression, it cannot on its own account for an aggregate reduction in salaries. Rather, it is the combination of the salary-setting process and asymmetries between firms that depresses competition.

It is perhaps easiest to see this in the context of an example. Imagine two worlds. In the first, firms draw their values independently and privately from a uniform distribution on $[0, N+1]$. In the second, firm values are drawn without replacement from $\{1, 2, ..., N\}$. The latter is our multiplication game model.

The environments are parallel in the following sense. In both, the expected value of the top firm is N, the expected value of the second firm is $N-1$, and so on. Moreover, the Vickrey allocations are identical in expectation. The nth firm has expected value n, matches with worker n, and pays the sum of the lower valuations, equal in expectation to $\sum_{k=1}^{n-1} k$. Given this, we denote aggregate expected surplus, profits and wages in the Vickrey auction by S, V and W.

Now consider what happens if the firms set salaries simultaneously, with the best worker going to the top offer and so on. In the first case, firms have symmetric beliefs about the values of their competitors. There is a symmetric pure strategy equilibrium, in which each firm makes an offer that depends monotonically on its value for quality. Expected salaries are compressed relative to the Vickrey auction: worker N expects a lower salary in equilibrium than in a Vickrey auction,[25] while worker 1 expects a positive salary under the match and a zero salary under the Vickrey auction. Nevertheless, because firms with higher values make higher offers, equilibrium is efficient. The expected surplus is again S. Moreover, the Revenue Equivalence Theorem implies that a firm with value Δ expects precisely the same profit in equilibrium as it does in a Vickrey auction. So aggregate firm profits and worker salaries are given by V and W in expectation.

In contrast, with asymmetric firms, the match leads to worker salaries that are significantly lower in aggregate and substantially compressed relative to the Vickrey allocation. Also, market surplus is lower. Thus, the combination of nondiscriminatory pricing and asymmetry are what generate a departure from competitive outcomes. This situation is summarized in Table 3.

[25] In the Vickrey auction, conditional on any top signal, the top firm will pay its expected Vickrey cost *conditional on its being the highest ranked firm.* In the match, conditional on the same signal, the firm will pay its unconditional expected payments in a Vickrey auction, averaging in the lower costs it will have in cases where it is outbid for the top worker(s). This amount will clearly be lower in the independent values case. The symmetric equilibrium is for a firm with value Δ to offer $p(\Delta) = \frac{1}{2}\frac{N-1}{N}(\Delta^2)$. If Δ^1 and Δ^N are the first and N^{th} order statistics, workers 1 and N expect salaries $p(\Delta^N)$ and $p(\Delta^1)$. A small amount of algebra shows that the expected highest offer is $\frac{N(N-1)}{2}\frac{N+1}{N+2}$, which is less than $\frac{N(N-1)}{2}$, the expected Vickrey salary for worker N. For the bottom worker the Vickrey salary is zero while the expected symmetric all-pay salary would be $\frac{N-1}{N+2}$.

Table 3. Asymmetry and Nondiscrimination

	Symmetric Firms		Asymmetric Firms	
	Vickrey	Match	Vickrey	Match
Market Surplus	S	S	S	$< S$
Firm Profits	V	V	V	$\gg V$
Worker Salaries	W	W	W	$\ll W$
Salary Compression	No	Some	No	Yes

We can connect this in a slightly more precise way to auction theory in the following way. The Vickrey auction is essentially a "second-price" environment, where to get a worker firms have to pay just enough to outbid their next competitor. In contrast, the match has an "all-pay" flavor. Auction theory tells us that the symmetric Vickrey and match auctions will both yield the same average profits for each type of firm. Therefore, average wages must be unchanged as well and the scope for compression is limited. In contrast, in an asymmetric match, equilibrium behavior leads to progressively greater expected profits for the more highly ranked firms, and therefore to progressively lower wages for the workers with whom the better firms match. What is remarkable is the degree of redistribution in the model relative to the very small amount of inefficiency.[26]

9. General Preferences

This section discusses extending our model to permit more general preferences. Two aspects of the model are natural to investigate. First, we assume workers care strictly about salary. Even if one takes a broad view of salaries as encompassing non-financial aspects of compensation, it is likely that idiosyncratic preferences of residents play a role in the residency labor market. Second, we assume firms hire a single worker. In practice, hospitals hire a cohort of residents each year, as do firms in many other entry-level labor markets.

[26] Baye, Kovenock and de Vries (1993) make another interesting observation about asymmetric all-pay auctions, which is that revenue (here wages) may actually increase if a high value bidder is excluded. They consider a single-unit auction, but their point would also apply here.

Heterogeneous Worker Preferences

Suppose workers to have heterogeneous preferences over firms, so that worker m's utility from a match with firm n is $p_n + u_{nm}$. The new issue is that from the workers' perspective, firms are now "differentiated products."[27] If the shocks are uncorrelated with the value of the match (as might be the case with "geographical" preferences) then standard pricing theory intuition is that this differentiation will increase the market power of each hospital and relax competition. The basic logic in that firms will try to exploit their market power over workers who idiosyncratically prefer them by lowering their offers; and in equilibrium salaries will be lower across the board.

While we do not have general results to this end, it seems roughly consistent with a number of examples we have worked out. An additional subtlety is that a new source of inefficiency may arise. If firm n makes a low offer targeted at some worker m that has a preference for n, then if for some reason n fails to attract m (either other firms happen to make better offers or because m's preference for n isn't as great as n expected), then n may be unable to attract any other worker with its offer. The possibility that positions will end up vacant generates a new source of inefficiency.

A second type of heterogeneous preference might be that better students might be willing to pay more to match with a better hospital than will less good students. An extreme case might be to reverse the asymmetry in our model and imagine that the hospitals are only interested in paying the lowest possible wage for a worker, while the value of a match of worker m to hospital n might be $n \cdot m$ plus the residency wage, less the reservation utility. In this case workers will receive their Vickrey wages and utilities, with the top firms extracting their share of the rents by paying the lowest wages.

An intermediate case of interest is where the value of a match was $\alpha \cdot n \cdot m$ plus a constant and minus the wage to the hospital and $\beta \cdot n \cdot m$ plus the wage minus a constant to the residents, $\beta \geq \alpha$. Further assume that there is a legal minimum wage which is high enough to insure that even the lowest worker would accept a match with the lowest quality hospital. Then it is an equilibrium for every hospital to offer the minimum wage. In this case workers would do well in utility terms — certainly better

[27] One case that does not lead to new issues is if workers have the identical cardinal preferences over firms. For instance, suppose every worker gets direct utility u_n from matching with firm n. Then firm n can offer $p_n - u_n$ and rather than p_n and the equilibrium we have considered is preserved.

than their Vickrey utility and perhaps better than in any competitive equilibrium — but we would still observe a compression in "non-prestige" compensation.

Multiple Hires

Allowing firms to hire multiple workers is easily accommodated in our model if each firm wants to hire the same number of workers, say k. Then the firm that makes the highest offer gets workers $N, N-1, ..., N-k+1$, the next offer attracts $N-k, ..., N-2k+1$, and so on. This model is strategically equivalent to a single worker per firm model where the quality of successive workers is $(k+1)/2$, $(3k+1)/2$, $(5k+1)/2$,... rather than $1, 2, 3, ...$[28]

If firms have different numbers of jobs, the nondiscriminatory nature of pricing works to the advantage of small firms. To see this, suppose there is a firm of size 5 and another of size 10. By beating the larger firm, the small firm realizes a total gain in worker quality of 50. The large firm realizes the same gain by beating the small firm. But the small firm pays only half as much for increasing its offer. This gives it a greater incentive to make aggressive offers. This issue may not be empirically relevant if firms of a given quality tend to be of similar size, for example if the top residency programs are larger and the lesser programs are all smaller.

10. Personalized Offers

In the context of the residency market, and in many other markets, it is natural to think of firms as making a single offer and hoping to attract the best match. Nevertheless, we would like to know what happens if firms make personalized offers prior to matching. We now show that this restores competitive equilibrium.

Suppose firms compete by simultaneously making offers to each worker. Let p_{nm} denote the salary that n offers to worker m, and let $\mathbf{p}_n = (p_{n1}, ..., p_{nN})$. Salary offers induce worker and firm preferences. A potential difficulty is that these preferences need not be identical, so matching is more complex. We assume that given offers

[28] This extension assumes surplus is additive across workers. Nondiscriminatory pricing can have significantly negative consequences if, for example, firms value having a few superstars. The problem is that they have no way of making a few special offers to attract these candidates.

$\mathbf{p}_1, ..., \mathbf{p}_N$, the matching process results in a stable assignment of workers to firms.[29] Suppose that firms understand this matching process. A set of offers is then an equilibrium if no firm could change its offer and increase its expected profit.

One equilibrium involves each firm making two non-zero offers. Each firm n offers workers n and $n+1$ their respective Vickrey salaries p_n^F and p_{n+1}^F. Firm N makes a single offer to worker N. Given these offers, the only stable matching is efficient and firms make Vickrey profits. Moreover, no firm can do better. Firm n cannot lower its offer to worker n without losing her to firm $n-1$. And obtaining worker $m \neq n$ would require offering m at least p_m^F, giving n less than her Vickrey profit.

While this is not the only personalized offer equilibrium, all others share the same outcome. We prove the next result in the Appendix.

Proposition 6 *With personalized offers, every equilibrium yields efficient matching and Vickrey profits and salaries.*

In theory, if other firms make only a single offer, a firm could benefit by making some personalized discriminatory offers. Yet participants in the residency match report that personalized contingent offers are quite uncommon. One reason may be that pre-match agreements are discouraged by NRMP rules: "...both applicants and programs may try to influence decisions in their favor, but any verbal or written contracts prior to the submission of Rank Order Lists is a violation of the Match" (NRMP Policies, Section 8.0).

Even in the absence of market rules limiting discrimination, however, there are reasons — concerns about fairness or internal equity, for instance — to believe firms may prefer offering similar compensation to entry-level workers. Law firm associates and consulting firm analysts are an example of this. So even if hospitals were permitted to make contingent personalized offers to prospective residents, they might not do so extensively, or for some period of time.

[29] Here the notion of stability is with respect to the fixed prices $\mathbf{p}_1, ..., \mathbf{p}_N$. That is, an assignment is stable if there is no firm n and worker m, who would prefer to leave their current partners and match at the salary p_{nm}.

11. Rejecting Matches

In the multiplication game every firm ends up with positive surplus from its match no matter how unlucky it gets. But this would not always be the case in the more general model. To take our baseball example, a team that bid a high salary in the hopes of signing Barry Bonds might be better off signing no one if they lost out for Bonds, than having to pay the same salary to Mario Mendoza. Therefore, one way for them to compete for Bonds in a match process would be to make a high offer and exclude Mendoza from their list of acceptable matches.[30]

Consider the following example: There are 2 firms, 1 and 2, that might match with workers of qualities 1 and 4. Value produced is the product of a firm's number and a worker's quality.

If both firms were required to match with someone then the equilibrium would be that firm 1 would bid 0 half the time and randomize uniformly in the range between 0 and 3 the rest of the time. Firm 2 would always randomize uniformly between 0 and 3. The lower quality worker would have an expected wage of .5 while the higher quality worker would have an expected wage of 1.75 and a maximum wage equal to his Vickrey wage of 3.

However, if the firms are allowed to exclude the lower quality worker when they make a high bid, and obtain a utility of zero from not matching, the equilibrium changes. Now, for firm 2 $g_2(p) = \frac{1}{3}$ for $p \leq 1$ and $g_2(p) = \frac{1}{(4-p)^2}$ for $1 \leq p \leq 3$. That is, firm 2 must bid more aggressively when it bids more than 1 to make such a bid less profitable for firm 1, which no longer need worry about a negative surplus if it matches with worker 1. Firm 1 will again bid 0 with probability .5 and must bid as before in the range up to 2: $g_1(p) = \frac{1}{6}$ for $0 \leq p \leq 2$. For prices between 2 and 3 firm 1 must be a bit more aggressive to make such bids less profitable for firm 2, which would now also exclude worker 1. So $g_1(p) = \frac{5}{(8-p)^2}$ for $2 \leq p \leq 3$. In this equilibrium the higher value worker is clearly better off than in the game with mandatory matching, because bidding is more aggressive and he is never excluded; his average wage rises to 1.95. The lower value worker would be worse off (average

[30] We thank Jeff Ely for developing a three firm and worker numerical example illustrating this point, and showing that equilibrium would in fact require that some bids will contain exclusions if this is an allowable strategy.

wage of about .19) largely because the exclusions mean that he receives no wages more than two thirds of the time. The exclusions reduce the average payment of firm 1 from .75 to .49, while the more aggressive bidding (partly offset by exclusions) raises the average payment by firm 2 from 1.50 to 1.65. Thus, while the exclusions improve the lot of the best worker, in this example expected wages fall overall (from 3 in the Vickrey case to 2.25 in the match without exclusions to 2.14 in the match with exclusions).

With three firms and workers the same principles apply, and it is straightforward though tedious to verify the general result that expected surplus still exceeds the Vickrey level for all but the bottom firm. However we have been unable to generalize these results to more matches.

12. Dynamic Competition

Our model assumes that firms make offers with no knowledge of the offers being made by other firms. While independent salary decisions may be a good assumption in some markets, in others one might argue that the salary-setting process is dynamic, so that firms have some ability to respond to competing offers. We now explore this possibility.[31]

Suppose first that firms set salaries through an ascending price auction. Prices start at zero and increase over time. At any point a firm may "drop out" and fix its offer at the current level. The process stops once every firm fixes its salary.

With ascending prices, there is an equilibrium that results in efficient matching. This equilibrium involves all firms offering a zero salary, but dropping out in sequence. In this equilibrium, each firm believes that all higher firms will stay in until it drops out. This belief is justified because all firms also believe that whatever has happened, lower ranked firms are always just on the verge of dropping out.

This equilibrium is only one of many, however, and efficiency is not guaranteed. Indeed, precisely the same logic can be used to construct an equilibrium where all

[31] Kamecke (1998) also proposes a dynamic model of salary setting prior to a centralized match. The key features of his model are that there is no "last date" for matching opponents' offers, and also some scope for personalized offers off the equilibrium path. In our context, his salary setting process would result in efficient matching and all workers receiving zero salary.

firms drop out immediately in any arbitrary order! There are also many inefficient mixed strategy equilibria. These have an interesting war of attrition flavor; and in some, firms may even do worse than in the Vickrey auction.

As an example, suppose there are two firms and two workers. Consider an equilibrium as follows. From the outset, Firm 1 drops out at a rate $1/\Delta_2$ and firm 2 at a rate $1/\Delta_1$. In this equilibrium, Firm 1 makes an expected profit equal to Δ_1; Firm 2 expects Δ_2, a profit strictly less than its Vickrey profit of $2\Delta_2 - \Delta_1$. With more firms, the possibilities expand. One possible equilibrium with three firms is for Firm 1 to drop out from the outset at rate $1/\Delta_2$ and Firm 2 at a rate $1/\Delta_1$. Once one firm, say 2, drops out, Firm 3 starts dropping out at a rate $1/\Delta_2$ and Firm 2 decreases its drop-out rate to $1/\Delta_3$. In this equilibrium, both Firms 2 and 3 have expected profit below their Vickrey surplus.

Proposition 7 *With ascending prices, there are both efficient and inefficient subgame perfect equilibria. Some of these equilibria involve firms making profits lower than their Vickrey profits.*

On the other hand, if firms set salaries through a descending price auction, the outcome is always efficient. Specifically, suppose prices start as some very high level. As time progresses, firms simultaneously lower their prices. At any point a firm may drop out, fixing its offer. The process terminates once all firms fix their prices or alternatively if prices reach zero. The next result is proved in the Appendix.

Proposition 8 *With descending prices, the unique subgame perfect pricing equilibrium results in efficient matching and Vickrey prices.*

A third model of dynamic price setting supposes that firms set prices in some prespecified sequential order. In some cases, this leads to efficient matching at Vickrey prices. For instance, this occurs if firms set prices in descending order, starting with Firm N and ending with Firm 1. More generally, however, sequential price offers can result in inefficient matching. For example, suppose that there are three firms and three workers and that firm 2 makes the first offer, followed by firms 1 and 3.[32] If $\Delta_1 + \Delta_2 > \Delta_3$, the subgame perfect equilibrium involves Firm 2 making an offer

[32] We thank Peter Coles for this example.

$p_2 = \Delta_3 + \varepsilon$. Given such an offer, both Firms 1 and 3 will offer zero. The result is that Firm 2 ends up with the most talented worker and Firm 3 with the second most talented. In this example, each worker ends up with a salary below his Vickrey salary. We suspect, but have not proved, that this is a general property of sequential offer games.

13. Conclusion

This paper has studied matching markets where firms compete by setting prices prior to matching. The "all-pay" nature of competition leads to greater profits, with the highest quality firms benefiting the most. The implication is that wages are both reduced and compressed, with the compression far beyond the mild amount that will occur in all-pay competition among symmetric firms with the same expected distribution of quality.

Appendix A: Derivation of the Equilibrium

This appendix fills in the details omitted from Section 3. We derive the equilibrium and prove uniqueness in a series of steps.

1. (No Atoms) No equilibrium distribution G_n can have an atom at $p > 0$.

Proof. Suppose firm m offers $p > 0$ with discrete probability. Then no firm $n \neq m$ could optimally make an offer in a small interval below p, say the interval $[p - \varepsilon, p)$ and no firm $n < m$ will offer p. But then firm m could not be optimizing since it could achieve a strictly higher payoff by offering $p - \varepsilon$ rather than p.

2. (No Aggregate Gaps) In equilibrium at least two firms offer each p between the minimum offer 0 and maximum $\overline{p}$.

Proof. If there was an interval where only one firm was active, this firm could not be optimizing. If there was an interval where no firms were active, then the lowest ranked firm active just above this interval could not be optimizing.

3. (Aggregate Offers) If $G_1, ..., G_n$ is an equilibrium, then $\sum_n g_n(p)$ is non-increasing in p.

Proof. Let J be the set of firms that make offers just below p. Then in order that these firms be willing to make offers just below p, it must be the case that for each $j \in J$, $\sum_{n \neq j} g_n(p^-) \geq \sum_{n \neq j} g_n(p^+)$. Summing over this inequality over all firms in J implies that $\sum_n g_n(p^-) \geq \sum_n g_n(p^+)$.

4. (No Gaps) Each equilibrium distribution G_n has interval support.

Proof. Suppose to the contrary that n makes offers just below p' and just above p'' but not in the interval (p', p''). Because n is optimizing, it must be the case that for any p in this interval:

$$\Delta_n \cdot \sum_{m \neq n} [G_m(p) - G_m(p')] \leq p - p'.$$

with equality when $p = p''$. Since n does not make offers in this interval, the above Lemma implies that $\sum_{m \neq n} g_m(p)$ is non-increasing in this interval. This implies that for any p in the gap, $\sum_{m \neq n} g_m(p) = \frac{1}{\Delta_n}$. Now, since n does make offers just above p'', it must also be the case that $\sum_{m \neq n} g_m(p) = \frac{1}{\Delta_n}$ for all p just above p''. But then $g_n(p) = 0$ just below p'' and $g_n(p) > 0$ just above p'' means that $\sum_m g_m(p)$ is equal to $\frac{1}{\Delta_n}$ just below p'' and is strictly greater than $\frac{1}{\Delta_n}$ just above p'', contradicting the previous Lemma.

5. (Monotonicity) If $G_1, ..., G_N$ is an equilibrium, and $n > m$, then $G_n(p) \leq G_m(p)$ for all p.

Proof. Established in the text.

6. (Price Distribution) Suppose that in equilibrium, firms $l, ..., m$ offer p. Then for each $n = l, ..., m$,

$$g_n(p) = \frac{1}{m - l} \sum_{k=l}^{m} \frac{1}{\Delta_k} - \frac{1}{\Delta_n}.$$

Proof. Established in the text.

7. (Supports) If $\overline{p}_m < \overline{p}$ is the highest offer made by some firm m, it must be lowest offer of some firm $n > m$.

Proof. Suppose that firms $m + 1, ..., n$ are active just above $\overline{p}_m$ and $m, ..., n$ are active just below $\overline{p}_m$. Then result 6 above implies that the aggregate offer rate just below $\overline{p}_m$ is:$\frac{1}{n-m} \sum_{k=m}^{n} \frac{1}{\Delta_k}$ and just above is $\frac{1}{n-m-1} \sum_{k=m+1}^{n} \frac{1}{\Delta_k}$. The latter is strictly greater contradicting the fact that the aggregate offer rate be non-increasing.

8. If m is the highest firm making offers on some interval, $l(m)$ is the least.

Proof. By monotonicity, the set of firms making offers is consecutive. If $l < l(m)$, then clearly $l, ..., m$ cannot be active since then $g_l(p) < 0$ on this interval — a contradiction. If instead $l, ..., m$ are active where $l > l(m)$, then for any p in this interval, $\sum_n g_n(p) > \frac{1}{\Delta_{l(m)}}$. Since the aggregate offer rate does not include $l(m)$ above $\overline{p}_{l(m)}$ and is non-increasing, it follows that $l(m)$ would do strictly better by offering a price in this interval or at the top of it than by offering $\overline{p}_{l(m)}$.

Proof of Proposition 1. We show that the conjectured strategies are the unique equilibrium. Suppose they are used by each firm $m \neq n$. Then on the interval where $g_n(p) > 0$, the aggregate density of opponent offers is $1/\Delta_n$ by construction. So n is indifferent between all offers in this region, the interval $(p_n, \overline{p}_n = p_{l^{-1}(n)}]$. If $p < p_n$, the aggregate density of opponent offers is strictly greater than $1/\Delta_n$, so offering p_n is strictly preferred to a lower offer. And if $p > \overline{p}_n$ the aggregate density of opponent offers is strictly less than $1/\Delta_n$, so offering such a high price cannot be optimal. So it is optimal for n to use the equilibrium strategy. In terms of uniqueness, it is quite easy to see that the maximum and minimum offers for each firm are uniquely pinned down as in the text. *Q.E.D.*

Appendix B: Omitted Proofs

Proof of Proposition 4. Let $G_1, ..., G_n$ be equilibrium strategies. Arguments similar to the above establish that these strategies have no atoms or gaps, and that if $n > m$, then $G_n(p) \leq G_m(p)$ for all p. The proof now follows the earlier argument for the linear case. By an analogous argument, $V_1 = \Pi_1 = v(1, 0)$. Now, note that

$$V_n - V_{n-1} = v(n, n) - v(n-1, n).$$

The difference in the Vickrey profits of n and $n-1$ is the difference in their value for worker n. Define $\hat{p}_n$ as the price that solves:

$$\Pi_n(\hat{p}_n) = v(n, n) - \hat{p}_n.$$

Such a price exists in the support of n's equilibrium strategy because at the lowest price n offers, n gets at best worker n (and potentially lower) and at the highest price n offers, n gets at worst worker n (and potentially higher), and because $\Pi_n(\cdot)$ will be continuous in p. Moreover, $n-1$ must also offer $\hat{p}_n$, and when it does, the distribution of worker quality it expects is worse than the distribution n expects in the sense of first order stochastic dominance. So,

$$\Pi_n - \Pi_{n-1} > v(n, n) - v(n-1, n) = V_n - V_{n-1},$$

completing the proof. *Q.E.D.*

Proof of Lemma 3. By the definition of $l(\cdot)$,

$$(\rho-1)\frac{1}{\Delta_{n-\rho}} \geq \frac{1}{\Delta_{n-\rho+1}} + ... + \frac{1}{\Delta_n} \geq (\rho-1)\frac{1}{\Delta_{n-\rho+1}}$$

Moreover, so long as $1/\Delta_n$ is convex in n:

$$\rho\frac{1}{2}\left(\frac{1}{\Delta_{n-\rho+1}} + \frac{1}{\Delta_n}\right) \geq \frac{1}{\Delta_{n-\rho+1}} + ... + \frac{1}{\Delta_n} \geq \rho\frac{1}{\Delta_{n-(\rho-1)/2}}.$$

Combining these inequalities and re-arranging:

$$\begin{aligned}\rho(\Delta_{n-(\rho-1)/2} - \Delta_{n-\rho}) &\geq \Delta_{n-(\rho-1)/2} \\ \rho(\Delta_n - \Delta_{n-\rho+1}) &\leq 2\Delta_n\end{aligned}$$

Substituting for Δ_n and re-arranging gives

$$\rho^2 + 2\rho - 1 \geq 2n \geq \rho^2 - \rho.$$

From here, it is easy to show that $\sqrt{2n}+1 > \rho > \sqrt{2n}-1$, so $\rho(n) \approx \sqrt{2n}$. *Q.E.D.*

Proof of Proposition 6. Note that any firm can ensure itself a profit of at least Δ_n by setting all its offers to zero. Now fix an equilibrium and suppose that in equilibrium worker n is assigned to firm k_n. Since k_1 must make at least Δ_{k_1}, $p_{k_1 1} = 0$. Moreover, to ensure that k_1 would not want to bid away worker 2, $p_{k_2 2} \geq \Delta_{k_1}$; by this same argument, $p_{k_{n+1}(n+1)} - p_{k_n n} \geq \Delta_{k_n}$. Since each firm n's profits must be at least Δ_n, it follows that $k_n = n$ for all n and matching is efficient. Moreover, in equilibrium $p_{11} = 0$ and more generally $p_{nn} \geq p_n^F$, so firms 2,...,n get no more than their Vickrey profits. In fact, $p_{nn} = p_n^F$. To see this, note that firm 2 can ensure itself its Vickrey profit by setting $p_{21} = 0$ and $p_{2k} = (k-1)\Delta_1$. With this offer, regardless of the offers of firms $n \neq 2$, any stable matching will involve Firm 2 matched to a higher worker than Firm 1. Thus firm 2 gets it Vickrey profit in equilibrium and $p_{22} = \Delta_1$. Given p_{11}, p_{22}, Firm 3 can ensure at least its Vickrey profit by setting $p_{31} = 0, p_{32} = \Delta_1$ and $p_{3n} = (n-2)\Delta_2 + \Delta_1$. So $p_{33} = \Delta_1 + \Delta_2$. Continuing the argument for firms $4, ..., n$ completes the proof. *Q.E.D.*

Proof of Proposition 8. Consider some subgame perfect equilibrium and suppose it results in worker n going to firm k_n at a price p_{k_n}. As firm k_1 can obtain its same match, worker 1, by decreasing its price to zero, $p_{k_1} = 0$ and k_1's equilibrium profits are Δ_1. Since k_1 also could have dropped out just above p_{k_2} and obtained at least worker 2, $p_{k_2} \geq \Delta_{k_1}$. By a similar argument, $p_{k_{n+1}} - p_{k_n} \geq \Delta_{k_n}$. From here, it is easy to see that because any firm n could ensure profits Δ_n by decreasing its price to zero, the only assignment consistent with equilibrium is that $k_n = n$. Finally, we have seen that $p_1 = 0$. Because firm 1 will not drop out above Δ_1 in order to secure worker 2, firm 2 must optimally set $p_2 = \Delta_1$. Continuing this argument shows that $p_n = \sum_{k<n} \Delta_k$ for each n. Thus if each firm n drops out at p_n^F, no firm has an incentive to deviate; also, no other strategies are consistent with equilibrium.[33] *Q.E.D.*

References

Avery, Christopher, Christine Jolls, Richard A. Posner, and Alvin E. Roth (2001), "The Market for Federal Judicial Law Clerks", University of Chicago Law Review, 68(3), Summer, 793-902.

Baye, Michael R., Dan Kovenock and Casper G. de Vries, "Rigging the Lobbying Process: An Application of the All-Pay Auction," *American Economic Review*, 83(1), March 1993, pp. 289–294.

Binmore, Ken and Paul Klemperer, "The Biggest Auction Ever: The Sale of the British 3G Telecom Licenses," *Economic Journal*, 112, March 2002, pp. C74–C96.

Chae, Sanders H., "Is the Match Illegal?" *The New England Journal of Medicine*, 348(4), January 23, 2003, pp. 352–356.

Crawford, Vincent P. and Elsie Marie Knoer, "Job Matching with Heterogeneous Firms and Workers," *Econometrica*, 49(2), March 1981, pp. 437–450.

Gale, David and Lloyd Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 1962, pp. 9–15.

[33] Strictly speaking, we have only shown that every pure strategy equilibrium must have a Vickrey outcome. But it is easy to check that there cannot be a mixed equilibrium.

Harsanyi, John, "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *Int. J. Game Theory*, 2, 1973, pp. 1–23.

Kamecke, Ulrich, "Wage Formation in a Centralized Matching Market," *International Economic Review*, 39(1), February 1998, pp. 33–53.

Kelso, Alexander S. and Vincent P. Crawford, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica,* 50(6), November 1982, pp. 1453–1504.

Koopmans, Tjalling C. and Martin Beckman, "Assignment Problems and the Location of Economic Activity," *Econometrica*, 25(1), January 1957, pp. 53–76.

Milgrom, Paul, "Matching with Contracts," Stanford University, March 2003.

Miller, Frances H. and Thomas L. Greaney, "The National Resident Matching Program and Antitrust Law," *Journal of the American Medical Association*, 289(7), February 19, 2003, pp. 913–918.

National Residency Matching Program, "Policies of the NRMP," July 2000, available at http://www.nrmp.org/res_match/policies.html.

Niederle, M. and A. Roth, "What are the Effects of a Match: Evidence from Internal Medicine," mimeo, 2003.

Roth, Alvin E., "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy,* 92, 1984, pp. 991–1016.

Roth, Alvin E. and Elliott Peranson, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89(4), September 1999, pp. 748–780.

Roth, Alvin E. and Xing, Xiaolin, "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, 84, September 1994, pp. 992–1044.