Matching and Price Competition

By JEREMY BULOW AND JONATHAN LEVIN*

We develop a model in which firms set impersonal 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. In light of our results, we discuss the performance of alternative institutions and the recent antitrust case against the National Resident Matching Program. (*JEL* D44, J41, L44)

A recent antitrust suit charged that the National Resident Matching Program (NRMP) 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 low, averaging around $40,000 per year, and compressed, and work hours are long, 80 hours a week in many programs.[2] While salary differentials are only one way in which residency programs might compete, the compression of salaries within programs, within specialties, and across fields is remarkable, compared to the variation in pay among more senior doctors.

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. There are two key elements: 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, and in which match surplus exhibits increasing differences so that higher-ranked hospitals place more value on attracting higher-ranked residents. Efficiency in this setting dictates an assortative matching of hospitals and residents. To model competition, we assume that hospitals make offers, with the hospital that makes the best offer 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

* Bulow: Graduate School of Business, Stanford University, Stanford, CA 94305-5015 (e-mail: jbulow@stanford.edu); Levin: Department of Economics, Stanford University, Stanford, CA 94305-6072 (e-mail: jdlevin@stanford.edu). We thank Susan Athey, Jeffrey Ely, Amy Finkelstein, Paul Klemperer, Paul Milgrom, Muriel Niederle, Alvin Roth, the editors, and an exceptional referee for helpful suggestions. Simon Board, Peter Coles, Michael D. Grubb, and Ron Siegel provided excellent research assistance. Levin gratefully acknowledges support from the Hoover Institution, the Alfred P. Sloan Foundation, and National Science Foundation grant SES-0349278.

[1] Alvin E. 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 Xiaolin Xing (1994), and Roth and Elliott Peranson (1999). The first theoretical study of matching algorithms is by David Gale and Lloyd 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 Vincent P. Crawford and Elsie Marie Knoer (1981) and Alexander S. Kelso and Crawford (1982). John Hatfield and Paul Milgrom (2005) unify and extend many of the central results in the literature.

[2] The antitrust case was dismissed in the summer of 2004 after Congress intervened on the side of the hospitals. Discussions of the case include Sanders H. Chae (2003) and Frances H. Miller and Thomas L. Greaney (2003).

bids. As in the auction analogy, we treat hospitals' offers as prices, but it is easy to reinterpret the offers as nonwage compensation or as investments hospitals make that provide utility to their residents. Regardless of the interpretation, the crucial feature is the "all-pay" element to 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 hospitals identical in their distributions for the value of obtaining the best residents, this system would lead to efficient matching and the same average wages as a system with personalized offers, 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, introduces a slight inefficiency and dampens competition.

In a competitive equilibrium, salaries adjust so that each hospital prefers to hire the resident who is its efficient match. The salary differential between two "adjacent" residents must lie between the extra amount that the lower and higher of the firms with which the residents will match would pay for the superior worker. So the surplus of the better hospital will exceed that of the lower hospital by an amount somewhere between the value of output the two firms obtain with the lesser worker and the value of output the two firms obtain with the superior worker.

With impersonal price setting followed by matching, there is a pricing equilibrium in mixed strategies. In this equilibrium, the expected surplus of the better hospital exceeds that of the lesser hospital by more than the difference in output with the superior worker. 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.[4] Therefore, the difference in the expected surplus of two adjacent hospitals reflects the maximum differential in a competitive 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 hospitals 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.[5] This accounts for the compression in the wage distribution.[6]

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

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

[6] 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.

While we frame the paper in terms of the residency match, we do not intend to cast negative aspersions on the NRMP. First, the match was designed for efficiency purposes; our model suggests that efficiency should be high, even relative to an idealized competitive equilibrium.[7] Second, we have chosen our assumptions for analytical simplicity and transparency, not as the most realistic possible model of the residency match. For instance, we abstract from a broad range of heterogeneity in preferences, such as locational preferences, that the NRMP process may be well-suited to address.

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 ranked order), employers may reasonably conclude that it is in their interest to pay the same wage to all incoming employees. It is this nondiscrimination feature that we regard as the focus of our study.[8]

The outline of the paper is as follows. In Section I we present a numerical example that illustrates the basic results in our model and an approach to solving for equilibrium. Section II briefly describes the model itself and Section III describes the pricing equilibrium. Sections IV and V describe the competitive equilibrium, the high profits, and the salary compression that the model produces. Section VI explains why the market has very good performance in terms of efficiency. Section VII 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 are due to the combination of the one wage restriction and the asymmetry of the firms. Section VIII discusses some extensions to the model and concludes.

I. A "Multiplication Game" Example

Consider a market with N firms (hospitals) 1, 2, ... , N, and a corresponding set of workers (residents), also labeled 1, 2, ... , N. Each firm is interested in matching with one worker and wants to maximize the value of its output less the amount it pays its worker. Firm n's output if it hires worker m has value $m \cdot n$, so we refer to this example as the "multiplication game." Workers are strictly interested in maximizing their wage and have a reservation wage of zero. All this information is common knowledge. Here we explicitly solve the example for $N = 3$.

In a competitive equilibrium, firms take the wages of each worker as given, each worker prefers to work, and each firm hires optimally at the given wages. A competitive equilibrium must have efficient matching, which here means that firm 1 will hire worker 1, firm 2 will hire worker 2, and firm 3 will hire worker 3. Worker 1's wage will be $p_1 \in [0, 1]$, so that firm and worker 1 each get positive surplus. Worker 2's wage must be high enough that firm 1 does not want to pay the extra amount needed to hire worker 2 over worker 1, but low enough that firm 2 does want to pay it. Therefore $p_2 - p_1$ must be greater than 1, firm 1's value for the incremental worker quality, but less than 2. So $p_2 \in [p_1 + 1, p_1 + 2]$. Similarly, worker 3's wage will be $p_3 \in [p_2 + 2, p_2 + 3]$. The lowest competitive wages are therefore 0, 1, and 3. At these wages, the hospitals receive their maximum competitive surpluses of 1, 3, and 6.[9] The total surplus is 14.

In our model of the match, each firm simultaneously offers a wage. These wages are observed by the workers prior to matching. At the matching stage, each firm ranks the workers in order of their ability and each worker ranks the firms in order of their offers. Given the align-

[7] In practice, some arguably similar markets without centralized matches perform quite poorly. See, for example, Christopher Avery et al. (2001) on the law clerkship market. It is also worth noting that in the context of medical fellowships, where a centralized matching process is used for some but not all subspecialties, salaries appear to be low across the board and do not seem to depend much on whether the market is centralized (Muriel Niederle and Roth, 2003). See also Ulrich Kamecke (1998) for an alternative model of low salaries in matching markets.

[8] The model also has some relevance for the study of multi-unit auctions and price discrimination in product markets (see Bulow and Levin, 2003, for discussion). Within firms, the model also suggests why tournaments might allow the most talented workers to obtain the most responsible jobs and a disproportionate share of the firm's surplus. It suggests that adopting such a pay mechanism might help the firm attract more of the most talented workers. We have not pursued this extension in any detail, however.

[9] 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.

ment of preferences, we do not specify an exact process for the match (it could, for instance, be an NRMP-style algorithm); we simply assume that worker 3 ends up with the firm that offered the highest wage, worker 2 with the firm that offered the second-highest wage, and worker 1 with the firm offering the lowest wage. What we will show in the context of this example is a set of results—subcompetitive 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.[10] If there were a pure strategy equilibrium, the middle bidder would offer only a fraction of a penny above the low bidder, and the high bidder only a fraction of a penny above the medium bidder. But then the lower two bidders would benefit by raising their bids to attract the top worker. This same logic implies that in a mixed equilibrium there will be no "atoms" in firm strategies, except possibly at a salary of zero. We also know that zero must be the lowest salary offered. A firm making the lowest offer is sure to obtain the lowest worker, so having the lowest offer be strictly positive is inconsistent with profit maximization.[11] What is more, every salary between zero and the maximum must be potentially offered by at least two firms. If no firms make offers on some range of salaries, no firm will make an offer at the top of the range, as it would prefer to make an offer at the bottom. Similarly, if only one firm makes offers on some range, it would do better to always make offers at the bottom of the range. Finally, we prove in the Appendix that each firm will randomize over an interval of prices.

Given these preliminaries, the only real issue is whether, at the top and bottom wages, only two or all three firms will randomize. It is easiest to start at the top. If only two firms randomize over the highest wages, it must be firms 2 and 3 because they care more about quality. If they are the only active firms, they must randomize with densities $q_2 = 1/3$ and $q_3 = 1/2$ in order that both are indifferent over making different offers in the range. Because $q_3 + q_2 = 5/6$, firm 1 gains less than one unit of quality for every extra dollar spent in this range, so it has no incentive to compete against firms 2 and 3 at the highest salary levels.[12]

Because $q_3 = 1/2 > q_2$, the range of wages over which 2 and 3 randomize must be from the maximum wage, call it $\bar{p}$, down to $\bar{p} - 2$. Within the range $[\bar{p} - 2, \bar{p}]$, firm 3 will exhaust its total bidding probability (a density of 1/2 times a range of 2). Firm 2, however, will exhaust only 2/3 of its probability (a density of 1/3 over a range of 2). This leaves firm 2 with a probability mass of 1/3 to employ over a range in which it competes with firm 1.

In the range where 1 and 2 compete $q_1 = 1/2$ and $q_2 = 1$ to make the other indifferent between making different offers. Given 2's density and its available probability, the length of the range must be 1/3. Since we know that the minimum wage must be zero, this means that 1 and 2 compete in a range of (0, 1/3] and therefore 2 and 3 compete in a range of [1/3, 7/3]. Competing with firm 2 over the range (0, 1/3] exhausts only $(1/2) \cdot (1/3) = 1/6$ of firm 1's probability mass, which implies that it will bid 0 with probability 5/6.

[10] An alternative approach would be to allow some randomness in hospital valuations, so that, for example, hospital j might value an extra unit of worker quality as an amount in the range $[j, j + \varepsilon]$, $\varepsilon \leq 1$, the exact amount known only to it. This would still allow for unambiguous rankings of hospitals and would not affect the expected competitive equilibrium prices, but would imply that each hospital would choose a pure strategy, bidding an amount that looked random to its competitors but was based on the hospital's private information about its value. In the limit, as $\varepsilon \to 0$, the equilibrium of the pure strategy game would be just the equilibrium of the mixed strategy game. The reason we take the mixed strategy approach is that it leads to simpler exposition and cleaner proofs. In the mixed strategy equilibrium, hospitals choose piecewise linear bidding strategies, while the pure strategy equilibrium is characterized by the piecewise solution of sets of nonlinear differential equations. Michael Grubb used Matlab to calculate the equilibrium in the pure strategy game with $\varepsilon = 1$, and with different numbers of firms. The numerical results were extremely similar to the mixed strategy equilibrium and virtually the same with larger numbers of firms.

[11] 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.

[12] Another way to say this is to note that if we solved for the q_1, q_2, and q_3 densities that would simultaneously make all firms indifferent across a range of prices, we would find $q_1 < 0$.

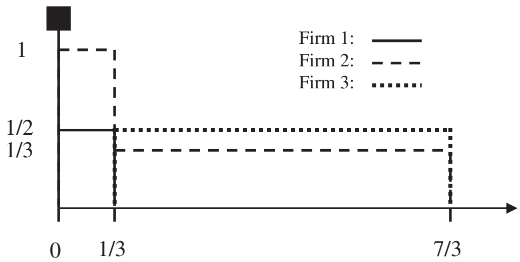

FIGURE 1. EQUILIBRIUM IN THE MULTIPLICATION GAME

The equilibrium in the three-firm example 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

Expected wages		Expected 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)

Note: Competitive outcomes in parentheses.

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. Results are similar with larger numbers of firms. In the multiplication game with $N = 16$, expected output in our game declines from a competitive level of 1,496 to 1,487, while wages decline by over 20 times that amount, from between 680 and 816 in a competitive equilibrium to 496. By comparison, collusion among the firms on a zero wage and random assignment of workers would yield 50 cents of efficiency loss for every dollar of wage decline.[13]

Not only do firms do better on average than in competitive equilibrium, every firm benefits. In the firm optimal competitive equilibrium, firm n hires worker n at a competitive salary p_n, leaving it with profit $n \cdot n - p_n$. In the match, firm n may not get worker n, but there is a price $\hat{p}_n$ in the support of firm n's offer distribution at which firm n gets worker n in expectation. Therefore firm n's equilibrium profit is $n \cdot n - \hat{p}_n$ and the difference between n's equilibrium profit and its competitive profit is $p_n - \hat{p}_n$. We found above that $(p_1, p_2, p_3) = (0, 1, 3)$, while in the match $(\hat{p}_1, \hat{p}_2, \hat{p}_3) = (0, 1/3, 7/3)$. So $p_n - \hat{p}_n \geq 0$ and every firm makes excess profit.

We want to explain why this is true and why the difference is larger for higher-ranked firms. Firm 1 is easy. In the match equilibrium, firm 1 gets worker 1 for certain by offering $\hat{p}_1 = 0$, which is also the price for worker 1 in competitive equilibrium. So $\hat{p}_1 = p_1$. Now the key feature of (the lowest) competitive prices is that each firm $n - 1$ is indifferent between paying p_n to hire worker n and paying p_{n-1} to hire worker $n - 1$. Thus $p_n - p_{n-1} = \Delta_{n-1} \cdot [n - (n - 1)] = \Delta_{n-1}$. In the match, the prices $\hat{p}_{n-1}$ and $\hat{p}_n$ are both in the support of firm $n - 1$'s offers, so firm $n - 1$ is indifferent between offering $\hat{p}_{n-1}$ and offering $\hat{p}_n$. But while paying $\hat{p}_{n-1}$ allows firm $n - 1$ to get worker $n - 1$ in expectation, paying $\hat{p}_n$ gets firm $n - 1$ an expected quality less than n. For instance, firm 1 expects worker quality 1/3 if it offers $\hat{p}_2 = 1/3$. The reason firm $n - 1$ expects lower quality than firm n conditional on offering $\hat{p}_n$ (or any other price) is that in equilibrium firm n makes higher offers than firm $n - 1$, so $n - 1$ faces tougher competition. The upshot is that $\hat{p}_n - \hat{p}_{n-1} = \Delta_{n-1} \cdot [\xi_{n-1,n} - (n - 1)]$, where $\xi_{n-1,n}$ is the expected worker quality that firm $n - 1$ gets by offering $\hat{p}_n$. But because $\xi_{n-1,n} \leq n$, it follows that $\hat{p}_n - \hat{p}_{n-1} \leq p_n - p_{n-1}$.

This logic demonstrates several points that will be true more generally. First, because $\hat{p}_1 = p_1$ and $\hat{p}_n - \hat{p}_{n-1} \leq p_n - p_{n-1}$, every firm makes at least as high a profit in the match as it does in a competitive equilibrium. Second, the price differentials cumulate, so that firm n makes all the additional profit of firm $n - 1$ plus a little more. Finally, the price difference for the top two firms is the same as in the competitive equilibrium. Because $\hat{p}_N$ must be the highest price offered in equilibrium, firm $N - 1$ and firm N both expect to get the highest-ranked worker by offering this price, so $\hat{p}_N - \hat{p}_{N-1} = p_N - p_{N-1}$. Apart from this, the differences are

[13] One might argue that any reasonable market mechanism would probably do very well compared to a random match, so that the real efficiency of an NRMP-type system is its ability to create such a market. Our point here is that the *ratio* of the amount redistributed from workers to firms to the amount of inefficiency is very large.

strict so long as the firms have different values for worker quality.

Another way to think of why the price a firm will pay for high quality will be low is to think about the equilibrium conditions that determine the price that workers can expect to receive for providing higher quality. In the three-firm example, firm 1 would not bid against 2 and 3 because the implicit price of increasing its quality was between 1 and 2 (1.2). In a 100-firm example, firm 100 competes only against firms 87 to 99 for the top worker, which means that the implicit price of quality for a nonbidder in the region is between 86 and 87. In a competitive equilibrium, we would observe firm 100 paying prices of 87 to 99 to beat out these same competitors.

The fact that higher-ranked firms make higher offers in equilibrium also leads to relatively efficient equilibrium matching. The key force here is that firms compete directly only against firms that are similar, so competition is localized. In the three-firm example, firm 3 always makes a higher offer than firm 1, so the best worker can fall only one place from his efficient assignment. With 100 firms, firm 100 always makes a higher offer than firms 1 through 86, so the best worker can fall no more than 13 places. In this case, firm 100 obtains an average quality of 96.75, and all other firms attain something even closer to their efficient quality. This suggests that larger markets will be very efficient, a point we develop in Section VI.

II. The Model

Our general model has N firms and N workers.[14] 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 $\Delta_N \geq \cdots \geq \Delta_1 \geq 0$. 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.

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 offered initially. 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. This assumption is inconsequential in the multiplication game, but one can construct examples where it is relevant for lower-ranked firms.[15]

Before solving for equilibrium, several points deserve emphasis. First, the model allows some or all of the firms to have identical value for worker quality. The most notable results, however, arise when there is some asymmetry between firms. Second, the multiplicative form of match surplus 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 each salary as encompassing job features such as responsibility, hours, and training. Indeed, one can view part of each firm's expenditure on its worker as an investment that makes the firm more attractive to workers.[16] It is also plausible that some firms are naturally more attractive to workers, and a simple version of this can be accommodated as well. If all workers derive additional utility u_n from working for firm n, we can let p_n denote the total utility firm n offers, of which it

[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.

[15] We thank Jeffrey Ely and Amy Finkelstein for bringing this point to our attention. Our working paper, Bulow and Levin (2003), provides a numerical example where firms may opt to pay nothing and not match ex post.

[16] In the medical residency case, the investment might be improved facilities, better senior physicians, more time for research or didactics, or more hospital support staff to relieve the administrative burden shared by residents.

pays $p_n - u_n$ as compensation. This modification will not change either the competitive equilibria or the pricing and matching equilibrium.[17] This being said, the model nonetheless is somewhat restrictive in assuming that workers have homogenous preferences rather than allowing for a broader degree of idiosyncrasy. We return to this point in the concluding section.

III. 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, we prove in the Appendix that in equilibrium each firm must randomize over an interval of prices.

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. 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. It follows that $1 - G_n(p) \geq 1 - G_m(p)$ for all p, establishing the claim.

FIGURE 2. PRICE SUPPORTS

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.

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 $l, \ldots, m$. Each of these firms must be just indifferent to changing its offer slightly away from p. So for each $n = l, \ldots, m$,

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

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

[17] It may be useful, however, for squaring the model with reality. For example, if u_n is viewed as a career benefit from being placed at firm n, the model could be consistent with a top-ranked university offering a relatively low salary and still attracting outstanding junior faculty.

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

Conveniently, the offer densities depend on the set of firms competing, but not on p.[18] Taking advantage of this, we define $q_n(l, m)$ to be firm n's offer density given firms $l, \ldots, m$ 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.

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

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

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. The algorithm generalizes the approach we used to solve the numerical example of the previous section. 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 $l(N), \ldots, N$ 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 $l(N-1), \ldots, N-1$ compete head to head; and each firm's offer density is given by $q_n(N-1)$.

More generally, suppose firms $m+1, \ldots, N$ "use up" their offer probability above p_{m+1}, but firm m does not. Then between p_m and p_{m+1}, firms $l(m), \ldots, m$ 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 $2, \ldots, N$. 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 $p_1, \ldots, p_{N+1}$ and $G_1(0)$ be defined as above, then for each firm n, and each nonempty 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.[19]

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

[19] 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).

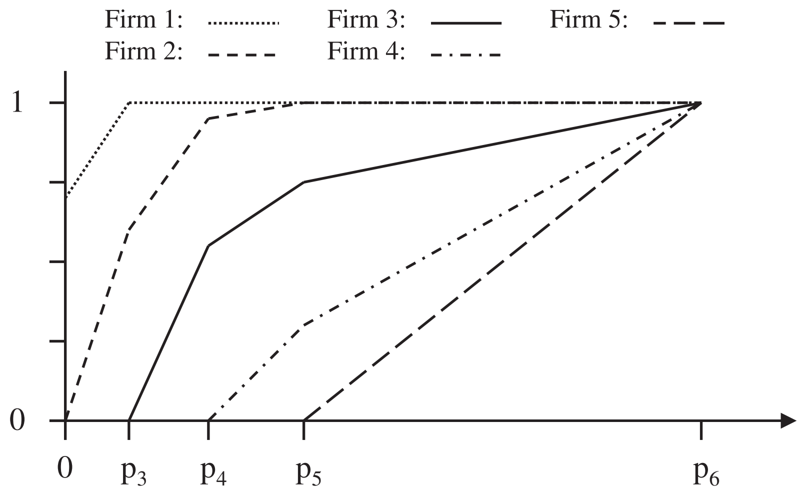

FIGURE 3. EQUILIBRIUM WITH FIVE FIRMS

IV. 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 and a pricing equilibrium with discriminatory offers.

A *competitive equilibrium* is a matching of firms and workers, and a salary for each worker, 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 whom it is matched to any other worker.

There is a range of competitive equilibria. Each involves efficient matching, but salaries vary. Extending our earlier example, firm 1 must pay worker 1 an individually rational salary $p_1 \in [0, \Delta_1]$. Also, the additional salary of worker n over worker $n-1$ must be such that firm $n-1$ prefers to hire worker $n-1$ while firm n prefers worker n. So 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 $p_n^F = \Sigma_{k<n} \Delta_k$, but not more than $p_n^W = \Sigma_{k \leq n} \Delta_k$. From the firms' perspective, the best competitive equilibrium has salaries $p_1^F, \ldots, p_N^F$; the worst has salaries $p_1^W, \ldots, p_N^W$.

The firm-optimal competitive equilibrium is the same outcome one would get from a Vickrey auction where the firms submit their values to a planner who efficiently allocates the workers and sets prices so that each firm's profit equals its marginal contribution to social surplus. For this reason, we refer to firm n's profit in the firm-preferred competitive equilibrium as its *Vickrey profit.*

The Vickrey or firm-optimal competitive outcome also arises as the equilibrium of the pricing followed by matching model if firms are allowed to make *discriminatory* offers. In this version of the model, each firm n makes a contingent offer to each worker, so firm n's offer is a vector $\mathbf{p}_n = (p_{n1}, \ldots, p_{nn})$. Matching follows taking these prices as fixed. Assuming the matching process always leads to a stable matching, Bulow and Levin (2003) show that any equilibrium in prices that does not involve weakly dominated strategies leads to the firm-optimal competitive equilibrium.

V. 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 profit in any competitive equilibrium. Second, worker salaries are lower in aggregate than their competitive 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.[20]

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 maximum competitive profit, or 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 $\Pi_n(\hat{p}_n) - \Pi_{n-1}(\hat{p}_n)$. Noting that offering $\hat{p}_n$ gets firm n an expected worker quality of n and firm $n - 1$ an expected worker quality of $n + G_n(\hat{p}_n) - G_{n-1}(\hat{p}_n)$, we have

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

Substituting and rearranging,

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

The first term is exactly $V_n - V_{n-1}$, the difference in the Vickrey profits of n and $n - 1$. The bracketed part of the second term is the difference between the worker quality that firms n and $n - 1$ expect conditional on offering $\hat{p}_n$; the entire second term is the value of this difference to firm $n - 1$. Because firm $n - 1$ makes lower offers than firm n in equilibrium, it expects lower worker quality than does firm n conditional on offering $\hat{p}_n$, so the second term is positive and is *strictly* positive provided that $\Delta_n > \Delta_{n-1}$ and $\hat{p} < \bar{p}$. Therefore, $\Pi_n - \Pi_{n-1} \geq V_n - V_{n-1}$, typically with strict inequality.

So equilibrium profit and Vickrey profit are the same for the lowest firm, but if $\Delta_2 > \Delta_1$, firm 2'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 difference between the Π_n and V_n increases for higher-ranked firms. For the very top two firms, the difference ceases to increase. That is, the extra profit that firm N makes over firm $N - 1$ in equilibrium is the same as the difference between their Vickrey profits. This occurs because $\hat{p}_N = \bar{p}$, and $G_{N-1}(\bar{p}) = G_N(\bar{p}) = 1$. We summarize as follows.

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

To reiterate, 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 than firm $n - 1$, it also expects, conditional on offering a given salary, to receive a better worker. This

[20] To see that firm n's expected worker quality is $1 + \sum_{k \neq n} G_k(p)$, note that n will get worker 1 if p is the lowest offer and will move up one worker for every competing firm that makes an offer below p. The competing offers are independent and the probability that k makes an offer below p is $G_k(p)$.

creates a larger profit differential between firms than in a competitive equilibrium where n's extra profit relative to $n - 1$ cannot exceed the difference in their values for worker n.

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 nonzero salary in equilibrium, which improves on her Vickrey salary of zero (though it may or may not be lower than her highest possible competitive salary). The best worker, however, not only expects a lower than competitive salary; in equilibrium she never receives an offer as high as her lowest possible competitive salary. We provide some quantitative examples in the next section to show that this difference is often quite 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 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.*

VI. 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. To expand on this point, we approximate both market efficiency and firm profits and show that the surplus redistribution is an order of magnitude larger than market inefficiency.

We focus for simplicity on the case where firm values are $\Delta_n = n/N^2$, so a match between firm n and worker m generates surplus $v(n, m) = (m \cdot n)/N^2$.[21] The normalization ensures that each match generates a surplus between 0 and 1, and that an efficient assignment always generates a per-firm surplus of approximately 1/3 provided N is sufficiently large.[22] This setup is equivalent to one where both firm values and worker qualities are uniformly located on the unit interval at locations $1/N, 2/N, \ldots, N/N$. In the latter interpretation, a larger market is one where the spacing of both firms and workers is denser.

We first address the extent to which competition is local. As part of our derivation of equilibrium, we characterized the highest- and lowest-ranked firms that make any given offer. We show in the Appendix that if firm n is the lowest firm to make some offer p, then the highest firm to offer p is approximately $n + \sqrt{2n}$. That is, the number of higher-ranked firms that 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 $|\rho(n) - \sqrt{2n}| < 1$.*

Now consider the expected efficiency loss in equilibrium relative to an efficient assortative match. If firm n makes a higher offer than firm $m > n$ in equilibrium, this creates a surplus loss of $\Delta_m - \Delta_n$ (because this "switch" moves n up one unit of worker quality, generating an addi-

[21] The results in this section generalize, with similar conclusions, provided $1/\Delta_n$ is convex in n.

[22] The last claim follows because $S(N) = 1/N \sum_{n=1}^{N} v(n, n) = 1/N \sum_{n=1}^{N} n^2/N^2 \approx 1/3$. Here and below, we employ the approximation $1/N \sum_{n=1}^{N} n^{\alpha}/N^{\alpha} \approx 1/(\alpha + 1)$.

tional Δ_n in surplus, but moves m down a corresponding unit, costing Δ_m). The expected per-firm inefficiency in equilibrium is the cost generated by all such displacements, weighted by the probability that they occur, and divided by the total number of firms. That is,

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

Our pool size result implies that the probability that firm n beats firm $n + k$ must be zero for any $k > \rho(n)$. Moreover, for any $0 < k < \rho(n)$, we know that in equilibrium firm $n + k$ must make stochastically higher offers than firm n. Therefore, 1/2 provides a (very) conservative upper bound on the probability that firm n beats firm $n + k$. Using these observations to substitute for the probabilities in the expression above, and noting that $\Delta_{n+k} - \Delta_n = k/N^2$, we have

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

That is, the per-firm inefficiency disappears at a fast $1/N$ rate. As we noted, the approximation is conservative; our numerical calculations suggest somewhat less inefficiency, though of the same order of magnitude.

Next consider the magnitude of redistribution. From the previous section we know that

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

How large is the difference between $G_m(\hat{p}_{m+1})$ and $G_{m+1}(\hat{p}_{m+1})$? Intuitively, in a large market, the support of firm m's offer distribution is very nearly $[p_m, p_{m+\rho(m)}]$, with $\hat{p}_m$ at the midpoint of this support. If we approximate the density of firm m's offer distribution as being linearly increasing, so $g_m(p) = (p - p_m)/(p_{m+\rho(m)} - p_m)$ (other functional forms such as a uniform density yield the same answer), we find that $G_m(\hat{p}_{m+1}) - G_{m+1}(\hat{p}_{m+1})$ is of the order $1/\rho(m)$. This intuitive argument is confirmed by numerical simulations. Thus

$$\Pi_n - V_n \approx \sum_{m=1}^{n-1}\Delta_m\cdot\frac{1}{\rho(m)} \approx \sum_{m=1}^{n-1}\frac{m}{N^2}\frac{1}{\sqrt{2m}} \approx \frac{\sqrt{2}n^{3/2}}{3N^2}.$$

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 least $1 - V_N = (N - 1)/(2N)$. In contrast, the highest salary that could *possibly* be offered to worker N in equilibrium is $1 - \Pi_N$. The difference between these salaries is $\Pi_N - V_N$, or approximately $\sqrt{2}/(3\sqrt{N})$.

For the average firm, the difference between equilibrium expected profit and competitive profit is

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

As the number of firms becomes large, the per-firm excess profits disappear, but at a $1/\sqrt{N}$ rate that is much slower than the rate at which the per-firm inefficiency disappears. The amount by which the average worker salary falls relative to the competitive benchmark, $E(N) + I(N)$, also disappears at the slower $1/\sqrt{N}$ rate.

So relative to competitive equilibrium, profits rise and salaries fall by an order of magnitude more than the change in surplus. In this sense, equilibrium generates far more redistribution than inefficiency. We summarize the discussion in the following proposition.

PROPOSITION 5: *Suppose that $\Delta_n = n/N^2$. As the market size increases, the per-firm inefficiency is of the order $1/N$, while the amounts by which the average firm profit exceeds the competitive level and the average worker salary fall short of the competitive level are of the order $1/\sqrt{N}$.*

As the market becomes very large, equilibrium becomes approximately efficient in the sense that virtually 100 percent of the total possible surplus is realized. Indeed, in the limiting case when firm values and worker qualities are continuously distributed on the unit interval, there is a pure strategy equilibrium in which each firm offers the salary it would pay in the firm-optimal competitive equilibrium, or Vickrey allocation.

It is perhaps useful to calibrate the magnitude of our results for reasonably sized markets. To this end, assume firm values are $\Delta_n = 100n/N^2$, where we multiply by 100 to avoid messy decimals. As the number of firms increases from 10 to 100 to 1,000, the per-firm efficient surplus stays roughly constant (38.5 with 10 firms, 33.8 with 10 firms, and 33.4 with 1,000 firms), as does the average firm Vickrey profit (22 with 10 firms, 17.2 with 100 firms, and 16.7 with 1,000 firms). The inefficiency from a match equilibrium falls rapidly from 0.4 per firm to 0.03 to 0.003, so that with 100 firms the inefficiency already is less than 1 percent of the possible surplus. Excess profits also fall, from 4.8 per firm to 2.2 to 0.8, but more slowly. With 100 firms, the average firm makes an equilibrium profit that is 13 percent more than its Vickrey profit.

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, more than twice the extra profit of the average firm. The top worker expects a wage of 49.5 in the Vickrey auction, and 43.9 in the match. This wage reduction is similar to what the top worker would suffer if the six top firms in the Vickrey auction, with values of 95 to 100, were replaced by firms that had values of zero. In absolute terms, her salary is depressed two and a half times the amount of the average worker.

VII. The Role of Nondiscrimination

It is tempting to attribute our results entirely to the fact that firms do not target their salary 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 $\{1, 2, \ldots, N\}$. 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,[23] 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 is what generates a departure from competitive

[23] In the match equilibrium, each firm optimally bids its expected, or average, Vickrey payment conditional on its signal. The averaging compresses the amount paid by the firms with ex post highest and lowest signals and so the salaries of the best and worst workers.

TABLE 2—ASYMMETRY AND NONDISCRIMINATION

	Symmetric firms		Asymmetric firms	
	Vickrey	Match	Vickrey	Match
Market surplus	S	S	S	$<S$
Firm profits	V	V	V	$\geqslant V$
Worker salaries	W	W	W	$\leqslant W$
Salary compression	No	Some	No	Yes

outcomes. This situation is summarized in Table 2.

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.[24]

VIII. Conclusion and Extensions

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

A natural question concerns the robustness of our conclusions to the simplifying assumptions of the model. One strong assumption is that the firms have perfect information about the values of their competitors. A variant of the model assumes that each firm n's marginal value for worker quality is an independent privately known draw from an atomless distribution on $[\underline{t}_n, \bar{t}_n]$. In this version of the model, there is a unique equilibrium in pure strategies. If the values of higher-ranked firms stochastically dominate those of lower-ranked firms, their offers will also be stochastically higher, as in the complete information model.

A special case of the private information model is where each firm n's value distribution is tightly concentrated around Δ_n. In this case, behavior in the pure strategy equilibrium is observationally equivalent to behavior in the mixed strategy equilibrium of our complete information model. The case where there is more uncertainty about competitors' values is harder to analyze analytically, but numerical computations confirm that our results remain valid in that setting.

A second extension that we have pursued with less success is to allow for more general preferences on the part of workers. It would be interesting to know if our results continue to hold in a setting where workers have idiosyncratic preferences over firms; so, for instance, worker m's value from matching with firm n is given by $p_n + \varepsilon_{mn}$, where ε_{mn} is an idiosyncratic component of utility that is privately known to worker m. Such a model adds an additional layer of realism, but also substantial complication because once firms are differentiated, they have some market power, making it nontrivial even to define an appropriate competitive benchmark. Finding richer tractable models of matching with noncooperative price setting is an avenue for future work.

APPENDIX A: DERIVATION OF THE EQUILIBRIUM

This Appendix fills in the details omitted from Section III. 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 opti-

[24] Michael Baye et al. (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.

mally 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 $\bar{p}$.

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 $G_1, \ldots, G_n$ is an equilibrium, then $\sum_n g_n(p)$ is nonincreasing in p.

PROOF:

Let J be the set of firms that make offers just below p. Then in order for these firms to 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 Lemma above implies that $\sum_{m \neq n} g_m(p)$ is nonincreasing in this interval. This implies that for any p in the gap, $\sum_{m \neq n} g_m(p) = 1/\Delta_n$. Now, since n does make offers just above p'', it must also be the case that $\sum_{m \neq n} g_m(p) = 1/\Delta_n$ 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 $1/\Delta_n$ just below p'' and is strictly greater than $1/\Delta_n$ just above p'', contradicting the previous Lemma.

5. (Monotonicity) If $G_1, \ldots, G_N$ 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 $l, \ldots, m$ offer p. Then for each $n = l, \ldots, m$,

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

PROOF:

Established in the text.

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

PROOF:

Suppose that firms $m + 1, \ldots, n$ are active just above $\bar{p}_m$, and $m, \ldots, n$ are active just below $\bar{p}_m$. Then result 6 above implies that the aggregate offer rate just below $\bar{p}_m$ is $1/(n - m) \sum_{k=m}^{n} (1/\Delta_k)$ and just above is $1/(n - m - 1) \sum_{k=m+1}^{n} (1/\Delta_k)$. The latter is strictly greater, contradicting the fact that the aggregate offer rate be nonincreasing.

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 $l, \ldots, m$ cannot be active since then $g_l(p) < 0$ on this interval—a contradiction. If, instead, $l, \ldots, m$ are active where $l > l(m)$, then for any p in this interval, $\sum_n g_n(p) > 1/\Delta_{l(m)}$. Since the aggregate offer rate does not include $l(m)$ above $\bar{p}_{l(m)}$ and is nonincreasing, 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 $\bar{p}_{l(m)}$.

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 $(p_n, \bar{p}_n = p_{l^{-1}(n)}]$. 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 $p > \bar{p}_n$, the aggregate density of opponent offers is strictly less than $1/\Delta_n$, so offering such a high price cannot be optimal. Thus, 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.

Appendix B: Omitted Proofs

PROOF OF PROPOSITION 4:

Let $G_1, \dots, G_n$ be equilibrium strategies. Arguments similar to those 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, $V_1 = \Pi_1 = v(1, 1)$. 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.

Let $h_n(m, p)$ denote the probability that firm n will match with worker m if it offers p and the other firms use their equilibrium offer strategies. With this notation, $\Pi_n(p) = \sum_m v(n, m)h_n(m, p) - p$. Now let $\hat{p}$ be the highest salary offered by firm $n - 1$ in equilibrium. If firm n were to offer $\hat{p}$, it would obtain at least worker n, so $h_n(m, p)$ puts no weight on $m < n$. We argue as follows:

$$\begin{aligned}
\Pi_n - \Pi_{n-1} &\geq \Pi_n(\hat{p}) - \Pi_{n-1}(\hat{p}) \\
&= \sum_m v(n, m)h_n(m, \hat{p}) \\
&\quad - \sum_m v(n - 1, m)h_{n-1}(m, \hat{p}) \\
&= \sum_m [v(n, m) - v(n - 1, m)]h_n(m, \hat{p}) \\
&\quad + \sum_m v(n - 1, m)[h_n(m, \hat{p}) - h_{n-1}(m, \hat{p})] \\
&\geq v(n, n) - v(n - 1, n) = V_n - V_{n-1}.
\end{aligned}$$

The first inequality holds because $\hat{p}$ is a potential offer for n, so $\Pi_n \geq \Pi_n(\hat{p})$. We then expand $\Pi_n(\hat{p}) - \Pi_{n-1}(\hat{p})$ into two terms: the difference in expected value to firm n and $n - 1$ from receiving workers according to $h_n(m, \hat{p})$ and the difference in expected value to firm $n - 1$ from receiving workers according to $h_n(m, \hat{p})$ and $h_{n-1}(m, \hat{p})$. The first term is at least $v(n, n) - v(n - 1, n)$ because of increasing differences and the fact that $h_n(m, \hat{p})$ puts weight only on $m \geq n$. The second term is nonnegative because $v(n - 1, m)$ is increasing in m and the distribution of workers that $n - 1$ expects from offering $\hat{p}$ is stochastically dominated by the distribution of workers that n expects.

PROOF OF LEMMA 3:

By the definition of $l(\cdot)$,

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

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

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

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

$$\geq \rho\,\frac{1}{\Delta_{n-(\rho-1)/2}}.$$

Combining these inequalities and rearranging:

$$\rho(\Delta_{n-(\rho-1)/2} - \Delta_{n-\rho}) \geq \Delta_{n-(\rho-1)/2}$$

$$\rho(\Delta_n - \Delta_{n-\rho+1}) \leq 2\Delta_n.$$

Substituting for Δ_n and rearranging gives

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

From here, it is easy to show that $\sqrt{2n} + 1 > \rho > \sqrt{2n} - 1$.

REFERENCES

Avery, Christopher; Jolls, Christine; Posner, Richard A. and Roth, Alvin E. "The Market for

Federal Judicial Law Clerks." *University of Chicago Law Review*, 2001, *68*(3), pp. 793–902.

Baye, Michael R.; Kovenock, Dan and de Vries, Casper G. "Rigging the Lobbying Process: An Application of the All-Pay Auction." *American Economic Review*, 1993, *83*(1), pp. 289–94.

Bulow, Jeremy and Levin, Jonathan. "Matching and Price Competition." Stanford Institute for Economic Policy Working Paper No. 03-06, 2003.

Chae, Sanders H. "Is the Match Illegal?" *The New England Journal of Medicine*, 2003, *348*(4), pp. 352–56.

Crawford, Vincent P. and Knoer, Elsie Marie. "Job Matching with Heterogeneous Firms and Workers." *Econometrica*, 1981, *49*(2), pp. 437–50.

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

Hatfield, John W. and Milgrom, Paul. "Matching with Contracts." *American Economic Review*, 2005, *95*(4), pp. 913–35.

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

Kelso, Alexander S., Jr. and Crawford, Vincent P. "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, 1982, *50*(6), pp. 1483–504.

Miller, Frances H. and Greaney, Thomas L. "The National Resident Matching Program and Antitrust Law." *Journal of the American Medical Association*, 2003, *289*(7), pp. 913–18.

Niederle, Muriel and Roth, Alvin E. "Relationship between Wages and Presence of a Match in Medical Fellowships." *Journal of American Medical Association*, 2003, *290*(9), pp. 1153–54.

Roth, Alvin E. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 1984, *92*(6), pp. 991–1016.

Roth, Alvin E. and Peranson, Elliott. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 1999, *89*(4), pp. 748–80.

Roth, Alvin E. and Xing, Xiaolin. "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions." *American Economic Review*, 1994, *84*(4), pp. 992–1044.