The Strategic Decentralization of Recruiting*

Yi Chen† Thomas Jungbauer‡

May 25, 2021

Abstract

We propose a model of strategic delegation in professional labor markets with heterogeneous workers. A big firm, competing with fringe firms, decides whether to exercise its market power to suppress wages. Alternatively, it may choose to delegate hiring to agents (divisions), thereby committing to bid more fiercely for workers. This reduces the incentive for fringe firms of going toe-to-toe with the big firm. In equilibrium, the big firm chooses to delegate unless (a) it is too large, (b) not productive enough, or (c) too productive. While a more productive big firm delegates more often, the optimal number of agents it delegates to decreases in the big firm's productivity. The presence of big firms does not substantially lower social welfare, unless its size exceeds the tipping point beyond which it chooses not to delegate. The introduction of a minimum wage in a professional labor market induces the big firm to delegate more aggressively, increasing match quality. Thus, social welfare may increase despite a drop in employment. (JEL C78, D44, J31, J42)

Keywords: strategic delegation, market power, decentralization, recruiting, minimum wage

*We are particularly grateful to Daniel Barron, Michael Baye, Jasmin Droege, Florian Ederer, George Georgiadis, Mikhail Golosov, Faruk Gul, Christian Haefke, Marina Halac, Justin Johnson, Bernhard Kasberger, Philipp Kircher, Fei Li, George Mailath, Margaret Meyer, Ignacio Monzon, Nicola Persico, Michael Powell, Marcel Preuss, Michael Ransom, Klaus Ritzberger, Larry Samuelson, Mark Satterthwaite, Christian Schmid, Adam Szeidl, Rakesh Vohra, Michael Waldman and Ning Yu for valuable discussions and comments on earlier drafts of the paper. In addition we are indebted to participants of various seminars and in particular those of the 12th World Congress of the Econometric Society, and the 30th Annual Congress of the European Economic Association. We would also like to thank Yujie Feng for excellent research assistance. All remaining errors are our own.

†Assistant Professor of Strategy and Business Economics at the Samuel Curtis Johnson Graduate School of Management at Cornell University, 114 East Avenue, Ithaca, NY-14853, yi.chen@cornell.edu;

‡Assistant Professor of Strategy and Business Economics at the Samuel Curtis Johnson Graduate School of Management at Cornell University, 114 East Avenue, Ithaca, NY-14853, jungbauer@cornell.edu;

1 Introduction

The dawn of a new age of big firms has ignited controversy about the effects of these behemoths on the economy. While much of the discussion revolves around the potential dangers of imperfect competition in downstream markets—such as overt pricing power or a lack of incentives to innovate—these companies equally wield tremendous power on input markets. The US business magazine Fast Company even hypothesizes that monopsony, not monopoly, power of these superfirms poses the greater threat to society and, ultimately, to companies themselves by means of antitrust law suits and other modes of backlash (Fleishman, 2019). Labor markets are among the foremost examples in which firms hold and exercise market power when acquiring inputs (Bhaskar et al., 2002; Ashenfelter et al., 2010). In fact, Azar et al. (2018) estimate that about 60% of labor markets in the US are highly concentrated, accounting for more than 20% of the overall workforce.

Big firms, however, do not always exercise their market power. In fact, they often voluntarily relinquish their market power when recruiting. Instead of leveraging their size by allocating decision rights to a central authority such as a human resources department, they decentralize and delegate hiring to independent divisions. When doing so, business units within the same company often end up competing for the same crop of workers. Recruiting platforms and business magazines even provide recommendations to job-seekers how to successfully apply for several jobs at the same company (Lindzon, 2020), and to firms how to handle competition for workers among divisions (McNeal, 2001). The recruiting platform Yello, for instance, states:

> [M]ultiple divisions of one company may attend the same career fairs and compete for the same talent. Operating independently, every business unit sends their own representatives to recruit candidates for division-specific job openings. This can result in ... competition amongst co-workers looking to recruit the same talent. (Yello, 2019)

A survey of leading retail, hospitality and restaurant chain managers found that 57% of companies managed hiring at the division level (AON Hewitt, 2010). These observations give rise to our first main question. Why do big firms decentralize recruiting, and thereby surrender market power?

Conventional wisdom suggests that bigger firms should delegate more often than smaller ones. Large corporations are impersonal and slow to act, and often struggle to leverage their employees' expertise (Steakley, 2013), while decentralized recruiting accelerates the pace of decision making with personal touch and local expertise (Gregory, 2019). However, a global survey of firms from 26 industries and 43 countries conducted by Mercer tells a different story. It shows that the relationship between firm size and decentralization is negative over a significant range (Mercer, 2017). The share of companies that relies on decentralized recruiting steadily decreases from 25% among those with 100-249 employees to a much lower 8% of companies with just below $5,000$ employees.[1] This leads to our second main question. Why do larger firms decentralize recruiting less often?[2]

In order to address these questions, we introduce a three-stage model of recruiting. A big firm, commanding a positive measure of jobs, competes against a continuum of fringe firms, each posting one job, for heterogeneous workers. We refer to this scenario as a quasi-monopsony.[3] In stage one, the big firm decides whether to delegate recruiting or not, and if so, to how many agents (divisions). In stage two, each agent posts a schedule of wages, one per job, while simultaneously each fringe firm posts a single wage. These agents—while put in action by the same big firm—act in their own best interest. Ultimately, in stage three, a reduced-form matching process assigns more able workers to jobs paying higher wages.[4] Production is supermodular in worker skill and firm productivity.[5] We invoke backwards induction and identify optimal delegation as well as wage bidding patterns. Beyond our two main questions, our model allows us to weigh in on two policy debates. First, to which extent are concerns about powerful firms in labor markets warranted, i.e., what are the effects of labor market power on total surplus and its distribution? And second, what are the consequences of introducing a minimum wage such as a union wage in a professional labor market when big firms command market power?

The key tension in our model pits the big firm's *market power* versus its *commitment power*. On the one hand, the big firm enjoys labor market power due to its size.[6] When

[1]The relationship reverses for companies with more than $5,000$ employees. Companies beyond this size are mostly multi-national enterprises whose divisions often do not even hire in the same geographic regions.

[2]A possible explanation for this phenomenon is that small firms do not have HR personnel and are therefore considered decentralized. This seems unlikely, however, for two distinct reasons. First, the Society for Human Resource Management for example recommends recruiting through HR specialists when companies reach 15 to 25 employees (Lennon, 2019). Second, firms below 100 employees actually decentralize less often than larger firms (Mercer, 2017). We suspect these firms to be too small for market power to play a role for the recruiting decision.

[3]See Autor et al. (2019) for an argument that developments such as the rise in platform competition as well as advances in information technology provoke such "winner takes most" scenarios giving rise to superstar firms. We later show that the main findings of our paper generalize to markets with multiple firms that hold market power.

[4]It is well known that wage setting is a common practice in plenty of professional labor markets (Bulow and Levin, 2006). Our reduced-form approach reflects a strategic wage setting model in which firms first set wages for their jobs. Then workers apply for jobs, and each firm indicates which worker it wants to hire.

[5]This is a common assumption in the labor matching literature. Recent results of Song et al. (2019) support the assumption of supermodularity between workers and firms in production as they find that an increased segregation of worker across firms—with respect to their ability levels—has led to an increase of overall wage variation.

[6]The market power of the big firm is determined by its size denoting the number of its job openings,

the big firm lowers the wages for some of its job openings, a positive externality arises which lowers the competitive pressure on its remaining vacancies. In other words, these job openings continue to acquire the same workers at lower wages, or enjoy more able workers at the same wages. This market power effect increases in the firm's size. On the other hand, the big firm benefits from committing to bid higher wages through delegation. When multiple agents for the big firm compete among themselves, wages increase. Fringe firms, when observing the big firm delegate, foresee this fierce level of competition and adjust their bids downwards to target less able workers. This, in turn, eases pressure on the big firm's agents. In other words, the big firm acts like a Stackelberg leader to crowd out competition for the best workers, acquiring superior ability without overly increasing its wages. Naturally, this commitment effect increases in the number of agents the big firm delegates to.

Not only does our model explain why big firms delegate recruiting, but in line with the real-world delegation pattern described above, it predicts that the big firm's tendency to decentralize decreases in its size. Intuitively, when the big firm commands a significant fraction of overall vacancies, the diminished ability to dampen wages due to its loss of market power outweighs any potential gains from commitment through delegation. Instead, a more moderately sized big firm does not enjoy the perks of significant market power in the first place, and hence considers delegation to be more attractive.

The big firm's relative productivity as compared to fringe firms plays a more subtle role for delegation. First, the big firm never delegates hiring if it is less productive than fringe firms. In this scenario, its productivity does not warrant an aggressive wage schedule, and the big firm hires the least able workers in the market at minimal wages. Second, the big firm opts against delegation as well whenever it is highly productive relative to fringe firms. In this case, the big firm acquires superior worker ability even without delegation. As such, there is no point in sacrificing market power, thereby raising wages. Therefore, the big firm delegates to multiple agents only if it is moderately productive, i.e., more than fringe firms but not by too much. As a result, the equilibrium delegation pattern is non-monotone in the big firm's productivity. Finally, due to the trade-off between market and commitment power, the big firm's benefits from delegation are maximized when it is just more productive than its fringe competitors. That is to say, a moderately productive big firm delegates less aggressively (to fewer agents) the more productive it is. The size threshold above which a moderate big firm seizes to delegate, on the other hand, increases in the big firm's productivity because delegating itself becomes more fruitful.

which we consider to be fixed throughout the paper. This is in line with many professional labor markets such as the markets for doctors, nurses, lawyers, professors or teachers, in which the need for personnel determines the number of vacancies.

These results generate welfare implications. To begin with, we claim that decentralization always weakly increases market efficiency. In particular, when the big firm is more productive than its fringe competitors, the competition among the delegated agents raises the overall wages offered by the big firm, and hence attracts more skilled workers to it. Due to supermodularity, the matching quality is improved. In contrast, when the big firm is less productive than the fringe firms, it always hires the least able workers and hence does not affect market efficiency, regardless of its delegation decision.

Therefore, a moderately productive and modestly sized big firm, which chooses to delegate in equilibrium, poses little threat to social welfare. Positively assortative matching (PAM) is achieved or almost achieved in that case. However, if its size exceeds the threshold beyond which it does not delegate, total surplus plummets due to a poor matching quality. Moreover, a big firm that is highly productive outbids most fringe firms even in the absence of decentralization, resulting in almost PAM. In summary, notable welfare loss arises only if a very large firm is moderately productive. However, even in this case, potential mismatch is limited due the large fraction of workers hired by the big firm.

In addition, we study the effects of a minimum wage, such as a union wage, in a professional labor market when firms command market power and workers are heterogeneous. In line with standard economics textbook intuition, we find that—when binding—a minimum wage unambiguously increases unemployment. This result is not immediate when market power exists (see e.g., Manning, 2006). More importantly, we unveil additional effects of a minimum wage that emerge in response to the endogenous delegation choice of the big firm in equilibrium. A minimum wage hurts the big firm only if it hires the least able workers. In this case, delegation to multiple agents—resulting in higher wages in exchange for more able workers—becomes relatively more attractive. This pushes the big firm towards delegation, thereby improving the matching between firms and workers. As a result, total surplus may increase despite the smaller number of matches, that is higher unemployment, in the market. It follows, that when firms command market power over heterogeneous workers, a minimum wage may not be a viable solution to address distribution concerns when big firms command significant labor market power. Moreover, due to the big firm's delegation and aggressive bidding, the (less productive) fringe firms end up hiring the least able workers, and unemployment may increase even further.

Finally, we support the validity of our findings by presenting three extensions of our model. First, we focus on a market with two big firms, a so-called duopsony. Not only do we find that our general intuition from the quasi-monopsony persists, but also that the less productive of the two firms has an incentive to hypothetically delegate to infinitely many agents. In other words, our duopsony model reinforces a quasi-monopsony market structure and shows that market power—not a single big firm—is the driving force behind our results. Second, we investigate the case in which each agent feels obliged to offer a single wage for all its jobs, motivated by the non-discrimination policy observed in many real-world entry-level job markets. We show that such a restriction does not alter the results presented in this paper. Third, we generalize the production function and show that the main intuition established throughout the paper is quite robust.

While real-world market observations motivate our main application, it is not the goal of this paper to generate the most realistic representation of labor markets. We abstract from matching frictions and uncertainty to focus on the tension between market and commitment power through delegation. The insights about strategic delegation in our model pertain to auctions and other general bidding environments for heterogeneous goods when players command market power. Key applications are narrow high-skill labor markets (doctors, lawyers, etc.), labor markets with geographical boundaries such as the markets for nurses and school teachers, but also multi-unit heterogeneous goods auctions such as those for communication frequencies. Our results on a minimum wage can be translated to a price floor a more general bidding environments.

The remainder of Section 1 embeds our paper into the relevant literature. Section 2 introduces the general model while Section 3 presents the equilibrium analysis and establishes our main findings. Section 4 analyzes the introduction of a minimum wage under a quasi-monopsony market structure while Section 5 introduces a second big firm as well as uniform wages, and extends our findings to general production functions. Finally, Section 6 concludes.

Related Literature This paper contributes to several literatures such as the streams about strategic delegation and divisionalization as well as the treatment of market power in labor markets.

(*Delegation*) It has been known since the seminal work of Schelling (1960) that a principal may prefer an agent acting in her stead—not despite, but because of the agent's objectives not coinciding with the the principal's. The classic paper of Fershtman and Judd (1987), for example, provides an analysis of how to incentivize managers under competition. Models of strategic delegation to principals cover an array of different scenarios.[7] Fauli-Oller and Giralt (1995) and Galunic and Eisenhardt (1996) both discuss that the delegation of decisions from a corporate entity to its divisions increases aggressiveness precisely if these divisions experience negative interdivisional spillovers. In our model—the first to apply this idea to labor markets and to describe the delegation pattern—negative interdivisional spillover arises

[7] For an excellent and comprehensive overview consult Sengul et al. (2012).

if one division hires more able workers than others.[8]

(*Divisionalization*) An idea closely related to our paper is divisionalization in output markets. Salant et al. (1983) observe that mergers, i.e., the mirror image of divisonalization, when setting quantities, lead to losses. Schwartz and Thompson (1986) and Veendorp (1991) show that splitting into competing divisions may prevent competitor entry. Polasky (1992) connects divisionalization with the desire of firms to be a Stackelberg leader. Finally, Baye et al. (1996a) and Baye et al. (1996b) analyze optimal splitting patterns in these games and show that corporate gains can exceed the sum of divisional losses from increased competition. A common denominator of these models is that, in quantity-setting games, merging firms (insiders) always lose while other firms (outsiders) gain. Also in standard price setting games, similarly, outsiders always gain more than insiders from merging (Creane and Davidson, 2004). In other words, in these games firms would like to split infinitely often if the cost of doing so was infinitesimally small. This is not the case in our model. Due to the tension of market and commitment power, when facing a set of fringe competitors, a big firm taking on a competitive fringe optimally wants to delegate to a finite number of agents.

(*Market Power*) A widespread phenomenon in labor markets is market power (Bhaskar et al., 2002; Ashenfelter et al., 2010). Specifically monopsony power has generated a variety of theoretical contributions and surveys (Boal and Ransom, 1997; Bhaskar and To, 1999; Bhaskar et al., 2002; Manning, 2006; Ashenfelter et al., 2010), as well as studies of specific examples such as the markets for teachers (Landon and Baird, 1971; Ransom and Sims, 2010), nurses (Buerhaus and Staiger, 1996; Staiger et al., 2010), academics (Ransom, 1993) and even professional athletes (Kahn, 2000). This literature has mainly focused on social welfare measured via unemployment and the effects of labor market policies in models with homogeneous workers (Manning, 2006; Bhaskar and To, 1999). Galenianos et al. (2011) is an exception in so far as it provides a study of the effects of market power of heterogeneous firms in the labor market on overall welfare and its distribution. Instead of focusing on dominant firms, however, they look at firms that are interested in hiring a single worker and the presence of market power is ensured by the finiteness of the market. Prager and Schmitt (2020) and Arnold (2020) provide empirical evidence that larger firms pay lower wages.

(*Properties*) Some other papers that share key elements with our approach are Burdett et al. (2001), in which the quality of the match depends on the distribution of buyers and sellers, and van den Berg and Ridder (1998) who argue that wage offer distributions are endogenous. Azevedo and Leshno (2016) use a framework that matches a discrete number of

[8]Birkinshaw (2002) argues that companies may draw value from internal competition among its subsidiaries or divisions by improving firm performance, for example through identifying superior processes or technologies. According to Birkinshaw et al. (2005), internal competition describes the competition between units within an organization for suppliers or internal customers.

firms with a continuum of workers but focus on supply and demand effects in large markets, mostly ignoring transfers.[9] Finally, Board et al. (2020) build a model of the labor market that—just as ours—clears from the top. Their main interest, however, is not the effect of market power as they match a continuum of firms and workers and focus on the effects of recruiting and screening skills.

(*Auctions*) To simplify the matching stage, we build on a stylized model of strategic wage setting related to the approach in Jungbauer (2021) generalizing Bulow and Levin (2006). By applying their logic to a continuum of workers we are able to establish pure strategy equilibria. Finally, since our model of the bidding stage is technically a multi-unit auction our insights may be relevant for auctions as well as contests more generally. The seminal paper for inefficiencies in multi-unit auctions is Ausubel et al. (2014). In their paper welfare loss arises due to demand reduction as a consequence of differential bid shading. Inefficient demand reduction may also arise in multi-unit uniform-price auctions (Levin, 2004). See Schwenen (2015)'s description of the New York capacity market for an example of these inefficiencies in real-world auctions.

2 The Model

In this section we introduce a three-stage game in the context of a professional labor market with heterogeneous workers. The primitives of the game are laid out in Subsection 2.1, the timeline of the three stages is illustrated in Subsection 2.2, and the solution concept is defined in Subsection 2.3.

2.1 Workers, Firms and Jobs

The two-sided market is characterized by workers and firms. One side is a continuum of heterogeneous workers with measure 1, indexed by $i \in [0, 1]$. On the other side, there is one *big firm* with a multitude of job openings, and a continuum of so-called *fringe firms*, each posting a single job. We define $s \in (0, 1)$ to be the *size* of the big firm, i.e., the total measure of its job openings. On the other hand, each fringe firm is atomless, while in aggregation their total measure is normalized to $1 - s$. The matching between workers and jobs is one-to-one, so that the market is balanced.[10]

[9] Another paper matching a discrete number of firms with a continuum of workers is Azevedo (2014), who shows firms' incentives to reduce capacity to obtain more able workers. The capacity choice literature does not split firms but reduces the size of a single entity. See also Sonmez (1997), Konishi and Unver (2006) and Kojima (2006).

[10] The assumption of a balanced market is convenient for exposition but inessential for our main results.

All jobs at fringe firms are assumed to be identical, and so are all jobs at the big firm. Worker i generates output $R_1(i)$ when matched with a job at the big firm, and $R_0(i)$ when matched with any fringe firm. We refer to $R_1(\cdot)$ and $R_0(\cdot)$ *output functions*. Furthermore, we assume that workers can be unambiguously sorted according to some "general ability," such that $R_j' > 0$ for $j = 0, 1$. For technical simplicity, we normalize $R_j(0) = 0$ and require R_j'' to exist. Moreover, we assume these output functions are assumed to be supermodular.

Definition 1 *Output functions R_0 and R_1 satisfy supermodularity if for any $i > i'$:*

$$\max\{R_0(i), R_1(i)\} + \min\{R_0(i'), R_1(i')\} \geqslant \max\{R_0(i'), R_1(i')\} + \min\{R_0(i), R_1(i)\}.$$

Supermodularity requires the difference in output between two jobs to increase with worker index. Under this assumption, the first best allocation of workers to jobs is positive assortative matching (PAM). We rule out externalities among workers at the same firm, such that the output is additive among worker-job pairs.[11] Finally, a vacant job generates output zero.

Throughout the main body of this paper we assume linear output functions for ease of exposition, i.e., $R_0(i) = i$ and $R_1(i) = \mu i$, where the parameter $\mu > 0$ denotes the relative productivity of the big firm compared to a fringe firm. Subsection 5.3 discusses non-linear output functions and generalizes the main results.

2.2 Delegation, Bidding and Matching

The game has three stages, the delegation, the bidding and the matching stage. In the first, the delegation stage, the big firm decides whether to delegate hiring to agents. If if does, it chooses an integer $M \geqslant 1$, representing the number of agents it delegates to. The special case of $M = 1$ corresponds to the big firm not delegating, i.e., it enters the second stage as a single entity. For ease of exposition, we focus on symmetric delegation, such that each agent $m \in \{1, 2, ..., M\}$ controls the same measure of job openings $s_m = \frac{s}{M}$ from the big firm. The delegation decision is publicly observable and committed. This assumption is supported by the evidence of divisions competing for workers discussed in Section 1, as well as by the fact that hiring typically constitutes a repeated game. Naturally, a firm would lose its credibility when undermining the authority of its divisions in the labor market.

In the second, the bidding stage, the M agents and a measure of $1 - s$ fringe firms simultaneously bid for workers. A fringe firm posts a single wage $w \geqslant 0$. An agent who is the m-th delegate from the big firm, posts a wage schedule $w_m(\cdot)$, where $w_m(\theta) \geqslant 0$ is

[11] In this model we assume away the heterogeneity caused by worker-job specific characteristics. We also disregard the possibility of teamwork to highlight the main insight. For worker allocation models that focus on complementarity, see the o-ring literature started by Kremer (1993).

the θ-th percentile wage among the set of jobs the agent is responsible for. In real-world markets, firms or firm divisions sometimes post a single wage for identical jobs to prevent tension among its workforce. We show in Subsection 5.2 that limiting agents to post equal wages for their respective set of job openings has no bearing on our main results.

Finally, the third, the matching stage, is mechanical. More able workers are matched with jobs paying higher wages. Formally, if worker i is matched with a job with wage w and worker i' is matched with a job with wage w', then $i \geqslant i'$ implies $w \geqslant w'$.[12] This is a reduced form outcome of a matching mechanism where workers uniformly seek higher wages while jobs always prefer more able workers at any given posted wage. The posting of impersonal wages, unconditional on worker ability, is customary in many professional labor markets, especially at the entry-level (see e.g., Niederle et al., 2006).[13]

2.3 Payoffs, Strategies, and the Solution Concept

After the matching, profits are realized. A worker-job pair (i, j) that pays wage w yields a net profit of

$$\pi = R_j(i) - w, \tag{1}$$

depending on whether the job is from a fringe firm ($j = 0$) or from the big firm ($j = 1$). A worker receives utility w when matched with a job paying w, and an outside option normalized to zero when unmatched.

A pure strategy of the big firm is an integer $M \geqslant 1$, i.e., the number of agents it delegates to. A pure strategy of a fringe firm is a mapping from the number of agents M into a non-negative wage $w \geqslant 0$. A pure strategy of an agent m is a mapping from the number of agents M into a non-negative increasing function $w_m(\cdot)$ defined on $[0, 1]$.

A subgame perfect equilibrium (SPE) of the game is a profile of strategies of all agents and fringe firms such that: (a) in the second stage, a fringe firm maximizes the profit of its single job, while each agent maximizes the total profits over the set of jobs the agent controls, and (b) in the first stage, the big firm maximizes the aggregate profit from all of its agents. Throughout the paper we focus on SPE in pure strategies, except when dealing with uniform wages in Subsection 5.2, in which we allow for randomization.

[12] If multiple agents bid identical wages assume a fair lottery to ensue.

[13] This reduced-form matching outcome may also model a market in which the allocation of workers to firms is such as the National Resident Matching Program. If firms/agents and workers are solely motivated by monetary considerations, then every firm wants to hire more able workers while all workers want to work for higher wages. Under such aligned preferences, deferred acceptance, serial dictatorship or top trading cycles are among the central algorithms that lead to the same outcome.

3 Equilibrium Analysis

We use backward induction to solve the game. As stage three is mechanical, we begin by analyzing the second stage, taking the matching outcome of the third stage as given.

3.1 Stage Two: the Bidding Game

Suppose the big firm has delegated hiring in the first stage to $M \geqslant 1$ agents, with $M = 1$ interpreted as not delegating. As delegation is publicly observable, the bidding stage starts a proper subgame. In order to fully characterize the Nash equilibria of this bidding subgame, we first establish some regularity properties of equilibrium behavior.

Lemma 1 (Regularity)
In any Nash equilibrium of the bidding subgame: (i) no set of positive wages with a positive measure is offered by a single agent only; (ii) given any $\varepsilon > 0$, there is a positive measure of jobs offering wages below ε; (iii) the support of all wages offered in the market is a closed interval; (iv) no wage strictly above zero is offered by a positive measure of jobs; and (v) a positive measure of jobs offers a wage of zero only if almost all these jobs are offered by one agent.

Intuitively, Lemma 1 establishes that the set of all wages offered in equilibrium is an interval starting at zero with no positive measure holes. Moreover, no positive wage is offered by a positive measure of jobs. This equilibrium behavior results in the following overall distribution of wages in the market.

Corollary 1 *Let F be the distribution of wages across all jobs in the market. In any Nash equilibrium of the bidding subgame, there is a $w_h > 0$ such that (i) the support of the wage distribution F is the interval $[0, w_h]$, and (ii) $F(w)$ is continuous and strictly increasing on $(0, w_h]$.*

Despite the regularities described in Lemma 1, every equilibrium of the bidding subgame has infinitely many variants as agents can modify their bidding behavior on zero-measure sets, while fringe firms may interchange their strategies. Since these modifications affect neither the equilibrium wage distribution nor firm profits, we ignore this trivial form of multiplicity and refer simply to a Nash equilibrium instead of a class of essentially equivalent equilibria.

Let $G(\cdot)$ be the cumulative distribution of wages offered by fringe firms in aggregate, i.e., $G(w)$ is the fraction of fringe firms whose wage does not exceed w. $H_m(\cdot)$, on the other hand, denotes the cumulative distribution of wages offered by agent $m \in \{1, 2, ..., M\}$. Since the wage schedule $w_m(\theta)$ indicates the θ-percentile wage among all job offers from agent m, $H_m(\cdot)$ is the inverse of $w_m(\cdot)$.

Each fringe firm chooses a wage w to maximize its profit

$$\pi_f = R_0 \left((1-s)G(w) + \frac{s}{M} \sum_{m=1}^{M} H_m(w) \right) - w, \tag{2}$$

where the sum of terms in the big parentheses denotes the total share of jobs paying less than w, given the wage offers of all other firms (see Corollary 1) in the market. Since the market is balanced, this sum also denotes the measure of workers hired at wages up to w. Therefore, the fringe firm ends up with a worker indexed exactly by this sum.

Agent $m \in \{1, 2, \ldots, M\}$, on the other hand, chooses an increasing wage schedule $w_m(\cdot)$ to maximize its total profits

$$\pi_m = \frac{s}{M} \int_0^1 \left[R_1 \left((1-s)G(w_m(\theta)) + \frac{s}{M}\theta + \frac{s}{M} \sum_{m' \neq m} H_{m'}(w_m(\theta)) \right) - w_m(\theta) \right] \mathrm{d}\theta. \tag{3}$$

The sum of terms in the big parentheses of Equation (3), as is the case for a fringe firm above, characterizes the measure of workers hired in the market at wages below the θ-quantile of agent m's wage schedule. The term $\frac{s}{M}\theta$, so to say, reflects the market power of agent m: the worker who receives the θ-quantile wage from agent m has an ability level higher than $\frac{s}{M}\theta$ coworkers at the same agent, regardless of the exact wage $w_m(\theta)$. This allows agent m to exploit its market power and lower all its wages. To calculate its total profits, agent m aggregates the profits it makes with every single job-worker pair and corrects by its measure $\frac{s}{M}$.

Equations (2) and (3) reveal the fundamental difference between the optimization problems an agent and fringe firms face. When maximizing its profits by setting a wage, a fringe firm simply evaluates the trade-off between output and cost (wage) of a signle-worker job pair. An agent of the big firm, on the other hand, when raising (lowering) the wages of any number of jobs, has to additionally internalize the effects of its actions on the remainder of its jobs. In other words, an agent has to consider the externalities of its bids on the success of all its jobs.

Since we already regard permutations of fringe firms and zero-measure modifications as non-essential, Proposition 1 below establishes that the equilibrium strategies of the agents are unique in the bidding subgame.[14]

[14] If we were to assume heterogeneous fringe firms, the equilibrium allocation of workers to fringe firms would be unique. Assuming homogeneous fringe firms does not affect the main findings of this paper but introduces unwanted algebraic complexity.

Proposition 1 (Uniqueness)
The equilibrium wage distributions, G and $\{H_m\}_{m=1}^{M}$, and the matching of workers to firms of the bidding subgame with M agents is unique up to a reallocation of workers across fringe firms.

Proposition 1 below establishes that the focus on symmetric equilibria is not an assumption, but rather the unique outcome of the bidding stage. In equilibrium, every agent m adopts the same strategy $w_m(\cdot) \equiv w(\cdot)$. As a consequence, the cumulative distribution of wages also coincide: $H_m(\cdot) \equiv H(\cdot)$ for all m. The unique Nash equilibrium outcome of the bidding subgame is therefore conveniently characterized by a pair of functions $G(\cdot)$ and $H(\cdot)$. In the equilibrium of the bidding subgame, all wages in the support of $G(\cdot)$ necessarily accrue equal profits, for if not there must be a fringe firm with a beneficial unilateral deviation. Moreover, the wage schedule $w(\cdot)$ maximizes the profit of each and every agent m. Depending on market characteristics, this equilibrium displays different patterns.

3.1.1 A Leading Big Firm ($\mu > 1$)

We start by characterizing the bidding equilibrium for the case in which the big firm is more productive than its fringe competitors, i.e., $\mu > 1$. On one hand, due to the supermodularity of the production function, agents exhibit an incentive to up their bids and, as a result, acquire more able workers. On the other hand, however, each agent still retains some market power that tempts it to bid lower. These tensions may generate wages that overlap with those of fringe firms, and, as a result, cause matching inefficiency. Lemma 2 summarizes all possible equilibrium scenarios based on parameters.

Lemma 2 (Bidding equilibrium: $\mu > 1$)
Suppose the big firm is more productive ($\mu > 1$) than its fringe competitors, and delegates its bidding to $M \geqslant 1$ agents. Then:
(i) If the big firm delegates to many agents ($M \geqslant \frac{\mu}{\mu-1}$), there exist cutoffs $w_l \equiv 1 - s$ and $w_h \equiv w_l + \frac{M-1}{M} s\mu$ such that:

$$\begin{aligned} G(w) &= \tfrac{w}{1-s}, & w \in [0, w_l], \\ H(w) &= \tfrac{M}{(M-1)\mu} \tfrac{w - w_l}{s}, & w \in [w_l, w_h]. \end{aligned}$$

(ii) If the big firm delegates to an intermediate number of agents ($\frac{\mu}{\mu-1} s \leqslant M < \frac{\mu}{\mu-1}$), there

exist $w_l \equiv 1 - \frac{s}{M}\frac{\mu}{\mu-1}$ *and* $w_h \equiv 1$ *such that:*

$$
\begin{aligned}
G(w) &= \begin{cases} \frac{w}{1-s}, & w \in [0, w_l], \\ G(w_l) + \frac{\mu - M(\mu-1)}{\mu}\frac{w-w_l}{1-s}, & w \in (w_l, w_h], \end{cases} \\
H(w) &= \frac{M(\mu-1)}{\mu}\frac{w-w_l}{s}, \qquad\qquad\qquad\quad w \in [w_l, w_h].
\end{aligned}
$$

(iii) If the big firm delegates to few agents ($M < \frac{\mu}{\mu-1}s$), *there exist* $w_l \equiv \frac{(M-1)\mu(s\mu - M(\mu-1))}{M(\mu - M(\mu-1))}$ *and* $w_h \equiv \frac{M-s}{M}\mu$ *such that:*

$$
\begin{aligned}
G(w) &= \frac{M-(M-1)\mu}{\mu}\frac{w-w_l}{1-s}, \qquad\qquad w \in [w_l, w_h], \\
H(w) &= \begin{cases} \frac{s\mu-(\mu-1)}{s}, & w = 0,\ M = 1, \\ \frac{M}{(M-1)\mu}\frac{w}{s}, & w \in [0, w_l],\ M > 1, \\ H(w_l) + \frac{M(\mu-1)}{\mu}\frac{w-w_l}{s}, & w \in (w_l, w_h]. \end{cases}
\end{aligned}
$$

Intuitively, if the big firm delegates to many agents, the resulting fierce competition among its agents drives their wages up to the extent that the lowest wage offered by agents exceeds the highest wage offered by the fringe firms (scenario (i) in Lemma 2). In this case, they no longer effectively compete with fringe firms. If the big firm delegates to an intermediate number of agents, its agents compete mildly in an "oligopsony." They utilize their market power to underbid such that their wage schedules partially overlap with wages offered fringe firms (scenario (ii)). Finally, if the big firm delegates to very few agents, the market power of each agent is overwhelming. They underbid so severely that a proportion of jobs feature wages lower than all wages offered by fringe firms (scenario (iii)). In particular, if the big firm does not delegate at all, it bids 0 for a positive measure of jobs.

Closer inspection of Lemma 2 reveals that the overall distribution of wages increases, in the sense of first order stochastic dominance, in the number of agents M the big firm delegates to, holding everything else fixed. This is a direct consequence of the increased level of competition among the big firm's agents.

Moreover, in equilibrium, agents never bid such that all their wages fall short of the wages posted by any fringe firm. This is because when $\mu > 1$, it is always tempting and affordable for the agents to compete with the fringe firms for more able workers. This intuition, however, is no longer true for a trailing big firm, i.e, $\mu \leqslant 1$, as discussed below.

3.1.2 A Trailing Big Firm ($\mu \leqslant 1$)

Now suppose that the big firm is less productive than its fringe competitors, i.e., $\mu \leqslant 1$. In this case, both the supermodularity of the production function and the market power work in the same direction to encourage lower bids, as is formally stated in Lemma 3 below.

Lemma 3 (Bidding equilibrium: $\mu \leqslant 1$)
Suppose the big firm is less productive ($\mu \leqslant 1$) than its fringe competitors, and delegates its bidding to $M \geqslant 1$ agents. Then, there exist cutoffs $w_l \equiv \frac{M-1}{M} s\mu$ and $w_h \equiv w_l + 1 - s$ such that:

$$
\begin{aligned}
G(w) &= \tfrac{w-w_l}{1-s}, \qquad w \in [w_l, w_h], \\
H(w) &= \begin{cases} 1, & w = 0,\ M = 1, \\ \tfrac{M}{(M-1)\mu}\tfrac{w}{s}, & w \in [0, w_l],\ M > 1. \end{cases}
\end{aligned}
$$

Lemma 3 establishes that the big firm's agents always underbid fringe firms, regardless of the number of agents. Due to their productivity disadvantage, agents value worker ability less than fringe firms. In addition, the agents' market power motivates them to lower their posted wages even further. Building on these forces, agents prefer hiring less able workers at very low wages, rather than acquiring more able ones at higher wages.

It is straightforward to see from Lemma 3 that, holding everything else fixed, the wage thresholds w_l and w_h also increase in the number of agents M. As a consequence, both wages paid by fringe firms as wells as those paid by agents of the big firm increase, while the allocation of workers to firms is unchanged.

3.2 Stage One: Optimal Delegation

Taking the unique equilibrium outcome of the bidding stage as given, we now focus on the first stage to characterize the optimal delegation strategy of the big firm. We find that the optimal number of agents depends on both the relative productivity μ and its size s. In order to simplify algebra, we refrain throughout the paper from making delegation costly. It is shown below that the the incentive to delegate is always bounded, i.e, the optimal choice of M is finite. As such, adding a delegation cost does not alter the qualitative insights of our analysis. Theorem 1 states that a big firm only delegates if it is moderately productive.

Theorem 1 (No delegation)
The big firm never delegates in equilibrium if it less productive ($\mu \leqslant 1$), or significantly more productive ($\mu > 4$) than its fringe competitors.

The intuition behind this result is as follows. Suppose the big firm is trailing, i.e., $\mu \leqslant 1$. Following Lemma 3, agents bid lower wages for all their jobs than any fringe firm, and hence always end up with all workers indexed from 0 to s. As a result, the average output of all jobs from the big firm is independent of the number of agents M. On the other hand, the average wage offered to these workers increases in M due to competition. It follows that a trailing big firm optimally never delegates. In other words, if the big firm is destined to match with the least able batch of workers independent of how many agents it delegates to, there is no point in delegating to multiple agents and competing wages upwards.

Now consider the opposite extreme, when the big firm is highly productive, i.e., $\mu > 4$. Then, any delegation to a number of agents $M \geqslant 2$ falls into scenario (i) of Lemma 2, such that the agents of the very productive big firm never directly compete with fringe firms. Specifically, by delegating to $M \geqslant 2$ agents, the big firm acquires all workers with indices from $1 - s$ to 1. Therefore, its average output is fixed whereas its average wage expenditure is increasing in $M \geqslant 2$. However, when $M = 1$, the subgame falls into scenario (ii) or (iii), where the agents' wage offers overlap with the fringe firms'. Since the big firm is highly productive, the choice of $M = 2$ is dominated by $M = 1$ because the highly productive agents compete fiercely with one another, so that the additional competition is more costly than its associated benefit from only a slightly more able workforce.

Finally, we dig into the interior case of a moderately leading big firm. That is to say, the big firm is more productive than its fringe competitors, but this advantage is limited, i.e., $1 < \mu < 4$. For such a moderately productive big firm, Theorem 2 summarizes our second observation that it delegates only if its size is not overwhelming.

Theorem 2 (Delegation)
When the big firm is moderately productive ($\mu \in (1, 4)$), there exists a cutoff size $s^*(\mu) \in (0, 1)$ and a number of agents $M^*(\mu) > 1$ such that the big firm delegates in equilibrium to $M^*(\mu)$ agents if $s < s^*(\mu)$, and does not delegate otherwise. Moreover, the cutoff size $s^*(\mu)$ is strictly increasing in μ, and the optimal number of agents $M^*(\mu)$ is decreasing in μ.

Theorem 2 reveals that the tendency to delegate decreases in the big firm's size. The intuition behind this result is based on the trade-off between the competing forces of market power and commitment power. Specifically, the big firm desires to be a Stackelberg leader in the market—setting its wage profile before its fringe competitors do—but lacks the ability to realize this outcome. Delegation to multiple agents is an imperfect yet effective way to commit to a more aggressive bidding strategy, thereby favorably influencing fringe firms' actions. Such a commitment by delegation is not without its cost, however, because the big firm relinquishes much of its market power to suppress wages. Balancing the tension between these two forces, a firm gives up market power in exchange for commitment power. A big

[15] In the picture, $\mu = 6/5$, and hence $\frac{\mu}{\mu-1} = 6$. The highest profit is obtained at $M = 6$ when $s < s^*(\mu) = 2/7$, and at $M = 1$ when $s \geqslant s^*(\mu)$.

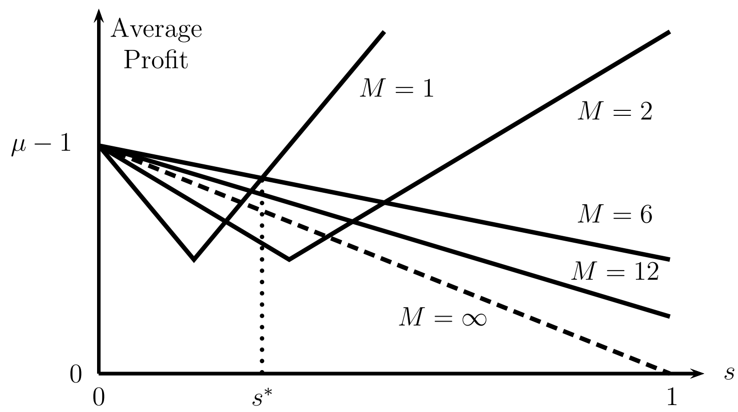


Figure 1: Profit comparison for different numbers of agents for a productive big firm.[15]

firm with small size never holds tremendous market power in the first place, and therefore the gains from commitment by delegation are more attractive. A big firm with large size, on the other hand, has too much market power to forsake. For such a nearly monopsonistic firm, delegation only leads to increased wages without significantly improving the ability of its workforce. The trade-offs between these two forces are illustrated in Figure 1.

When a big firm optimally delegates, the optimal number of agents $M^*(\mu) \geqslant 2$ is the result of the trade-off between the market power and commitment power. The intuition why it decreases in μ is as follows. When μ just exceeds 1, the big firm's productivity advantage over fringe firms is insignificant, which means they are not differentiated enough. Competition among the big firm's agents increases its average wage only mildly while it greatly fends off the threat from fringe firms, in terms of winning over the best workers. As μ increases further, however, the competition among its agents raises the wages by too much, more than offsetting the gains from a more able workforce.

Another result stated in Theorem 2 is that the cutoff $s^*(\mu)$, above which the big firm never delegates, increases in its relative productivity μ. The intuition behind this relationship is as follows. Market power is relatively more important for a less productive firm. As a consequence, holding firm size fixed, whenever a less productive big firm delegates, a more productive one does so as well. The reverse statement is not true. Figure 2 depicts the optimal cutoff $s^*(\mu)$ as a function of μ and indicates the optimal number of agents $M^*(\theta)$ the big firm optimally delegates to.

3.3 Output, Firm Profits and Wages

Total surplus is the aggregate output across all worker-job pairs. It is maximized (minimized) when the matching is positively (negatively) assortative, that is, more (less) able

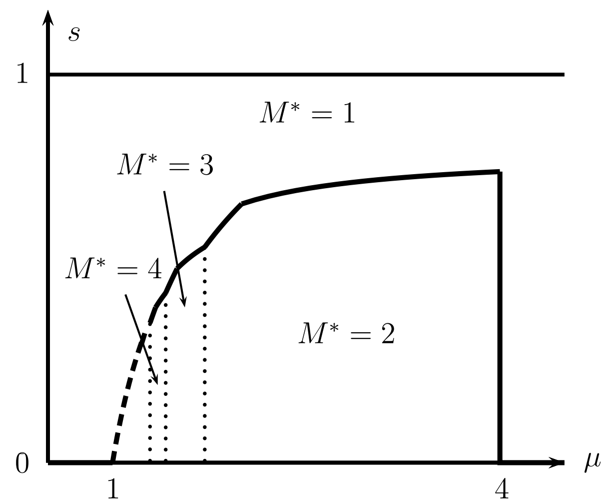


Figure 2: Characterization of stage-one decision in the (μ, s) space. The increasing curve drawn is s^*. The dashed part of the curve is omitted because it has infinitely many piece-wise segments.

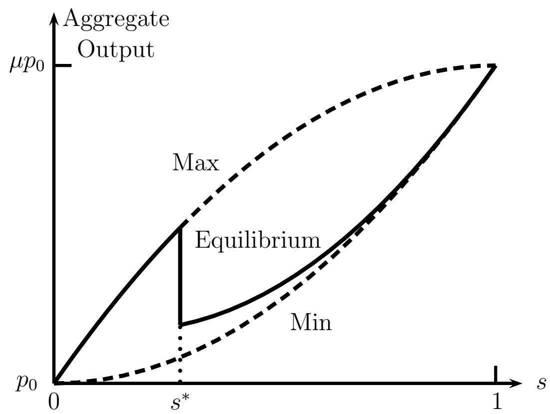
workers are hired by more productive firms. Therefore, the first best is achieved when positive assortative matching (PAM) occurs in equilibrium. We learned that a big firm always exhibits an incentive to exercise its market power, and only delegates if it is moderately productive. Whenever such a moderately big firm delegates, the competition among its agents causes the big firm to hire more able workers.

Lemma 4 (Monotonicity)
Total surplus weakly increases as the big firm delegates hiring to more agents.

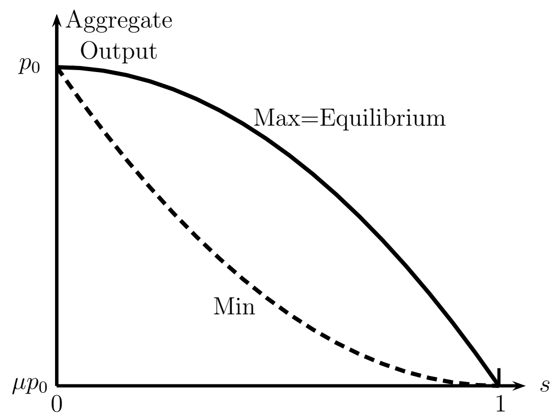
In equilibrium, welfare depends on the big firm's endogenous choice of delegation, which in turn is a function of the big firm's relative productivity and size. Inspection of how total surplus is generated in our model allows us to make the following observations.

Proposition 2 (Total surplus)
In equilibrium,
(i) if the big firm is highly productive ($\mu > 4$), total surplus strictly increases in its size, while the first best is never obtained.
(ii) If the big firm is moderately productive ($1 < \mu < 4$), total surplus strictly increases in its size whenever $s \neq s^*(\mu)$, but exhibits a discontinuity at $s^*(\mu)$, where it drops. The first best is obtained for some μ when $s < s^*(\mu)$, but is never obtained when $s > s^*(\mu)$.
(iii) If the big firm is less productive than its fringe competitors ($\mu \leqslant 1$), total surplus strictly decreases in its size, while the first best is always obtained.

Theorem 1 states that a highly productive big firm never delegates, thereby putting its market power to maximal use and keeping total surplus away from the first best. The total surplus increases in s because the big firm is more productive than the fringe firms. If, instead, the big firm is moderately productive, it may or may not delegate to multiple agents, depending on its size. In this case, the first best obtains or nearly obtains when $s < s^*(\mu)$ because the increased level of decentralization at the big firm improves the matching quality since it hires more able workers. As soon as $s > s^*(\mu)$, the big firm stops delegating, however, and suppresses wages by means of their market power, and results in suboptimal matching quality. Figure 3(a) illustrates a case in which the first best obtains only if the size of the big firm is below s^*. Finally, an unproductive big firm never delegates, ending up with the least able workers. Figure 3(b) depicts the total surplus when the big firm is unproductive ($\mu \leqslant 1$). Despite its market power, wage suppression does not lead to inefficient matching.



(a) $\mu > 1$



(b) $\mu < 1$

Figure 3: Total surplus and the first best outcome as a function of the size of the big firm. (a): Leading big firm, $\mu > 1$. (b): Trailing big firm, $\mu < 1$.[17]

What is noteworthy, although the first best strictly increases in $\mu > 1$, the equilibrium total surplus does not. Specifically, for $s < s^*(\mu)$, as μ increases, $M^*(\mu)$ steadily decreases until the big firm ceases to delegate. Whenever μ is in the vicinity a cutoff value where M^* jumps downward, so does total surplus. Above $\mu = 4$, total surplus always increases in μ since there are more jobs with a very high productivity.

Due to the supermodularity of the production function and the transferability of utility, comparing total surplus to the first best outcome is a formidable way to assess the quality of matching between firms and workers.

[17]Left panel: the case of $\mu = 6/5$ and $s^*(\mu) = 2/7$. Right panel: the case of $\mu = 4/5$.

Proposition 3 (Matching quality)
For every given productivity $\mu > 1$, there exists a threshold $\bar{s}(\mu) \in (0, 1]$ such that the matching quality, defined as the ratio between the total surplus and the first best, decreases in the big firm's size s when $s < \bar{s}(\mu)$.

Proposition 3 states that the quality of matching, or the "assortativeness" of the equilibrium allocation, in general declines the larger the big firm is. The intuition is straightforward. A larger big firm is less likely to delegate and relies more on its market power, thus underbidding relative to its productivity. For very high s, i.e., $s > \bar{s}(\mu)$, the direction can be reversed because the supermodularity of the production function becomes overwhelming as s approaches 1 and the potential for misallocation vanishes.

Next, we turn to firm profits. Since equilibrium profits are necessarily constant across the competitive fringe, the profits of the big firm and any fringe firm exhaustively describe firm profits. For the remainder of this section we extend the domain of $s^*(\cdot)$ and set $s^*(\mu) \equiv \frac{\mu-1}{\mu}$ for all $\mu \geqslant 4$. Proposition 4 below describes firm profits as a function of s and μ.

Proposition 4 (Firm profits)
(i) If the big firm is more productive ($\mu > 1$) than its fringe competitors, its per-job profit decreases in its size for $s < s^*(\mu)$, and increases for $s > s^*(\mu)$, and
(ii) the profit of a fringe firm increases in the big firm's size and decreases in the big firm's relative productivity.
(iii) If the big firm is less productive ($\mu \leqslant 1$) than its fringe competitors, its per-job profit increases in its size, and
(iv) the profit of a fringe firm strictly increases in the big firm's size while it is independent of the big firm's productivity.

When the big firm is more productive than the competitive fringe and its size does not exceed the cutoff s^*, it optimally delegates hiring to a number of agents who compete for the most able workers. As its size increases, it naturally hires also lower-ability workers, driving down the average profit. Fringe firms absorb the least able workers in this scenario, and the ensuing competition among them at low wages drives their profits down to zero.

On the other hand, when the big firm is more productive than its fringe competitors and its size exceeds the cutoff s^*, it never delegates and leverages its size to bid as low as possible. As its size increases, it recruits more workers of higher ability, increasing average profits. Fringe firms now earn positive profits because they compete for more able workers. What is noteworthy is that fringe firms benefit from a larger quasi-monopsonist, since a lack of delegation lowers wages across the entire market.

When the big firm is less productive than the competitive fringe, however, it does not delegate, and hires workers indexed in $[0, s]$. Clearly its average profit increases with size because it hires marginally more able workers as it grows. The fringe firms, however, also benefit from a larger big firm because they end up hiring more able workers without an increase in wages.

Last but not least, we conclude this section by a discussion of worker surplus, i.e., wages, as a function of the relative productivity of the big firm μ and its size s. We denote the equilibrium overall wage distribution by $F^s_\mu(w)$, with μ and s as parameters. We adhere to the standard definition of first-order stochastic dominance among probability distributions.

Definition 2 *A distribution $F(x)$ first-order stochastically dominates ($\succeq_{FOSD}$) distribution $\hat{F}(x)$ if and only if $F(x) \leqslant \hat{F}(x)$ for all x.*

Proposition 5 below establishes comparative statics for the equilibrium wage distribution.

Proposition 5 (Wages)
(i) If the big firm is less productive than its fringe competitors ($\mu \leqslant 1$), all wages decrease in its size s, i.e., $s < s'$ implies $F^s_\mu(w) \succeq_{FOSD} F^{s'}_\mu(w)$.
(ii) If the big firm is moderately productive ($\mu \in (1, 4)$), and delegates in equilibrium ($s < s^*(\mu)$), all wages increase in its size s, i.e., $s < s' < s^*(\mu)$ implies $F^s_\mu(w) \preceq_{FOSD} F^{s'}_\mu(w)$.
(iii) If the big firm is highly productive ($\mu > 4$), or moderately productive ($\mu \in (1, 4)$) and does not delegate in equilibrium ($s > s^*(\mu)$), all wages decrease in its size s, i.e., $s^*(\mu) < s < s'$ implies $F^s_\mu(w) \succeq_{FOSD} F^{s'}_\mu(w)$.
(iv) Finally, $F^s_\mu(w)$ approaches its point-wise infimum as $s \uparrow \frac{3}{4}$ and $\mu \uparrow 4$. The average wage is bounded above at $\frac{25}{32}$.

According to Proposition 5, an increase in the size of a trailing big firm always hurts workers. This is because more workers are hired at zero wage. What is more, the wage distribution is independent of the big firm's productivity. This follows from the fact that a trailing big firm never delegates to a single agent and bids zero wage regardless of its productivity. As a consequence, the equilibrium distribution of wages only depends on the big firm's size and the productivity of the competitive fringe.

If the big firm is more productive than its fringe competitors, i.e., $\mu > 1$, workers benefit from a larger big firm as long as it delegates to multiple agents. When the big firm's size s increases but falls short of the threshold $s^*(\mu)$, two effects arise. Since μ exceeds 1, increasing the number of jobs at the big firm increases the first best outcome of the market. On the other hand, as s increases, the big firm delegates bidding to a smaller number of agents, and thus, wages fall and the matching quality in the market worsens. Proposition 5 (ii) establishes that the first effect dominates the second. In other words, as long as the big firm delegates, workers receive some share from an increase in total surplus.

Proposition 5 (iii) states that the same logic does not apply when the big firm does not delegate. Then, any increase in total surplus is split between the big firm and its fringe competitors while workers' compensation declines. Finally, the effect of the productivity of the big firm on worker compensation is bounded. As soon as μ passes a certain threshold, productivity gains are swallowed by the big firm, but not passed on to workers. Although in equilibrium it is sometimes in the big firm's interest to delegate to multiple agents, this self-induced competition is never fierce enough to raise average wages significantly and it is very much the productivity of the competitive fringe that is crucial for equilibrium worker compensation.

4 Minimum wage

Up to this point we have assumed that all workers have an outside option of zero. In other words, zero is the lower bound of wages in the market. What happens if we deliberately increase the lowest level of remuneration in the market, i.e., more generally, introduce a price floor? Note that this section does not address the effects of introducing a minimum wage in the overall economy, but the consequences of a price floor such as a union wage in a professional labor market with heterogeneous workers. We have learned in Section 3 that market power of big firms in general skews surplus distribution decisively towards firms. In fact, a big firm exercising its market power may remunerate a portion of workers at the lower bound wage. A common labor market intervention geared towards supporting lower ability workers is the introduction of a minimum or union wage.

Economics textbook intuition tells us that the introduction of a minimum wage in a competitive market increases unemployment since firms re-evaluate their hiring decisions and adjust their input mix accordingly. It is well known that the same logic does not necessarily apply to environments in which individual firms hold market power. In a true monopsony, for example, a legally binding minimum wage should even increase employment (see e.g., Ashenfelter et al., 2010). Bhaskar and To (1999) argue that under monopsonistic competition, i.e., several firms that command market power competing with each other, a minimum wage raises employment per firm but causes firm exit. Its overall effect on social welfare is positive if industry employment increases, and ambiguous if it does not. This logic, however, relies on a free-entry/zero-profit condition, an assumption that is at odds with the reality of many professional labor markets in which firms command market power. Consider for example local health care systems in the market for nurses (Staiger et al., 2010), highly profitable employers that often enjoy market power.

In our model, the introduction of a minimum wage induces a similar direct effect as it does in competitive markets. Suppose we were to introduce a strictly positive minimum wage $\underline{w}$. According to the output functions, a fringe firm never hires a worker with index $i < \underline{w}$ because $\pi = R_0(i) - \underline{w} = i - \underline{w} < 0$. Similarly, an agent from the big firm never hires a worker with index $i < \frac{\underline{w}}{\mu}$.

As a consequence, unemployment necessarily increases from zero to a positive level, and we would expect a lower total surplus due to the reduced number of matched worker-job pairs. This is not necessarily true, however. A more complex strategic effect arises in our model as the delegation decision is endogenous. This strategic effect influences the competitiveness in the bidding game, and subsequently, indirectly determines the matching qualities and surplus allocation.

Lemma 5 (Minimum wage and delegation)
If the big firm is moderately productive ($\mu \in (1, 4)$), and the minimum wage $\underline{w}$ is sufficiently small, there exists a cutoff size $s^*_{\underline{w}}(\mu)$, such that the big firm does not delegate if $s > s^*_{\underline{w}}(\mu)$, and delegates to $M^*(\mu)$ agents if $s < s^*_{\underline{w}}(\mu)$. Moreover, the cutoff $s^*_{\underline{w}}(\mu)$ is increasing in the minimum wage $\underline{w}$.

Lemma 5 teaches us that—holding everything else fixed—the level of the minimum wage determines the cutoff size below which the big firm prefers to delegate to multiple agents. In fact, the higher the minimum wage, the more likely the big firm is to delegate. As we know from Proposition 2, delegation increases total matching surplus for all μ and s. As it turns out, this effect may even dominate the loss in surplus due to increased unemployment.

Theorem 3 (Comparative statics)
If the big firm is moderately productive ($\mu \in (1, 4)$), an increase in the minimum wage $\underline{w}$ always raises unemployment. However, total surplus increases discretely if this increase induces the big firm to delegate to multiple agents instead of one.

In fact, there are two forces at play that increase unemployment. First, some workers are not able enough to be hired by any firm at the minimum wage. Secondly, a big firm that delegates to multiple agents, rather than posting a single wage schedule increases wages by competition among its agents. As a consequence, it is now the less productive fringe firms that hire the workers of the lowest ability. These, however, are even less likely to hire workers of lower ability. That being said, the improved matching between firms and workers in response to a big firm's delegating may generate more surplus than is lost due to the reduced number of matches.

To summarize, the introduction of a minimum wage in our model has two consequences, a direct effect and a strategic effect. The direct effect arises since some workers are not able enough to warrant being hired under a binding minimum wage, increases unemployment and lowers total surplus due to the lost number of matches. The strategic effect, on the other hand, arises if the minimum wage is high enough to push the big firm to delegate. In this case, the big firm hires more able workers and the fact that the weaker fringe firms now hire the worst workers causes additional unemployment. The positive matching effect, however, may even dominate the welfare loss due to lost matches, and total surplus may ultimately fall or rise. In other words, a low minimum wage raises unemployment among the least able and reduces overall surplus. A higher minimum wage reduces unemployment among lower ability workers even further. In case total surplus increases, any additional gains are split among firms and and high ability workers.

5 Extensions

Below we support the validity of the main results presented in this paper by establishing that the underlying intuition is robust to alternative setups. First, we introduce a second big firm to show that market power and not the existence of a single big firm is the driving force behind our results. Thereafter, we analyze the case in which agents do not wage-discriminate among the workers they hire, a common restriction in many professional labor markets, and find that our main results remain unchanged. Finally, we consider more general output functions to replicate the same insights.

5.1 Strategic Delegation in a Duopsony

In this subsection, we replace the competitive fringe with a second big firm commanding market power. We refer to this scenario as a "duopsony." To be more specific, there are two big firms 1 and 2 with respective sizes s_1 and $s_2 \equiv 1 - s_1$. Since there are no fringe firms, we refer to the two big firms simply as "firms" without confusion. The output of a job in firm $j = 1, 2$ when matched with worker i is $R_j(i)$, and again, for simplicity we assume $R_2(i) = i$ and $R_1(i) = \mu i$. Without loss of generality, $\mu \geqslant 1$ is interpreted as the relative productivity of (the more efficient) firm 1 compared to firm 2. The delegation stage is no longer a decision problem for a single firm. Instead, two firms simultaneously decide $M_1 \geqslant 1$ and $M_2 \geqslant 1$, the number of agents they want to delegate to. Thereafter, in the bidding stage, each of the $M_1 + M_2$ agents simultaneously posts a wage schedule $w_{j,m}(\cdot)$, where $m \in \{1, \ldots, M_j\}$ and $j = 1, 2$. Finally, more able workers are matched with jobs that pay higher wages.

Let $H_{j,m}(\cdot)$ $(j = 1, 2)$ denote the cumulative wage offer distribution of agent m from firm j. The wage schedule $w_{j,m}$ is the inverse of $H_{j,m}$. Taking all other agents' schedules as given, an agent m from firm j chooses a wage schedule $w_{j,m}(\cdot)$ to maximize its total profit

$$\pi_{j,m} = \frac{s_j}{M_j}\int_0^1 \left[R_j\left(\sum_{(j',m')\neq(j,m)} \frac{s_{j'}}{M_{j'}} H_{j',m'}(w_{j,m}(\theta)) + \frac{s_j}{M_j}\theta\right) - w_{j,m}(\theta)\right] \mathrm{d}\theta, \tag{4}$$

subject to the constraint that $w_{j,m}$ is increasing. As can be seen from the expression in square brackets in Equation (4), each firm now takes into account how many workers are hired by all agents of both firms when setting its θ-quantile wage. Lemma 6 below is the counterpart of Lemma 2, describing the wage distributions in the bidding subgame among M_1 agents of firm 1 and M_2 agents of firm 2.

Lemma 6 (Bidding equilibrium)
Suppose firm j delegates to M_j agents, $j = 1, 2$. Then,
(i) if the big firm delegates to many agents ($M_1 \geqslant \frac{\mu}{\mu-1}$), there exist $w_l \equiv \frac{M_2-1}{M_2}(1 - s_1)$ and $w_h \equiv w_l + \frac{M_1-1}{M_1} s_1\mu$ such that:

$$\begin{aligned} H_1(w) &= \tfrac{M_1}{(M_1-1)\mu}\tfrac{w-w_l}{s_1}, \quad && w \in [w_l, w_h], \\ H_2(w) &= \tfrac{M_2}{M_2-1}\tfrac{w}{1-s_1}, && w \in [0, w_l]. \end{aligned}$$

(ii) If the big firm delegates to an intermediate number of agents ($\frac{\mu}{\mu-1+\frac{1-s_1}{M_2}} s_1 \leqslant M_1 < \frac{\mu}{\mu-1}$), there exist $w_l \equiv \frac{(M_2-1)((1-s_1+M_2(\mu-1))M_1-M_2 s_1\mu)}{M_1 M_2(1+M_2(\mu-1))}$ and $w_h \equiv \frac{M_2-1+s_1}{M_2}$ such that:

$$\begin{aligned} H_1(w) &= \tfrac{M_1((\mu-1)M_2+1)}{(M_1+M_2-1)\mu}\tfrac{w-w_l}{s_1}, && w \in [w_l, w_h], \\ H_2(w) &= \begin{cases} \tfrac{M_2}{M_2-1}\tfrac{w}{1-s_1}, & w \in [0, w_l], \\ H_2(w_l) + \tfrac{M_2(\mu-(\mu-1)M_1)}{(M_1+M_2-1)\mu}\tfrac{w-w_l}{1-s_1}, & w \in (w_l, w_h]. \end{cases} \end{aligned}$$

(iii) If the big firm delegates to few agents ($M_1 < \frac{\mu}{\mu-1+\frac{1-s_1}{M_2}} s_1$), there exist $w_l \equiv \frac{(M_1-1)(M_2 s_1\mu - M_1(1-s_1+M_2(\mu-1)))}{M_1 M_2(M_1-(M_1-1)\mu)}\mu$ and $w_h \equiv \frac{M_1-s_1}{M_1}\mu$ such that:

$$\begin{aligned} H_1(w) &= \begin{cases} \tfrac{M_2 s_1\mu-(1-s_1+M_2(\mu-1))}{M_2 s_1}, & w = 0,\ M = 1, \\ \tfrac{M_1}{(M_1-1)\mu}\tfrac{w}{s_1}, & w \in [0, w_l],\ M > 1, \\ H_1(w_l) + \tfrac{M_1(M_2(\mu-1)+1)}{(M_1+M_2-1)\mu}\tfrac{w-w_l}{s_1}, & w \in (w_l, w_h], \end{cases} \\ H_2(w) &= \tfrac{M_1-(M_1-1)\mu}{(M_1+M_2-1)\mu}\tfrac{w-w_l}{1-s_1}, \qquad w \in [w_l, w_h]. \end{aligned}$$

Closer inspection of Lemma 6 reveals that the equilibrium of the bidding subgame between two big firms described by Lemma 6 resembles the equilibrium of the bidding subgame between a big firm and its competitive fringe. In either case, if a productive firm delegates bidding to many agents, it does not directly bid against its less productive competitors, but instead offers only wages above the wages set by other firms. If the number of agents is intermediate, the firm bids against its competitors but they also offer wages below. And, finally, if a productive firm only delegates to very few agents, it bids directly against its competitors but also offers lower wages exclusively. The robustness of the intuition underlying the bidding equilibrium establishes that the driving forces in our model do not rely on the specific market constellation described in the basic model in Section 2. Proposition 6 builds on the outcome of the bidding stage in a duopsony and describes the subgame-perfect equilibrium of the overall game.

Proposition 6 (Delegation with two big firms)
If firm 1 *is highly productive* ($\mu > 4$)*, it never delegates. In this case, firm* 2 *does not delegate if firm 1 is small* ($s_1 < \frac{2\mu-1}{2\mu+1}$)*, and otherwise delegates to infinitely many agents. If firm* 1 *is moderately productive* ($\mu \in (1,4)$)*, there are cutoffs* $0 < s^*(\mu) < s^{**}(\mu) < 1$ *and an integer* $M^*(\mu) \geqslant 2$ *such that:*
(i) If firm 1 *is smaller than* $s^*(\mu)$*, it optimally delegates to* $M^*(\mu)$ *agents while firm* 2 *does not delegate.*
(ii) If firm 1 *is larger than* $s^{**}(\mu)$*, it does not delegate while firm* 2 *delegates to infinitely many agents.*

Therefore, if one big firm is sufficiently large and more productive than the other, the delegation equilibrium resembles a quasi-monopsony, with one dominant firm and infinitely many small ones. Moreover, we can see from Proposition 6 (i) that a more productive big firm may delegate to many agents whereas the other big firm does not. While we do not claim that labor market repercussions are necessarily powerful enough to influence firm size distribution in a given industry, strategic incentives in the labor market reinforce the "winner takes most" logic discussed in Section 1.

Finally, if a rather small but productive firm competes with a less productive but large firm, neither one delegates to agents. The intuition underlying this scenario is as follows. The larger productive firm values its market power much more than any commitment power since it does not want to post higher wages due to its lack of superior productivity. On the other hand, the smaller firm is so productive, that splitting would provoke the ensuing competition to increase wages too much.

In summary, the intuition gained from an analysis of a duopsony with strategic delegation does not fundamentally differ from the intuition we obtained in a quasi-monopsony. That is to say, the driving forces behind equilibrium behavior of firms that command market power appear to be fairly robust to model specifications.

5.2 Uniform Wages

In real-world markets, firms often face stark incentives not to pay employees in the same position different wages. This is especially true for incoming workforce. Not discriminating between workers by means of remuneration may help firms to avoid tension among its employees and potential legal repercussions. In addition, once the practice of discriminating wages becomes common knowledge among prospective employees, the firm may reluctantly find itself individually negotiating wages.

Since we do not model any of these potential benefits of uniform wages, we supplement our treatment of quasi-monopsonies in this subsection by analyzing a scenario in which outside forces compel the big firm to post a single wage for all its jobs within a division. In terms of the main model, we assume that every agent in the bidding stage can only offer a single wage w_m for all the jobs under its control as opposed to an increasing wage schedule $w_m(\cdot)$. As a consequence, the equilibrium of the bidding subgame features randomization by agents. We assume the firms and its agents to be risk-neutral.

Abusing notation, let $H_m(w)$ now denote the probability (instead of a fraction) that agent m's wage offer is below w. A risk-neutral firm only cares about the expected output. Therefore, a fringe firm chooses w to maximize its profits:

$$\begin{aligned} \pi &= (1-s)G(w) + \frac{s}{M}\mathbb{E}\tilde{k}(w) - w, \\ &= (1-s)G(w) + \frac{s}{M}\sum_{m=1}^{M} H_m(w) - w, \end{aligned} \tag{5}$$

where $\tilde{k}(w)$ is the random variable indicating the number of agents who bid below w. Note that it reads exactly the same as Equation (2). An agent m, on the other hand, chooses a single wage w_m to maximize the expected profits:

$$\begin{aligned} \pi_m &= \frac{s}{M}\int_0^1 \left[\left((1-s)G(w_m) + \frac{s}{M}\theta + \frac{s}{M}\mathbb{E}\tilde{k}_{-1}(w_m)\right)\mu - w_m\right] \mathrm{d}\theta, \\ &= \frac{s}{M}\int_0^1 \left[\left((1-s)G(w_m) + \frac{s}{M}\theta + \frac{s}{M}\sum_{m'\neq m} H_{m'}(w_m)\right)\mu - w_m\right] \mathrm{d}\theta, \end{aligned} \tag{6}$$

where $\tilde{k}_{-1}(w)$ is the random variable indicating the number of the other $M-1$ agents who bid below w. This expression is the same as Equation (3) once we replace $w_m(\theta)$ with w_m.

It turns out that there is a one-to-one mapping between a pure-strategy equilibrium of the bidding game in which agents offer a wage schedule and and a mixed-strategy equilibrium of the game described in this subsection in which agents offer a single wage. As a consequence, our analysis in Section 3 carries over as long as we reinterpret H as the probability distribution function of a mixed strategy. Additionally, this finding aligns well with firms setting a uniform wage for its incoming workforce in real-world markets.[18] Formally, Proposition 7 states the link between the two settings.

Proposition 7 (Equivalence)
The wage schedules G and $\{H_m\}_{m=1}^{M}$ constitute a Nash equilibrium of the bidding subgame in which each agent is free to offer one wage per job, if and only if they constitute a Nash equilibrium in the bidding subgame in which each agent can only set a single wage for all its jobs, reinterpreting the H_m's as the distribution of a random wage w_m.

Since the expected payoffs of the big firm in the bidding stage are identical to those described in Section 2, it directly follows that the analysis of the delegation stage also carries over. As a result, the intuition presented up to this point remains valid if outside forces compel firms to set a single wage per division.

5.3 Non-Linear Output Functions

So far, for simplicity, we have assumed output functions to be linear. For non-linear cases, the bidding subgame may or may not display randomization in equilibrium, depending on the curvature of these functions.

Specifically, in this subsection we only assume $R_1(i) = \mu R_0(i)$, without imposing linearity. The parameter $\mu > 0$ still captures the relative productivity of the big firm. Rewrite Equations (2) and (3) as objectives for fringe firms and agents, allowing for mixed strategies:

$$\begin{aligned}
\pi_f &= \mathbb{E}\left[R_0\left((1-s)G(w) + \frac{s}{M}\sum_{m=1}^{M} H_m(w)\right)\right] - w, \\
\pi_m &= \frac{s}{M}\int_0^1 \left(\mathbb{E}\left[R_1\left((1-s)G(w_m(\theta)) + \frac{s}{M}\theta + \frac{s}{M}\sum_{m'\neq m} H_{m'}(w_m(\theta))\right)\right] - w_m(\theta)\right) \mathrm{d}\theta.
\end{aligned}$$

Without linearity, the expectation operator cannot move inside the R_j functions, and therefore the mixed strategies of other firms or agents are not equivalent to some "expected"

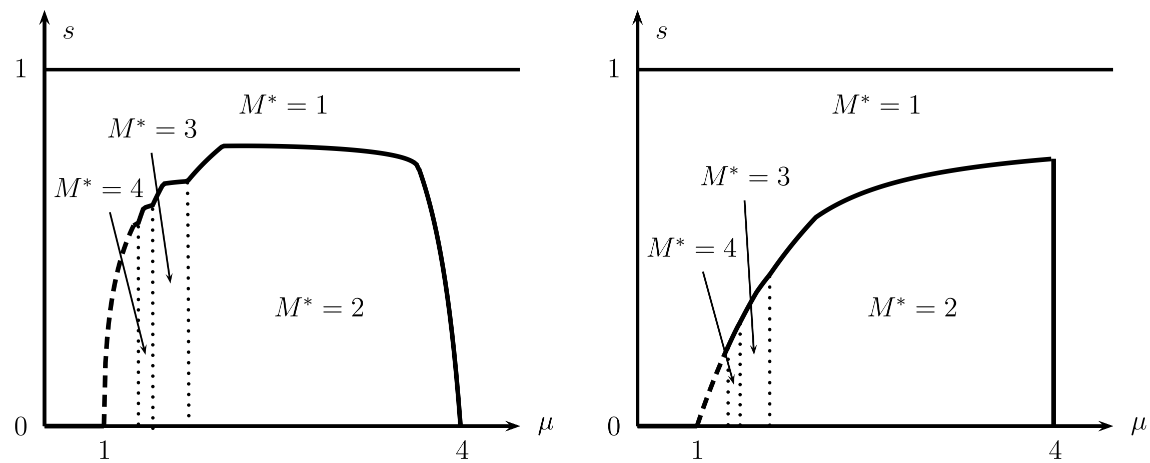
[18] See Jungbauer (2021) for a similar argument for professional labor markets in which multiple firms enjoy varying levels of market power.

distribution functions. To study whether a pure or mixed strategy equilibrium arises, we consider two scenarios.

First, consider the case where R_0 is convex. Assuming the existence of a pure strategy equilibrium, we follow the same procedure as in the main model. It turns out that the properties of the bidding stage in essence follow Lemmas 2 and 3. A key finding, though, is that a pure strategy equilibrium exists if and only if R_0 (and hence R_1) is convex. The proposition below formally states this result.

Proposition 8 *Let $R_1(i) = \mu R_0(i)$ for all $i \in [0,1]$, where $\mu > 0$. A pure strategy equilibrium exists in the bidding subgame for all $M \geqslant 1$ if and only if R_0 is convex.*

With the bidding subgame solved, the optimal delegation in the first stage is readily obtainable by comparing outcomes for different numbers of agents. In Figure 4(a) we plot the optimal delegation rule for output function $R_0(i) = i^2$, as a comparison to Figure 2 in the main model.



(a) $R_0(i) = i^2$ (b) $R_0(i) = 1-(1-i)^2$

Figure 4: Optimal delgation rule for non-linear R_0. (a): A convex output function $R_0(i) = i^2$ with pure strategy bidding. (b): A concave output function $R_0(i) = 1-(1-i)^2$ with mixed strategy bidding.

Next, consider the case where R_0 is concave. From Proposition 8 we already know that no pure strategy equilibrium exists if the output functions are not convex. Here, we conjecture a mixed strategy equilibrium where each agent bids a uniform wage, but this wage is distributed according to $H(\cdot)$. By posting a wage, the number of rival agents it outbids is a random variable with binomial distribution. While explicit-form solutions are not attainable, we nevertheless plot the numerically optimal delegation in Figure 4(b) to show the same pattern, with a concave output function $R_0(i) = 1 - (1-i)^2$. This suggests that our intuition of the market power/commitment power trade-off is robust even without a pure strategy equilibrium.

Finally, there are general cases where R_0 is neither convex nor concave. Based on previous analysis, we know that no pure strategy equilibrium exists. More interestingly, purely mixed equilibrium does not arise either. Our conjecture is a hybrid of the two extremes: a pure strategy with a spectrum of different wage offers for a portion of jobs where R_0 is convex, and randomization of a single wage for the rest of jobs where R_0 is concave. In this sense, randomization serves to "iron out" concave parts of the output function.

6 Concluding Remarks

In many professional labor markets big firms command market power competing with smaller rivals for heterogeneous workers. We refer to this scenario as a quasi-monopsony. An example that fits the mold of a superstar firm having emerged victorious from a "winner takes most" contest is Amazon. From a professional labor market perspective, Amazon started to hire a significant fraction of empirical microeconomists in recent years (Athey and Luca, 2019). More traditional examples of quasi-monopsonies are regional healthcare systems in the market for registered nurses (Staiger et al., 2010) or public school districts when hiring teachers (Ransom and Sims, 2010).

The presence of such dominating firms has triggered a discussion whether market power in labor markets is a matter of grave concern for social welfare and its distribution among firms and workers. Rather than coordinating their efforts centrally, some big firms, however, choose to decentralize hiring to agents such as divisions, stores or other forms of business units. This delegation lowers the labor market power of the big firm since agents act in their own best interest and often end up competing with each other for workers.

Building on these observations, we build a novel model that explains why and when big firms choose to delegate hiring, and explains why, counter-intuitively, sometimes bigger firms are less likely to decentralize recruiting. A big firm faces a trade-off between commitment and market power. If it delegates bidding to multiple agents or divisions, the ensuing competition increases wages. In equilibrium, however, the firm's competitors anticipate this increase and lower their own wages. As a result, the big firm hires more able workers at only moderately increased wages. On the other hand, not delegating allows the firm to exercise market power and dampen wages across the market.

The big firm never delegates if its productivity is low or particularly high. A moderately productive big firm optimizes the tension between commitment and market power by choosing to delegate to a finite number of multiple agents. We show that strategic delegation always increases total surplus as well as the average wage in the market. The sole source of substantial welfare loss in our model is a big firm with an exorbitant degree of market power. This firm never delegates as the loss of market power outweighs any upside from added commitment power. In general, whenever the big firm exercises its market power, its fringe competitors also gain from the lowered market wages. If the big firm is sufficiently productive, its wage schedule overlaps with the bids of fringe firms. It follows that our model predicts different firms or divisions hiring from the same crop of workers. This markedly differs from many other models in which one firm hires strictly more able workers than another one or vice versa.

We show that the introduction of a minimum or union wage does not appear to be viable instrument of redistribution in professional labor markets with heterogeneous workers. In fact, it may even divert welfare from workers to firms. We also show in our duopsony extension that our insights are not limited to a scenario with a single big firm. Moreover, if firms prefer to pay uniform wages to an incoming workforce, our results remain valid. The same is true for general production functions.

It is certainly not our claim that our results supersede alternative findings about the welfare consequences of labor market power and minimum or union wages. We rather urge the profession not to ignore the effects of market power and competition through delegation on overall welfare and its distribution between wages and firm profits.

Finally, we would like to invite empirical studies of our testable implications for labor markets in which one or a few firms hold considerable power. Among these are: Everything else the same, (i) a dominant firm with low per worker productivity hires a distinct set of workers with lower ability (education, tenure, etc.). (ii) A dominant firm with decentralized hiring offers higher wages. (iii) The likelihood of big firms to delegate hiring is reverse U-shaped in their productivity.

Appendix: Proofs

Proof of Lemma 1. (i) Any subset of $\mathbb{R}$ with a positive measure contains a subset J with $m(J) > 0$ such that the smallest interval I for which $J \subseteq I$ satisfies $m(J) = m(I)$. Now suppose there is a set $X \subseteq \mathbb{R}$ with $m(X) > 0$ such that only agent m offers wages in X. It follows that there is an interval $(\underline{x}, \overline{x})$ such that only m offers wages in $(\underline{x}, \overline{x})$ for a positive measure of jobs. But then m cannot be optimizing since there must be an $\epsilon < \overline{x} - \underline{x}$ such that m prefers to bid for all these jobs in $(\underline{x}, \underline{x} + \epsilon)$ instead, thereby lowering the wage sum but hiring the identical set of workers ignoring measure zero cases.

(ii) Suppose that $w_0 > 0$ is the smallest wage such that wages in $[w_0, w_0 + \epsilon)$ are offered for a positive measure of jobs for an arbitrarily small $\epsilon > 0$. If no agent offers w_0 for a positive measure of jobs, and there is a fringe firm bidding arbitrarily close to w_0 from above, this agent would be better off bidding 0 instead. If there is a single agent posting w_0 for a positive measure of jobs, this agent improves as well by bidding 0 for all these jobs. If more than one agent bids w_0 for a positive measure of jobs, either one benefits by increasing its wage by an infinitesimal amount. Thus, at least two agents m and l both bid wages in wages in $[w_0, w_0 + \epsilon)$ for an arbitrarily small $\epsilon > 0$ without offering w_0 for a positive measure of jobs. Let ϵ approach 0. Then the benefit of an agent from outbidding its rival(s) in this interval approaches 0 as well. Hence, there must be an $\epsilon > 0$ such that agent m benefits from bidding 0 instead of wages in $[w_0, w_0 + \epsilon)$. If it was one agent competing with a positive measure set of fringe firms instead, the fringe firm bidding the lowest wage benefits from dropping its wage to 0.

(iii) Assume the statement is not true. Then, there exists an interval $X = (\underline{x}, \overline{x})$ such that no positive measure of jobs offers wages in X but both below and above. An appropriate variation of the argument in the proof of (ii) above applies.

(iv) Suppose a positive measure of fringe firms offers $w_0 > 0$. Each of them would benefit by raising its wage offer infinitesimally. Thus, an agent bids w_0 for a positive measure of jobs. By (iii) we know that there is a positive measure of jobs bidding wages in $(w_0 - \epsilon, w_0)$ for an arbitrarily small $\epsilon > 0$. If these jobs are offered by fringe firms, such an agent offering a wage arbitrarily close to w_0 benefits from exceeding w_0 infinitesimally with its wage bid instead. If an agent m covers the interval, there must be interval small enough right below w_0 such that exceeding w_0 for all jobs in this interval raises profits for m.

(v) Suppose 0 is offered by a positive measure of jobs. It is immediate that none of them belongs to a fringe firm. Thus, suppose multiple agents offer 0 for a positive measure of jobs. Then, there must be an $\epsilon > 0$ such that one of them would benefit from bidding ϵ instead of 0 for all these jobs. ■

Proof of Corollary 1. (i) By Lemma 1 (iii) the set of all wages offered in equilibrium has the same measure as the smallest interval that contains it. It follows that there is no gap of positive measure between the smallest and largest wages offered in equilibrium. And therefore, the support of F takes on the form of an interval. Moreover, due to Lemma 1 (ii), this interval contains values arbitrarily close to 0 and by Lemma 1 (v) may contain 0 itself. Let w_h be the supremum of the set of all wages offered in equilibrium. Without loss of generality, we can include the endpoints 0 and w_h in the support.

(ii) Since there is no gap in the support of the equilibrium wage distribution F, it follows immediately that it is strictly increasing. In addition, by Lemma 1 (iv), $F(w)$ does not exhibit any upward jumps at $w > 0$. ■

Proof of Proposition 1. Let $\left(G, \{H_m\}_{m=1}^{M}\right)$ be an equilibrium of the bidding subgame in which all agents bid identical wage schedules. Proposition 7 below establishes that this collection $\left(G, \{H_m\}_{m=1}^{M}\right)$ is a subgame equilibrium of the bidding stage if and only if it is a also subgame equilibrium when all agents can only post a single uniform wage, when H_m is reinterpreted as the cdf of the random wage w_m for all $m = 1, 2, ..., M$. Consider $\left(G, \{H_m\}_{m=1}^{M}\right)$ to be an equilibrium of this uniform wage bidding subgame. Then, the collection $\left(G', \{H'_m\}_{m=1}^M\right)$ with $H'_m = H'_n \ \forall (m,n) \in \{1,2,...,M\}^2$, $\sum_{m=1}^{M} H'_m = \sum_{m=1}^{M} H_m$ and $G' = G$ necessarily constitutes an equilibrium of the uniform wage bidding subgame as well.

By Proposition 7 this implies that $\left(G', \{H'_m\}_{m=1}^M\right)$ is also an equilibrium of the bidding subgame in which every agent posts a wage schedule. It follows that for every equilibrium of the bidding subgame there is an equivalent equilibrium in which all agents act identically. Lemmas 2 and 3 construct the equilibrium of the bidding subgame in which agents post identical wage schedules. This equilibrium is unique up to measure 0 events. The claim follows. ■

Proof of Lemma 2. Focusing on symmetric equilibrium, the FOC of (2) w.r.t. w and FOC of (3) w.r.t. $w_m(\theta)$ read:

$$\begin{aligned} (1-s)G'(w) + sH'(w) - 1 &= 0, && (7)\\ \left((1-s)G'(w_m(\theta)) + s\frac{M-1}{M}H'(w_m(\theta))\right)\mu - 1 &= 0. && (8) \end{aligned}$$

Here we temporarily ignore the constraint that $w_m(\cdot)$ is increasing, and will verify later.

(i) For $M \geqslant \frac{\mu}{\mu-1}$, conjecture that fringe firms always bid lower than agents. Lemma 1 implies that there must exist $0 < w_l < w_h$ such that G has support on $[0, w_l]$ and H has support on $[w_l, w_h]$. Applying $H'(w) = 0$ to (7) and imposing condition $G(0) = 0$, we have:

$$G'(w) = \frac{1}{1-s}, \ \forall \ w \in [0, 1-s].$$

Define $w_l \equiv 1 - s$. Applying $G'(w) = 0$ to (8) and imposing condition $H(w_l) = 0$, we have:

$$H'(w) = \frac{M}{(M-1)s\mu}, \ \forall \ w \in \left[w_l, w_l + \frac{M-1}{M}s\mu\right].$$

We verify the conjecture by the following steps. First, $G' > 0$ and $H' > 0$. Second, since H is increasing on its support, the wage schedule $w(\cdot) = H^{-1}(\cdot)$ is increasing too, satisfying the monotonicity constraint. Third, $\frac{\mathrm{d}\pi}{\mathrm{d}w} = \frac{M}{(M-1)\mu} - 1 < 0$ when $w > w_l$, so that a fringe firm does not deviate to above w_l. Finally, $\frac{\mathrm{d}\pi_m}{\mathrm{d}w(\theta)} = \mu - 1 > 0$ when $w < w_l$, so that an agent does not deviate to below w_l.

(ii) For $\frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}$, conjecture that fringe firms bid from 0 to some w_h, while agents bid from some $w_l > 0$ to w_h. Applying $H'(w) = 0$ to (7) and imposing condition $G(0) = 0$, we have:

$$G'(w) = \frac{1}{1-s}, \ \forall \ w \in [0, w_l],$$

where w_l is to be determined. For $w \in [w_l, w_h]$, solving (7) and (8) simultaneously and imposing $G(w_l) = \frac{w_l}{1-s}$, $H(w_l) = 0$, $G(w_h) = h(w_h) = 1$, we have:

$$G'(w) = \frac{\mu - M(\mu-1)}{(1-s)\mu}, \ H'(w) = \frac{M(\mu-1)}{s\mu}, \ \forall \ w \in [w_l, w_h], \tag{9}$$

where $w_l = 1 - \frac{s}{M}\frac{\mu}{\mu-1}$, $w_h = 1$. We verify the conjecture by the following steps. First, $G' > 0$ and $H' > 0$. Second, the wage schedule $w(\cdot) = H^{-1}(\cdot)$ is increasing, satisfying the monotonicity constraint. Finally, $\frac{\mathrm{d}\pi_m}{\mathrm{d}w(\theta)} = \mu - 1 > 0$ when $w < w_l$, so that an agent does not deviate to below w_l.

(iii) For $M < \frac{\mu}{\mu-1}s$, conjecture that agents bid from 0 to some w_h, while fringe firms bid from some $w_l > 0$ to w_h. Applying $G'(w) = 0$ to (7) and imposing condition $H(0_-) = 0$, we have:

$$H'(w) = \frac{M}{(M-1)s\mu}, \ \forall\, w \in [0, w_l],$$

where w_l is to be determined. For $w \in [w_l, w_h]$, solving (7) and (8) simultaneously and imposing $G(w_l) = 0$, $H(w_l) = \frac{M}{(M-1)s\mu}w_l$, $G(w_h) = H(w_h) = 1$, we have the same solution for G' and H' as in (9), and $w_l = \frac{(M-2)\mu(s\mu - M(\mu-1))}{M(\mu - M(\mu-1))}$, $w_h = \frac{M-s}{M}\mu$. We verify the conjecture by the following steps. First, $G' > 0$ and $H' > 0$. Second, the wage schedule $w(\cdot) = H^{-1}(\cdot)$ is indeed increasing. Finally, $\frac{\mathrm{d}\pi}{\mathrm{d}w} = \frac{M}{(M-1)\mu} - 1 > 0$ when $w < w_l$, so that a fringe firm does not deviate to below w_l. ■

Proof of Lemma 3. For $\mu \leqslant 1$, conjecture that fringe firms always bid higher than agents. Lemma 1 implies that there must exist $0 < w_l < w_h$ such that G has support on $[w_l, w_h]$ and H has support on $[0, w_l]$. Applying $G'(w) = 0$ to (8) and imposing condition $H(0_-) = 0$, we have:

$$H'(w) = \frac{M}{(M-1)s\mu}, \ \forall\, w \in \left[0, \frac{M-1}{M}s\mu\right].$$

Define $w_l \equiv \frac{M-1}{M}s\mu$. Applying $H'(w) = 0$ to (7) and imposing condition $G(w_l) = 0$, we have:

$$G'(w) = \frac{1}{1-s}, \ \forall\, w \in [w_l, w_l + 1 - s].$$

We verify the conjecture by the following steps. First, $G' > 0$ and $H' > 0$. Second, the wage schedule $w(\cdot) = H^{-1}(\cdot)$ is increasing. Third, $\frac{\mathrm{d}\pi_m}{\mathrm{d}w(\theta)} = \mu - 1 < 0$ when $w > w_l$, so that an agent does not deviate to above w_l. Finally, $\frac{\mathrm{d}\pi}{\mathrm{d}w} = \frac{M}{(M-1)\mu} - 1 > 0$ when $w < w_l$, so that a fringe firm does not deviate to below w_l. ■

Proof of Theorem 1. Suppose $\mu \leqslant 1$. According to Lemma 3, the per-job profit of an agent (also of the big firm due to symmetry) is obtained by letting $w(\theta) = w_l$ because the agent is indifferent among all wages on the support:

$$\left((1-s)0 + \frac{s}{M}\frac{1}{2} + \frac{(M-1)s}{M}0\right)\mu - 0 = \frac{s\mu}{2M},$$

which applies to all $M \geqslant 1$. It is decreasing in M. Hence, it is optimal to set $M = 1$.

Suppose $\mu > 4$, which means $\frac{\mu}{\mu-1} < 2$. If $M \geqslant 2$, it must belong to scenario (i) of Lemma 2. In that case, the per-job profit of an agent is:

$$\left((1-s)1 + \frac{s}{M}\frac{1}{2} + \frac{(M-1)s}{M}0\right)\mu - 1 - s = \frac{s\mu}{2M} + (1-s)(\mu-1),$$

which is decreasing in M. Therefore, we only need to compare the per-job profit when $M = 2$ and when $M = 1$. At $M = 1$, the per-job profit is

$$\left((1-s)1 + \frac{s}{M}\frac{1}{2} + \frac{(M-1)s}{M}1\right)\mu - 1 = (\mu - 1) - \frac{s\mu}{2}$$

if $s \leqslant \frac{\mu-1}{\mu}$, and is

$$\left((1-s)1 + \frac{s}{M}\frac{1}{2} + \frac{(M-1)s}{M}1\right)\mu - \frac{M-s}{M}\mu = \frac{s\mu}{2}$$

if $s > \frac{\mu-1}{\mu}$. In either case, the per-job profit for $M = 2$ is always lower than for $M = 1$ when $\mu > 4$. ■

Proof of Theorem 2. Define a sequence $\{\mu_k\}_{k=1}^{\infty}$ where $\mu_k = \frac{2k(k+1)}{2k^2-1}$, $\forall\ k$. It is strictly decreasing with $\mu_1 = 4$ and $\lim_{k\to\infty}\mu_k = 1$. Moreover, it holds for all k that $\mu_{k+1} < \frac{k+1}{k} < \mu_k$, and hence we can partition the interval $(1, 4)$ into consecutive subintervals: $\bigcup_{k=1}^{\infty}\left(\left[\mu_{k+1}, \frac{k+1}{k}\right) \cup \left[\frac{k+1}{k}, \mu_k\right)\right)$.

Define

$$\begin{aligned} M^*(\mu) &\equiv k+1, \text{ if } \mu \in [\mu_{k+1}, \mu_k), \\ s^*(\mu) &\equiv \begin{cases} \frac{2M^*(\mu)(\mu-1)}{(M^*(\mu)+1)\mu} & \text{if } \mu \in \left[\mu_{k+1}, \frac{k+1}{k}\right), \\ \frac{2M^*(\mu)(\mu-1)}{M^*(\mu)(3\mu-2)-\mu} & \text{if } \mu \in \left[\frac{k+1}{k}, \mu_k\right). \end{cases} \end{aligned}$$

In the following, I show that the delegation strategy described in the proposition is optimal. By choosing different M, the per-job profit of an agent is (see proof of Theorem 1):

$$\begin{cases} \frac{s\mu}{2M} + (1-s)(\mu-1) & \text{if } M \geqslant \frac{\mu}{\mu-1}, \\ (\mu-1) - \frac{s\mu}{2M} & \text{if } \frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}, \\ \frac{s\mu}{2M} & \text{if } M < \frac{\mu}{\mu-1}s. \end{cases}$$

Suppose an arbitrary $\mu \in (1, 4)$ is such that $\mu_{k+1} \leqslant \mu < \frac{k+1}{k} < \mu_k$ for some k. When $s < \frac{2M^*(\mu)(\mu-1)}{(M^*(\mu)+1)\mu}$, the choice $M = k+1$ dominates all M below, and the choice $M = k+2$ dominates all M above. Since $\mu < \frac{k+1}{k}$, the choice $M = k+1$ dominates $M = k+2$. When $s \geqslant \frac{2M^*(\mu)(\mu-1)}{(M^*(\mu)+1)\mu}$, the choice $M = 1$ dominates all other M. Therefore, the claimed s^* and M^* are indeed optimal.

Suppose an arbitrary $\mu \in (1, 4)$ is such that $\mu_{k+1} < \frac{k+1}{k} \leqslant \mu < \mu_k$ for some k. When $s < \frac{2M^*(\mu)(\mu-1)}{M^*(\mu)(3\mu-2)-\mu}$, the choice $M = k$ dominates all M below, and the choice $M = k+1$ dominates all M above. Since $\mu > \frac{k+1}{k}$, the choice $M = k+1$ dominates $M = k$. When $s \geqslant \frac{2M^*(\mu)(\mu-1)}{M^*(\mu)(3\mu-2)-\mu}$, the choice $M = 1$ dominates all other M. Therefore, the claimed s^* and M^* are indeed optimal.

Next, M^* is weakly decreasing by definition. s^* is continuous because its values match at all boundaries of the subintervals. s^* is strictly increasing because it is so on each subinterval. As $\mu \downarrow 1$, we must have $\lim_{\mu\downarrow 1} M^*(\mu) = \infty$ because $\lim_{k\to\infty}\mu_k = 1$. Also, $s^*\left(\frac{k+1}{k}\right) = \frac{2}{2+k} \to 0$

as $\frac{k+1}{k} \to 1$, which means $\lim_{\mu \downarrow 1} s^*(\mu) = 0$. As $\mu \uparrow 4$, we know that $\frac{1+1}{1} < \mu < \mu_1$. Hence, $M^*(\mu) = 2$ by definition. Also, $s^*(\mu) = \frac{4(\mu-1)}{2(3\mu-2)-\mu} \to \frac{3}{4}$. ■

Proof of Lemma 4. In the bidding game, if the big firm delegates to (exogenously determined) M agents, then the total surplus is:

$$\begin{cases} \frac{1}{2}(1+s(2-s)(\mu-1)) & \text{if } M \geqslant \frac{\mu}{\mu-1}, \mu > 1, \\ \frac{1}{2M}(M+2Ms(\mu-1)-s^2\mu) & \text{if } \frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}, \mu > 1, \\ \frac{\mu(1+s^2(\mu-1))-M(\mu-1)(2+2s(\mu-1)-\mu)}{2M-2(M-1)\mu} & \text{if } M < \frac{\mu}{\mu-1}s, \mu > 1, \\ \frac{1}{2}(s^2\mu+1-s^2), & \text{if } \mu \leqslant 1. \end{cases}$$

It holds that total surplus does not depend on M if $\mu < 1$. For $\mu > 1$, total surplus is increasing in M (as a continuous variable) whenever $\frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}$ or $M < \frac{\mu}{\mu-1}$. It levels off when $M \geqslant \frac{\mu}{\mu-1}$. Moreover, as M crosses $\frac{\mu}{\mu-1}s$ or $\frac{\mu}{\mu-1}$ from below, total surplus changes continuously. Restricting M to integers, total surplus is still increasing in M. ■

Proof of Proposition 2. According to Lemma 2, if $\mu > 1$, agents always bid higher than fringe firms only if $M \geqslant \frac{\mu}{\mu-1}$. If $M = 1$, this is not possible. According to Lemma 3, if $\mu \leqslant 1$, there will be only one agent (the big firm itself) and it always bids zero, acquiring the worst workers. Therefore, first best is achieved in this case.

In equilibrium, (i) if $\mu > 4$, the big firm does not delegate, according to Proposition 1. Therefore, total surplus is $\frac{\mu(1+s^2(\mu-1))-M(\mu-1)(2+2s(\mu-1)-\mu)}{2M-2(M-1)\mu}$ if $s > \frac{\mu-1}{\mu}$, and is $\frac{1}{2}(1+2s(\mu-1)-s^2\mu)$ if $s < \frac{\mu-1}{\mu}$. Both segments are increasing in s, and their values match at $s = \frac{\mu-1}{\mu}$.

(ii) if $\mu \in (1,4)$, we have either $M \geqslant \frac{\mu}{\mu-1}$ or $\frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}$ when $s < s^*(\mu)$. The former case applies when $\mu_{k+1} < \frac{k+1}{k} \leqslant \mu < \mu_k$. In this case, agents always bid higher than fringe firms, obtaining first best. Total surplus increases in s because first best does so. The latter case applies when $\mu_{k+1} \leqslant \mu < \frac{k+1}{k} < \mu_k$. In this case, total surplus falls short of first best because the bidding of agents overlaps with that of fringe firms. It has a derivative of $\mu - 1 - \frac{s\mu}{M} > \frac{(M-1)(\mu-1)}{M+1} > 0$ for $s < s^*(\mu)$. We also know that $M = 1$ when $x > s^*(\mu)$, in which case total surplus has a derivative of $(\mu-1)(1-(1-s)\mu) > \frac{(\mu-1)^2}{2\mu-1} > 0$ since $s > s^*(\mu) > \frac{2(\mu-1)}{2\mu-1}$. Total surplus exhibits a downward jump at $s^*(\mu)$ because M drops to 1.

(iii) if $\mu \leqslant 1$, total surplus has a derivative of $s(\mu-1) < 0$, meaning that it is decreasing in s. ■

Proof of Proposition 3. We divide the proof in the three cases, (i) $mu < 1$, (ii) $\mu \in (1,4)$ and $\mu > 4$.

(i) If $\mu < 1$ the equilibrium allocation is always efficient and the claim trivially holds.

(ii) For the case in which $s > s^*(\mu)$ the appropriate variant of the proof in (iii) below applies. For $s < s^*(\mu)$ the argument is as follows. If $M \geqslant \frac{\mu}{\mu-1}$, agents always outbid fringe firms. As a consequence, the claim holds trivially. As $\frac{\mu}{\mu-1}s \leqslant M$, following the proof of Proposition 2, one can show that the derivative of the relative surplus declines in s.

(iii) If $\mu > 4$ and $s < \frac{\mu-1}{\mu}$ total surplus is $\frac{1}{2}\left(1+2s(\mu-1)-s^2\mu\right)$ while the first best amounts to $\frac{1}{2}\left(1+s^2(\mu-1)\right)$. The derivative of relative surplus with respect to s is negative for all $\mu > 4$. If $s < \frac{\mu-1}{\mu}$, total surplus equals $\frac{\mu(1+s^2(\mu-1))-M(\mu-1)(2+2s(\mu-1)-\mu)}{2M-2(M-1)\mu}$. Then there exists an $\bar{s}(\mu) \in \left(\frac{\mu-1}{\mu}, 1\right)$ such that the derivative of relative surplus with respect to s is negative for $s < \bar{s}(\mu)$ and exceeds 0 for every $s > \bar{s}(\mu)$. ■

Proof of Proposition 4. (i) Given $\mu > 1$, if $s < s^*(\mu)$, then either $M \geqslant \frac{\mu}{\mu-1}$ or $\frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}$. In the former case, the per-job profit is $(1-s)(\mu-1) + \frac{s\mu}{2M}$, which has a derivative of $(1-\mu) + \frac{\mu}{2M} < -\frac{1}{2}(\mu-1)$. In the latter, the per-job profit is $(\mu-1) - \frac{s\mu}{2M}$, which has a derivative of $-\frac{\mu}{2M} < 0$. If $s > s^*(\mu)$, then $M = 1$, and the per-job profit is $\frac{s\mu}{2}$, increasing in s.

(ii) The profit of a fringe firm is zero unless the lower end of the support of G is above zero. Indeed, the profit is

$$\pi = \begin{cases} 0 & \text{if } s < s^*(\mu), \\ s\mu - (\mu-1) & \text{if } s \geqslant s^*(\mu). \end{cases}$$

Therefore, it stays at zero for lower s. For $\mu \in (1,4)$, it jumps up to $s^*(\mu)\mu - (\mu-1) > 0$ and keeps increasing in s. For $\mu \geqslant 4$, it is continuous at $s = s^*(\mu)$ and keeps increasing in s. Since $s^*(\mu)$ increases in μ, an increase in μ results in a larger set of s such that the profit for fringe firms is zero. For s still above $s^*(\mu)$, the profit has a derivative of $-(1-s) < 0$.

(iii) Given $\mu \leqslant 1$, the per-job profit of the big firm is $\frac{1}{2}s\mu$, which is increasing in s.

(iv) The profit of a fringe firm is s, which is increasing in s and independent of μ. ■

Proof of Proposition 5. (i) If $\mu \leqslant 1$, the only agent bids at zero, while fringe firms bid uniformly between 0 and $1-s$. Therefore,

$$F(w) = \begin{cases} 0 & \text{if } w < 0, \\ s + w & \text{if } 0 \leqslant w \leqslant 1-s, \\ 1 & \text{if } w > 1-s. \end{cases}$$

It weakly increases in s everywhere, meaning $F^s_\mu(w) \succsim_{FOSD} F^{s'}_\mu(w)$ whenever $s < s'$. Moreover, it does not depend on μ.

(ii) Suppose $\mu > 1$ and $s < s^*(\mu)$. If μ is such that $\mu \geqslant 4$, or $\mu < 4$ and $\mu_{k+1} \leqslant \mu < \frac{k+1}{k} < \mu_k$ for some $k \geqslant 1$, then scenario (ii) of Lemma 2 gives:

$$F(w) = \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \leqslant w \leqslant 1, \\ 1 & \text{if } w > 1. \end{cases}$$

It is independent of s for $s < s^*(\mu)$. If μ is such that $\mu < 4$ and $\mu_{k+1} < \frac{k+1}{k} \leqslant \mu < \mu_k$ for some $k \geqslant 1$, then scenario (i) of Lemma 2 gives:

$$F(w) = \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \leqslant w < 1-s, \\ \frac{Mw}{(M-1)\mu} - \frac{(1-s)(\mu-M(\mu-1))}{(M-1)\mu} & \text{if } 1-s \leqslant w \leqslant 1-s+s\mu\frac{M-1}{M}, \\ 1 & \text{if } w > 1-s+s\mu\frac{M-1}{M}. \end{cases}$$

This piece-wise linear function is everywhere decreasing in s.

(iii) Suppose $\mu > 1$ and $s \geqslant s^*(\mu)$. Scenario (iii) of Lemma 2 gives:

$$F(w) = \begin{cases} 0 & \text{if } w < 0, \\ \frac{Mw}{(M-1)\mu} & \text{if } 0 \leqslant w < \frac{(M-1)\mu(s\mu - M(\mu-1))}{M(\mu - M(\mu-1))}, \\ 1 - \mu + \frac{s\mu}{M} + w & \text{if } \frac{(M-1)\mu(s\mu - M(\mu-1))}{M(\mu - M(\mu-1))} \leqslant w \leqslant \frac{M-s}{M}\mu, \\ 1 & \text{if } w > \frac{M-s}{M}\mu. \end{cases}$$

This piece-wise linear function is everywhere increasing in s.

(iv) Let

$$\tilde{F} \equiv \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \leqslant w \leqslant 1, \\ 1 & \text{if } w > 1. \end{cases}$$

From the statements above, we know that $F^s_\mu \geqslant \tilde{F}$ if $\mu \leqslant 1$ or if $\mu > 1$ and $s > s^*(\mu)$; $F^s_\mu = \tilde{F}$ if $\mu \geqslant 4$, or if $\mu \in (1,4)$ and $\mu_{k+1} \leqslant \mu < \frac{k+1}{k} < \mu_k$ for some $k \geqslant 1$; $F^s_\mu < \tilde{F}$ if $\mu \in (1,4)$ and $\mu_{k+1} < \frac{k+1}{k} \leqslant \mu < \mu_k$ for some $k \geqslant 1$. In the last case, F^s_μ is decreasing in s for all $s < s^*(\mu)$. Fixing a $k \geqslant 1$, F^s_μ is decreasing in μ. Therefore, the infimum of F^s_μ, given k, is achieved by letting $\mu \uparrow \mu_k$ and $s \uparrow s^*(\mu_k) = \frac{1+2k}{(1+k)^2}$. Rewriting the limiting F^s_μ for these μ and s in terms of k, we have:

$$F(w) = \begin{cases} 0 & \text{if } w < 0, \\ w & \text{if } 0 \leqslant w < \frac{k^2}{(1+k)^2}, \\ \frac{Mw}{(M-1)\mu} - \frac{(1-s)(\mu - M(\mu-1))}{(M-1)\mu} & \text{if } \frac{k^2}{(1+k)^2} \leqslant w \leqslant \frac{k^2(1+4k+2k^2)}{(1+k)^2(2k^2-1)}, \\ 1 & \text{if } w > \frac{k^2(1+4k+2k^2)}{(1+k)^2(2k^2-1)}. \end{cases}$$

which in everywhere increasing in k. As a result, the infimum of F^s_μ is achieved with $k = 1$, $\mu \uparrow \mu_1 = 4$, and $s \uparrow s^*(\mu_1) = \frac{3}{4}$.

At this limiting distribution of wages, we have the average wage at $\int w \mathrm{d}F^s_\mu(w) = \frac{25}{32}$ as the supremum. ■

Proof of Lemma 6. Focusing on symmetric equilibrium, the FOC of (4) w.r.t. $w_{j,m}(\theta)$ $(j = 1, 2)$ reads:

$$\left(s_{-j} H'_{-j}(w_{j,m}(\theta)) + s_j \frac{M_j - 1}{M_j} H'_j(w_{j,m}(\theta)) \right) \mu - 1 = 0. \tag{10}$$

Again, we ignore the constraint that $w_{j,m}$ is increasing, and will verify later.

(i) For $M_1 \geqslant \frac{\mu}{\mu-1}$, conjecture that agents from firm 2 always bid lower than agents from firm 1, i.e., there exist $0 < w_l < w_h$ such that H_2 has support on $[0, w_l]$ and H_1 has support on $[w_l, w_h]$. Applying $H'_1(w) = 0$ and $j = 2$ to (10) and imposing condition $H_2(0) = 0$, we have:

$$H'_2(w) = \frac{M_2}{(M_2 - 1)(1 - s_1)}, \ \forall\, w \in [0, 1 - s_1].$$

Define $w_l \equiv \frac{M_2 - 1}{M_2}(1 - s_1)$. Applying $H'_2(w) = 0$ to (10) and imposing condition $H_1(w_l) = 0$,

we have:
$$H_1'(w) = \frac{M_1}{(M_1-1)s_1\mu}, \ \forall\ w \in \left[w_l, w_l + \frac{M_1-1}{M_1}s_1\mu\right].$$
We verify the conjecture by the following steps. First, H_j is increasing on its support, and the wage schedule $w_j(\cdot) = H_j^{-1}(\cdot)$ is increasing too, satisfying the monotonicity constraint. Second, $\frac{\mathrm{d}\pi_{2,m}}{\mathrm{d}w_{2,m}(\theta)} = \frac{M_1}{(M_1-1)\mu} - 1 < 0$ when $w_{2,m} > w_l$, so that an agent from firm 2 does not deviate to above w_l. Finally, $\frac{\mathrm{d}\pi_{1,m}}{\mathrm{d}w_{1,m}(\theta)} = \frac{M_2\mu}{M_2-1} - 1 > 0$ when $w_{1,m} < w_l$, so that an agent from firm 1 does not deviate to below w_l.

(ii) For $\frac{\mu}{\mu-1+\frac{1-s_1}{M_2}}s_1 \leqslant M_1 < \frac{\mu}{\mu-1}$, conjecture that agents from firm 2 bid from 0 to some w_h, while agents from firm 1 bid from some $w_l > 0$ to w_h. Applying $H_1'(w) = 0$ to (10) and imposing condition $H_2(0) = 0$, we have:
$$H_2'(w) = \frac{M_2}{(M_2-1)(1-s_1)}, \ \forall\ w \in [0, w_l],$$
where w_l is to be determined. For $w \in [w_l, w_h]$, solving (10) for both $j = 1, 2$ simultaneously and imposing $H_2(w_l) = \frac{M_2 w_l}{(M_2-1)(1-s_1)}, \ H_1(w_l) = 0, \ H_2(w_h) = H_1(w_h) = 1$, we have:
$$H_2'(w) = \frac{M_2(\mu - M_1(\mu-1))}{(M_1+M_2-1)(1-s_1)\mu}, \ H_1'(w) = \frac{M_1(1+M_2(\mu-1))}{(M_1+M_2-1)s_1\mu}, \ \forall\ w \in [w_l, w_h], \quad (11)$$
where $w_l \equiv \frac{(M_2-1)((1-s_1+M_2(\mu-1))M_1 - M_2 s_1\mu)}{M_1 M_2(1+M_2(\mu-1))}, \ w_h \equiv \frac{M_2-1+s_1}{M_2}$. We verify the conjecture by the following steps. First, $H_j' > 0$ on the support, and the wage schedule $w_j(\cdot) = H_j^{-1}(\cdot)$ is increasing, satisfying the monotonicity constraint. Second, $\frac{\mathrm{d}\pi_{1,m}}{\mathrm{d}w_{1,m}(\theta)} = \frac{M_2}{M_2-1}\mu - 1 > 0$ when $w < w_l$, so that an agent from firm 1 does not deviate to below w_l.

(iii) For $M_1 < \frac{\mu}{\mu-1+\frac{1-s_1}{M_2}}s_1$, conjecture that agents from firm 1 bid from 0 to some w_h, while fringe firms bid from some $w_l > 0$ to w_h. Applying $H_2'(w) = 0$ to (10) and imposing condition $H_1(0_-) = 0$, we have:
$$H_1'(w) = \frac{M_1}{(M_1-1)s_1\mu}, \ \forall\ w \in [0, w_l],$$
where w_l is to be determined. For $w \in [w_l, w_h]$, solving (10) for $j = 1, 2$ simultaneously and imposing $H_2(w_l) = 0, \ H_1(w_l) = \frac{M_1}{(M_1-1)s_1\mu}w_l, \ H_2(w_h) = H_1(w_h) = 1$, we have the same solution for H_2' and H_1' as in (11), and $w_l \equiv \frac{(M_1-1)(M_2 s_1\mu - M_1(1-s_1+M_2(\mu-1)))}{M_1 M_2(M_1-(M_1-1)\mu)}\mu, \ w_h \equiv \frac{M_1-s_1}{M_1}\mu$. We verify the conjecture by the following steps. First, H_j is strictly increasing on the support, and the wage schedule $w_j(\cdot) = H_j^{-1}(\cdot)$ is increasing too. Second, $\frac{\mathrm{d}\pi_{2,m}}{\mathrm{d}w_{2,m}(\theta)} = \frac{M_1}{(M_1-1)\mu} - 1 > 0$ when $w < w_l$, so that an agent from firm 2 does not deviate to below w_l. ■

Proof of Proposition 6. Suppose $\mu > 4$, which means $\frac{\mu}{\mu-1} < 2$. Fixing M_2, if $M_1 \geqslant 2$, firm 1's per-job profit is
$$\left((1-s_1) + \frac{s_1}{M_1}\frac{1}{2}\right)\mu - \frac{M_2-1}{M_2}(1-s_1) = \frac{s_1\mu}{2M_1} + (1-s_1)\left(\mu - \frac{M_2-1}{M_2}\right),$$

which decreases in M_1. Therefore, we only need to compare $M_1 = 1$ with $M_1 = 2$. At $M_1 = 1$, the per-job profit is

$$\begin{cases} \left(1 - \frac{s_1}{M_1}\frac{1}{2}\right)\mu - \frac{M_2-1+s_1}{M_2} = \mu - \frac{M_2-1+s_1}{M_2} - \frac{s_1\mu}{2} & \text{if } s_1 \leqslant \frac{\mu-1+1/M_2}{\mu+1/M_2}, \\ \left(1 - \frac{s_1}{M_1}\frac{1}{2}\right)\mu - \frac{M_1-s_1}{M_1}\mu = \frac{s_1\mu}{2} & \text{if } s_1 > \frac{\mu-1+1/M_2}{\mu+1/M_2}. \end{cases} \tag{12}$$

The choice of $M_1 = 1$ dominates $M_1 = 2$ in both cases if $\mu > 4$. Given firm 1's dominant strategy $M_1 = 1$, firm 2's per-job profit reads:

$$\begin{cases} \left(1 - \frac{1-s_1}{M_2}\frac{1}{2}\right) - \frac{M_2-1+s_1}{M_2} = \frac{1-s_1}{2M_2} & \text{if } s_1 \leqslant \frac{\mu-1+1/M_2}{\mu+1/M_2}, \\ \left(1 - \frac{1-s_1}{M_2}\frac{1}{2}\right) - \frac{M_1-s_1}{M_1}\mu = 1 - \mu(1-s_1) - \frac{1-s_1}{2M_2} & \text{if } s_1 > \frac{\mu-1+1/M_2}{\mu+1/M_2}. \end{cases}$$

Given the piece-wise objective function, it is optimal for firm 2 to set $M_2 = 1$ if $s_1 \leqslant \frac{2\mu-1}{2\mu+1}$, and $M_2 = \infty$ if $s_1 > \frac{2\mu-1}{2\mu+1}$.

For $\mu \in (1, 4)$, define a sequence $\{\mu_k\}_{k=1}^{\infty}$ and functions s^* and M^* as we do in the proof of Theorem 2. In addition, define

$$s^{**}(\mu) \equiv \begin{cases} \frac{2M^*(\mu)\mu}{M^*(\mu)(\mu+2)+\mu} & \text{if } \mu \in \left[\mu_{k+1}, \frac{k+1}{k}\right), \\ \frac{2M^*(\mu)}{3M^*(\mu)-1} & \text{if } \mu \in \left[\frac{k+1}{k}, \mu_k\right). \end{cases}$$

Algebra shows that $s^*(\mu) < s^{**}(\mu)$ for every $\mu \in (1, 4)$, and that s^{**} is weakly increasing in μ.

If $s < s^*(\mu)$, then the per-job profit of firm 1 when delegating to $M^*(\mu)$ agents is

$$\begin{cases} \mu - \frac{M_2-1+s_1}{M_2} - \frac{s_1\mu}{2M^*(\mu)} & \text{if } \mu \in \left[\mu_{k+1}, \frac{k+1}{k}\right), \\ \frac{s_1\mu}{2M^*(\mu)} + (1-s_1)\left(\mu - \frac{M_2-1}{M_2}\right) & \text{if } \mu \in \left[\frac{k+1}{k}, \mu_k\right). \end{cases} \tag{13}$$

Compared to (12), delegating to $M^*(\mu)$ agents is better than not delegating when $s_1 < s^*(\mu)$, regardless of M_2. On the other hand, firm 2 compares $M_2 = 1$ with $M_2 = \infty$, given the fact that $M_1 = M^*(\mu) \geqslant 2$ and $s_1 < s^*(\mu)$. Algebra shows that $M_2 = 1$ is always preferred. Therefore, the unique equilibrium is $M_1 = M^*(\mu)$, $M_2 = 1$.

If $s_1 > s^{**}(\mu)$, then it holds for all M_2 that $s_1 > \frac{\mu-1+1/M_2}{\mu+1/M_2}$. By not delegating, the per-job profit of firm 1 is $\frac{s_1\mu}{2}$. By delegating to $M_1 \geqslant 2$ such that scenario (i) or (ii) applies, then according to (13), firm 1 receives lower profit than $M_1 = 1$ regardless of M_2. On the other hand, firm 2 compares $M_2 = 1$ with $M_2 = \infty$, given the fact that $M_1 = 1$ and $s_1 > s^{**}(\mu) > \frac{\mu-1+1/M_2}{\mu+1/M_2}$. The per-job profit is $1-\mu(1-s_1)-\frac{1-s_1}{2M_2}$, increasing in M_2. Therefore, the unique equilibrium is $M_1 = 1$, $M_2 = \infty$. ■

Proof of Proposition 7. "Only if" part. Suppose $\left(G, \{H_m\}_{m=1}^{M}\right)$ is a Nash equilibrium in the bidding stage of the main model. A fringe firm choosing w in the support of G always maximizes (2), while choosing a wage outside the support does not improve. Since the right-hand side of (2) is the same as the right-hand side of (5), the optimality is guaranteed. An agent m choosing an increasing wage schedule $w_m = H_m^{-1}$ maximizes (3), where the FOC w.r.t. any particular $w_m(\theta)$ on the support of H_m requires $(1-s)G'(w_m(\theta))+\frac{s}{M}\sum_{m'\neq m}H'_{m'}(w_m(\theta))=\frac{1}{\mu}$. Since this holds for all $w_m(\theta)$ on the support, we have

$$(1-s)(G(w_m(\theta))-G(w))+\frac{s}{M}\sum_{m'\neq m}(H_{m'}(w_m(\theta))-H_{m'}(w))=\frac{1}{\mu}(w_m(\theta)-w),$$

for any w on the support of H_m. Therefore, by replacing the wage schedule $w_m(\theta)$ with a single wage w on the support, the agent m achieves the same profit, which is optimal given G and $\{H_{m'}\}_{m'\neq m}$. Due to indifference, agent m randomizes w exactly according to H_m.

"If" part. Suppose $\left(G,\{H_m\}_{m=1}^M\right)$ is a mixed strategy Nash equilibrium in the bidding stage of extension 5.2. That is, a fringe firm choosing w in the support of G always maximizes (5), while choosing a wage outside the support does not improve. Since the right-hand side of (5) is the same as the right-hand side of (2), the optimality is guaranteed. An agent m choosing a single wage w_m maximizes (6), and is indifferent among all wages on the support of H_m. In particular, a wage schedule $w_m(\cdot)=H_m^{-1}(\cdot)$ is a convex combination of all wages on the support, also achieving maximum profit given the wage distributions G and $\{H_{m'}\}_{m'\neq m}$. ■

Proof of Lemma 5. Before finding the optimal delegation strategy in the first stage, we describe the equilibrium of the bidding game given parameters. For any fixed $\mu\in(1,4)$, when varying M and s, we have seven scenarios.

(1) If $M\geqslant\frac{\mu}{\mu-1}$ and $s<1-\underline{w}$, then conjecture that agents always bid higher than fringe firms, agents fill their capacity, and some fringe firms do not hire. Therefore, there exist $\underline{w}<w_l<w_h$ such that G has support on $[\underline{w},w_l]$ and H has support on $[w_l,w_h]$. Along with the usual FOC, we also impose the "free entry" condition that all fringe firms earn zero profit. As a result, $w_l\equiv 1-s$, $w_h\equiv w_l+\frac{M-1}{M}s\mu$, $G'(w)=\frac{1}{1-s-\underline{w}}$ for $w\in[\underline{w},w_l]$, and $H'(w)=\frac{M}{(M-1)s\mu}$ for $w\in[w_l,w_h]$. The per-job profit of an agent is $\mu-1-s+\frac{s\mu}{2M}$. The measure of unemployed workers is $\underline{w}$. The total output is $\frac{1}{2}(1+s(2-s)(\mu-1))-\frac{\underline{w}^2}{2}$.

(2) If $\frac{s\mu}{\mu-1}\frac{1}{1-\underline{w}}\leqslant M<\frac{\mu}{\mu-1}$, then conjecture that fringe firms bid from $\underline{w}$ to some w_h while agents bid from some $w_l>\underline{w}$ to w_h, agents fill their capacity, and some fringe firms do not hire. Therefore, G has support on $[\underline{w},w_h]$ and H has support on $[w_l,w_h]$. The free entry condition implies that all fringe firms earn zero profit. As a result, $w_l\equiv\frac{M(\mu-1)-s\mu}{M(\mu-1)}$, $w_h\equiv 1$, $G'(w)=\frac{1}{1-s-\underline{w}}$ for $w\in[\underline{w},w_l]$, $G'(w)=\frac{\mu-M(\mu-1)}{\mu(1-s-\underline{w})}$ and $H'(w)=\frac{M(\mu-1)}{s\mu}$ for $w\in[w_l,w_h]$. The per-job profit of an agent is $\mu-1-\frac{s\mu}{2M}$. The measure of unemployed workers is again $\underline{w}$. The total output is $\frac{1}{2}+s(\mu-1)-\frac{s^2\mu}{2M}-\frac{\underline{w}^2}{2}$.

(3) If $\frac{s\mu}{\mu-1}\leqslant M<\frac{s\mu}{\mu-1}\frac{1}{1-\underline{w}}$ and $s<\frac{M(1-\underline{w})}{\mu(N-1)}$, then conjecture that fringe firms bid from some $w_l>\underline{w}$ to some w_h while agents bid from $\underline{w}$ to w_h, agents fill their capacity, and some fringe firms do not hire. Therefore, G has support on $[w_l,w_h]$ and H has support on $[\underline{w},w_h]$. The free entry condition implies that all fringe firms earn zero profit. As a result, $w_l\equiv\frac{M\underline{w}-(M-1)(M(\mu-1)-s\mu)}{M(M-(M-1)\mu)}$, $w_h\equiv 1$, $H'(w)=\frac{M}{(M-1)s\mu}$ for $w\in[\underline{w},w_l]$, $G'(w)=\frac{M(M-(M-1)\mu)}{M(1-\underline{w})-(M-1)s\mu}$ and $H'(w)=\frac{M(\mu-1)}{s\mu}$ for $w\in[w_l,w_h]$. The per-job profit of an agent is again $(\mu-1)-\frac{s\mu}{2M}$. The measure of unemployed workers is now $1-\frac{s}{M}-\frac{1}{\mu}+\frac{\underline{w}}{\mu}\in(0,1-s)$. The total surplus is

$\frac{s^2\mu(1-M(\mu-1)(M-1))}{2(\mu(M-1)-M)M^2} + \frac{(s((\mu-1)(-1+(\mu-1)(M-1)M)-\underline{w})}{(\mu(M-1)-M)M} + \frac{(1-\underline{w})((1+\mu(M-2)-M)+(-1+(\mu-1)M)\underline{w})}{2\mu(\mu(M-1)-M)}$.

(4) If $\frac{s\mu-\underline{w}}{\mu-1} \leqslant M < \frac{s\mu}{\mu-1}$ and $M < \frac{\mu-\underline{w}}{\mu-1}$, conjecture that fringe firms bid from some $w_l > \underline{w}$ to some w_h while agents bid from $\underline{w}$ to w_h, agents do not fill their capacity, and some fringe firms do not hire. Therefore, G has support on $[w_l, w_h]$ and H has support on $[\underline{w}, w_h]$. The free entry condition implies (a) All fringe firms earn zero profit, and (b) agents earn zero profit from the lowest wage they offer. As a result, $w_l \equiv \frac{\underline{w}}{\mu-M(\mu-1)}$, $w_h \equiv 1$, $H'(w) = \frac{1}{(M-1)(\mu-1)}$ for $w \in [\underline{w}, w_l]$, $G'(w) = \frac{M-(M-1)\mu}{M-(M-1)\mu-\underline{w}}$ and $H'(w) = 1$ for $w \in [w_l, w_h]$. Unemployed workers come from two sources: $M\frac{\mu-1}{\mu} - a + \frac{\underline{w}}{\mu}$ are given up by fringe firms, and $a - M\frac{\mu-1}{\mu}$ are given up by agents. The per-job profit of an agent is now $\frac{1}{2}(\mu-1)$. In total, the measure of unemployed workers is $\frac{\underline{w}}{\mu}$. The total surplus is $\frac{\mu+(\mu-1)^2M}{2\mu} + \frac{(1+M-\mu M)\underline{w}^2}{2\mu(\mu(M-1)-M)}$.

(5) If $M < \frac{s\mu-\underline{w}}{\mu-1}$, then conjecture that fringe firms bid from some $w_l > \underline{w}$ to some w_h while agents bid from $\underline{w}$ to w_h, agents do not fill their capacity, and all fringe firms hire. Therefore, G has support on $[w_l, w_h]$ and H has support on $[\underline{w}, w_h]$. The free entry condition requires agents to earn zero profit from the lowest wage they offer. As a result, $w_l \equiv \frac{(M-1)\mu(s\mu-M(\mu-1))}{M(\mu-M(\mu-1))} + \frac{\underline{w}}{M}$, $w_h \equiv \frac{M-s}{M}\mu + \frac{\underline{w}}{M}$, $H'(w) = \frac{M}{(M-1)(s\mu-\underline{w})}$ for $w \in [\underline{w}, w_l]$, $G'(w) = \frac{M-(M-1)\mu}{(1-s)\mu}$ and $H'(w) = \frac{M(\mu-1)}{s\mu-\underline{w}}$ for $w \in [w_l, w_h]$. The per-job profit of an agent is now $\frac{s\mu-\underline{w}}{2M}$. The measure of unemployed workers is again $\frac{\underline{w}}{\mu}$. The total surplus is $\frac{\mu+(\mu-1)^2M}{2\mu} + \frac{(1+M-\mu M)\underline{w}^2}{2\mu(\mu(M-1)-M)}$.

(6) If $1-\underline{w} \leqslant s < 1-\frac{\underline{w}}{\mu}$ and $s \geqslant \frac{M(1-\underline{w})}{\mu(M-1)}$, then conjecture that fringe firms drop out of the market while agents bid from $\underline{w}$ to some w_h with full capacity. Therefore, H has support on $[\underline{w}, w_h]$. As a result, $w_h \equiv \frac{s\mu(M-1)}{M} + \underline{w}$, $H'(w) = \frac{M}{s\mu(M-1)}$ for $w \in [\underline{w}, w_h]$. The per-job profit of an agent is now $(1-s)\mu + \frac{s\mu}{2M} - \underline{w}$. The measure of unemployed workers is $1-s$, which is the measure of fringe firms. The total surplus is $\left(1-\frac{1}{2}s\right)s\mu$.

(7) If $s \geqslant 1-\frac{\underline{w}}{\mu}$ and $M \geqslant \frac{\mu-\underline{w}}{\mu-1}$, then conjecture that fringe firms drop out of the market while agents bid from $\underline{w}$ to some w_h with some idle capacity. Therefore, H has support on $[\underline{w}, w_h]$. As a result, $w_h \equiv \frac{s\mu(M-1)}{M} + \underline{w}$, $H'(w) = \frac{M}{s\mu(M-1)}$ for $w \in [\underline{w}, w_h]$. The per-job profit of an agent is now $\frac{\mu-\underline{w}}{2M}$. The measure of unemployed workers is $\frac{\underline{w}}{\mu}$. The total surplus is $\frac{\mu}{2} - \frac{\underline{w}^2}{2\mu}$.

The proof for optimal delegation decision is similar to the proof of Theorem 2. Again, define $\{\mu_k\}_{k=1}^{\infty}$ where $\mu_k = \frac{2k(k+2)}{2k^2-1}$, and partition the interval $(1,4)$ into consecutive subintervals: $(1,4) = \bigcup_{k=1}^{\infty}\left(\left[\mu_{k+1}, \frac{k+1}{k}\right) \cup \left[\frac{k+1}{k}, \mu_k\right)\right)$. Define:

$$
\begin{aligned}
M^*(\mu) &\equiv k+1, \text{ if } \mu \in [\mu_{k+1}, \mu_k), \\
s^*_{\underline{w}}(\mu) &\equiv \begin{cases} \frac{2M^*(\mu)(\mu-1)+M^*(\mu)\underline{w}}{(M^*(\mu)+1)\mu} & \text{if } \mu \in \left[\mu_{k+1}, \frac{k+1}{k}\right), \\ \frac{2M^*(\mu)(\mu-1)+M*(\mu)\underline{w}}{M^*(\mu)(3\mu-2)-\mu} & \text{if } \mu \in \left[\frac{k+1}{k}, \mu_k\right), k \geqslant 2, \text{ or } \mu \in [2, \tilde{\mu}), \\ \frac{2(\mu-1)}{3\mu-4} & \text{if } \mu \in [\tilde{\mu}, 4). \end{cases}
\end{aligned}
$$

where $\tilde{\mu} \equiv \frac{5-3\underline{w}+\sqrt{9-14\underline{w}+9\underline{w}^2}}{2} \in (2,4)$. Notice that $0 < s^*_{\underline{w}}(\mu) < 1$ for $\underline{w} \in \left(0, \frac{2}{9}\right)$ because $\frac{2M^*(\mu)(M^*(\mu)+1)}{2M^*(\mu)^2-1} \leqslant \mu \leqslant \frac{2M^*(\mu)(M^*(\mu)-1)}{2M^*(\mu)^2-1}$. As a result, $s^*_{\underline{w}}(\mu)$ increases in $\underline{w}$ for all $\mu \in (1,4)$.

Given the seven scenarios which exhaust all possibilities, it is pure algebra to verify that the delegation decision described by M^* and $s^*_{\underline{w}}$ is optimal. ■

Proof of Theorem 3. As $\underline{w}$ grows, there are two effects on the unemployment rate. First, *fixing* M, the unemployment rate increases with $\underline{w}$. Second, the *endogenous* choice of M by the big firm is more likely to be $M^*(\mu)$ instead of 1 when $\underline{w}$ grows. The two effects work in the same direction.

Specifically, according to the seven scenarios in the proof of Lemma 5, the unemployment rate is:

$$\begin{cases} \underline{w} & \text{in Scenarios (1) (2),} \\ 1-\frac{s}{M}-\frac{1}{\mu}+\frac{\underline{w}}{\mu} & \text{in Scenario (3),} \\ 1-s & \text{in Scenario (6),} \\ \frac{\underline{w}}{\mu} & \text{in Scenarios (4) (5) (7).} \end{cases} \tag{14}$$

Each segment of the above is increasing in $\underline{w}$, and the entire function is continuous in (μ, M). By Lemma 5, we know that unemployment is $\underline{w}$ (Scenarios (1) or (2)) when $s < s^*_{\underline{w}}(\mu)$, and is $\frac{\underline{w}}{\mu}$ (Scenarios (4) or (5)) when $s > s^*_{\underline{w}}(\mu)$. Now suppose $\underline{w}' > \underline{w}$. Either M stays the same, in which case unemployment increases from $\underline{w}$ to $\underline{w}'$ or from $\frac{\underline{w}}{\mu}$ to $\frac{\underline{w}'}{\mu}$; or M jumps from 1 to $M^*(\mu)$, in which case unemployment increases from $\frac{\underline{w}}{\mu}$ to $\underline{w}'$.

The total surplus could increase when $\underline{w}$ grows. To see this, fix $\mu \in (1,4)$ and fix an $s = s^*_{\underline{w}}(\mu)$ for some $\underline{w} > 0$. Since $s^*_{\underline{w}}(\mu)$ increases in $\underline{w}$, we know that $s > s^*_{\underline{w}'}(\mu)$ and $M = 1$ for all $\underline{w}' < \underline{w}$, and $s < s^*_{\underline{w}'}(\mu)$ and $M = M^*(\mu) > 1$ for all $\underline{w}' > \underline{w}$. As $\underline{w}'$ crosses $\underline{w}$ from below, M jumps from 1 to $M^*(\mu)$, resulting in a discrete increase in total surplus because it is increasing in M. Due to the jump, there exists a $\delta > 0$ such that the total surplus increases as $\underline{w}'$ moves from $(\underline{w}-\delta, \underline{w})$ to $(\underline{w}, \underline{w}+\delta)$. ■

Proof of Proposition 8. "If" part is shown by construction. FOC of (2) w.r.t. w and FOC of (3) w.r.t. $w_m(\theta)$ read:

$$\begin{aligned} (1-s)G'(w) + sH'(w) &= \frac{1}{R_0'((1-s)G(w)+sH(w))}, \\ (1-s)G'(w_m(\theta)) + s\frac{M-1}{M}H'(w_m(\theta)) &= \frac{1}{\mu R_0'\left((1-s)G(w_m(\theta)) + s\frac{M-1}{M}H(w_m(\theta)) + \frac{s}{M}\theta\right)}. \end{aligned}$$

In the following, we reconstruct the bidding equilibrium in Lemmas 2 and 3 for all parameter ranges, except that R_0 is non-linear.

(i) For $\mu > 1$ and $M \geqslant \frac{\mu}{\mu-1}$, let $w_l \equiv R_0(1-s)$ and $w_h \equiv w_l + \frac{M-1}{M}\mu(R_0(1) - R_0(1-s))$. Then:

$$\begin{aligned} G(w) &= \frac{R_0^{-1}(w)}{1-s}, & w \in [0, w_l], \\ H(w) &= \frac{1}{s\mu}R_0^{-1}\left(\frac{M}{M-1}(w-w_l) + \mu R_0(1-s)\right) - \frac{1-s}{s}, & w \in [w_l, w_h]. \end{aligned}$$

(ii) For $\mu > 1$ and $\frac{\mu}{\mu-1}s \leqslant M < \frac{\mu}{\mu-1}$, let $w_l \equiv R_0\left(1 - \frac{s}{M}\frac{\mu}{\mu-1}\right)$ and $w_h \equiv R_0(1)$. Then:

$$
\begin{aligned}
G(w) &= \begin{cases} \frac{R_0^{-1}(w)}{1-s}, & w \in [0, w_l], \\ G(w_l) + \frac{\mu - M(\mu-1)}{\mu}\frac{R_0^{-1}(w) - R_0^{-1}(w_l)}{1-s}, & w \in (w_l, w_h], \end{cases} \\
H(w) &= \frac{M(\mu-1)}{\mu}\frac{R_0^{-1}(w) - R_0^{-1}(w_l)}{s}, \qquad w \in [w_l, w_h].
\end{aligned}
$$

(iii) For $\mu > 1$ and $M < \frac{\mu}{\mu-1}s$, let $w_l \equiv \frac{(M-1)\mu}{M}R_0\left(\frac{s\mu - M(\mu-1)}{\mu - M(\mu-1)}\right)$ and $w_h \equiv w_l + R_0(1) - R_0\left(\frac{s\mu - M(\mu-1)}{\mu - M(\mu-1)}\right)$. Then:

$$
\begin{aligned}
G(w) &= \frac{(\mu - M(\mu-1))R_0^{-1}\left(w - w_l + R_0\left(\frac{s\mu - M(\mu-1)}{\mu - M(\mu-1)}\right)\right) - (s\mu - M(\mu-1))}{(1-s)\mu}, \qquad w \in [w_l, w_h], \\
H(w) &= \begin{cases} \frac{R_0^{-1}\left(\frac{M}{M-1}w\right)}{s\mu}, & w \in [0, w_l], \\ 1 - \frac{M(\mu-1)}{s\mu}(1 - R_0^{-1}(w - w_l + R_0(sH(w_l)))), & w \in (w_l, w_h]. \end{cases}
\end{aligned}
$$

(iv) For $\mu \leqslant 1$, let $w_l \equiv \frac{M-1}{M}\mu R_0(s)$ and $w_h \equiv w_l + R_0(1) - R_0(s)$. Then:

$$
\begin{aligned}
G(w) &= \frac{R_0^{-1}(w - w_l + R_0(s)) - s}{1-s}, \quad w \in [w_l, w_h], \\
H(w) &= \frac{R_0^{-1}\left(\frac{Mw}{(M-1)\mu}\right)}{s}, \quad w \in [0, w_l].
\end{aligned}
$$

Verification of the equilibrium follows the same steps, although an important key difference is that an agent is no longer indifferent among all wages in the support. Instead, we show that posting $w(\theta)$ for the $\theta-$ quantile job is strictly better than posting $w(\theta')$. The integrand of the objective for θ, when posting wage $w(\theta')$, reads:

$$
\mu R_0(1 - s + \frac{s}{M}(\theta + (M-1)H(w(\theta')))) - w(\theta') = \mu R_0(1 - s + \frac{s}{M}(\theta + (M-1)\theta')) - H^{-1}(\theta').
$$

Taking derivative w.r.t. θ', we have

$$
\begin{aligned}
&\frac{\left((1-s)G'(H^{-1}(\theta')) + s\frac{M-1}{M}H'(H^{-1}(\theta'))\right)\mu R_0'\left((1-s)G(H^{-1}(\theta')) + \frac{s}{M}\theta + \frac{M-1}{M}\theta'\right) - 1}{H'(H^{-1}(\theta'))} \\
&= \frac{1}{H'(H^{-1}(\theta'))}\left(\frac{R_0'\left((1-s)G(H^{-1}(\theta')) + \frac{s}{M}\theta + \frac{M-1}{M}\theta'\right)}{R_0'((1-s)G(H^{-1}(\theta')) + s\theta')} - 1\right),
\end{aligned}
$$

where the equality comes from the FOC. Since $H' > 0$ on its support and $R_0'(\cdot)$ is increasing, we know the above is negative when $\theta' > \theta$, and is positive when $\theta' < \theta$. Therefore, the proposed equilibrium is verified.

"Only if" part. Consider any agent. In order for its $\theta-$ quantile wage not to post wage

$w(\theta')$, we need the following conditions to hold for any $\theta, \theta' \in [0,1]$:

$$\begin{aligned}
& \mu R_0\left((1-s)G(w(\theta)) + s\frac{M-1}{M}H(w(\theta)) + \frac{s}{M}\theta\right) - w(\theta) \\
\geqslant\ & \mu R_0\left((1-s)G(w(\theta')) + s\frac{M-1}{M}H(w(\theta')) + \frac{s}{M}\theta\right) - w(\theta'), \\
& \mu R_0\left((1-s)G(w(\theta')) + s\frac{M-1}{M}H(w(\theta')) + \frac{s}{M}\theta'\right) - w(\theta') \\
\geqslant\ & \mu R_0\left((1-s)G(w(\theta)) + s\frac{M-1}{M}H(w(\theta)) + \frac{s}{M}\theta'\right) - w(\theta).
\end{aligned}$$

Adding up these inequalities, dividing by $(\theta' - \theta)^2$, taking limit as $\theta' \to \theta$ and plugging in FOC, we have:

$$\frac{s}{MH'(w(\theta))}\frac{R_0''((1-s)G(w(\theta)) + s\theta)}{R_0'((1-s)G(w(\theta)) + s\theta)} \geqslant 0.$$

Since $H' > 0$ on support, $R_0' > 0$ by assumption, we have $R_0'' \geqslant 0$.

■

References

AON Hewitt (2010). *2010 State of Hiring Survey*. London: AON Hewitt.

Arnold, D. (2020). Mergers and acquisitions, local labor market concentration, and worker outcomes. *mimeo*.

Ashenfelter, O. C., H. Farber, and M. R. Ransom (2010). Labor market monopsony. *Journal of Labor Economics 28*, 203–210.

Athey, S. and M. Luca (2019). Economists (and economics) in tech companies. *Journal of Economic Perspectives 33*, 209–230.

Ausubel, L. M., P. Cramton, M. Pycia, M. Rostek, and M. Weretka (2014). Demand reduction and inefficiency in multi-unit auctions. *Review of Economic Studies 81*, 1366–1400.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2019). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics (forthcoming)*.

Azar, J. A., I. Marinescu, M. I. Steinbaum, and B. Taska (2018). Concentration in US labor markets: Evidence from online vacancy data. *NBER Working Paper*.

Azevedo, E. M. (2014). Imperfect competition in two-sided matching markets. *Games and Economic Behavior 83*, 207–223.

Azevedo, E. M. and J. D. Leshno (2016). A supply and demand framework for two-sided matching markets. *Journal of Political Economy 124*, 1235–1268.

Baye, M. R., K. J. Crocker, and J. Ju (1996a). Divisionalization and franchising incentives with integral competing units. *Economics Letters 50*, 429–435.

Baye, M. R., K. J. Crocker, and J. Ju (1996b). Divisionalization, franchising and divestiture incentives in oligopoly. *The American Economic Review 86*, 223–236.

Bhaskar, V., A. Manning, and T. To (2002). Oligopsony and monopsonistic competition in labor markets. *Journal of Economic Perspectives 16*, 155–174.

Bhaskar, V. and T. To (1999). Minimum wages for Ronald McDonald monopsonies: A theory of monopsonistic competition. *The Economic Journal 109*, 190–203.

Birkinshaw, J. (2002). Strategies for managing internal competition. *IEEE Engineering Management Review 30*, 90–99.

Birkinshaw, J., N. Hood, and S. Young (2005). Subsidiary entrepreneurship, internal and external competitive forces, and subsidiary performance. *International Business Review 14*, 227–248.

Boal, W. M. and M. R. Ransom (1997). Monopsony in the labor market. *Journal of Economic Literature 35*, 86–112.

Board, S., M. Meyer-ter Vehn, and T. Sadzik (2020). Recruiting talent. *mimeo*.

Buerhaus, P. I. and D. O. Staiger (1996). Managed care and the nurse workforce. *Journal of the American Medical Association 276*, 1487–1493.

Bulow, J. and J. Levin (2006). Matching and price competition. *The American Economic Review 96*, 652–668.

Burdett, K., S. Shi, and R. Wright (2001). Pricing and matching with frictions. *Journal of Political Economy 109*, 1060–1085.

Creane, A. and C. Davidson (2004). Multidivisional firms, internal competition and the merger paradox. *Canadian Journal of Economics 37*, 951–977.

Fauli-Oller, R. and M. Giralt (1995). Competition and cooperation within a multidivisional firm. *The Journal of Industrial Economics 43*, 77–99.

Fershtman, C. and K. L. Judd (1987). Equilibrium incentives in oligopoly. *American Economic Review 77*, 927–940.

Fleishman, G. (2019, June 7). Monopsony gives tech giants enormous power—but could be their undoing. *Fast Company*.

Galenianos, M., P. Kircher, and G. Virag (2011). Market power and efficiency in a search model. *International Economic Review 52*, 85–103.

Galunic, D. C. and K. M. Eisenhardt (1996). The evolution of intracorporate domains: Divisional charter losses in high-technology, multidivisional corporations. *Organization Science 7*, 255–282.

Gregory, M. (2019). *7 Pros and Cons of Decentralized Hiring and Why Your ATS is Responsible.* https://www.smartrecruiters.com/blog/pros-and-cons-of-decentralized-hiring (accessed October 22, 2020).

Jungbauer, T. (2021). Strategic bidding with fixed quotas. *mimeo*.

Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives 14*, 75–94.

Kojima, F. (2006). Mixed strategies in games of capacity manipulation in hospital-intern markets. *Social Choice and Welfare 27*, 25–28.

Konishi, H. and U. Unver (2006). Games of capacity manipulation in hospital-intern markets. *Social Choice and Welfare 27*, 3–24.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics 108*, 551–575.

Landon, J. H. and R. N. Baird (1971). Monopsony in the market for public school teachers. *The American Economic Review 61*, 966–979.

Lennon, C. (2019). *My Small Business Doesn't Have an HR Department—Is That OK?* https://www.allbusiness.com/my-small-business-doesnt-have-hr-department-114005-1.html (accessed October 25, 2020).

Levin, D. (2004). The competitiveness of joint bidding in multi-unit uniform-price auctions. *RAND Journal of Economics 35*, 373–385.

Lindzon, J. (2020, March 20). How many positions should you apply for at one company? *Fast Company*.

Manning, A. (2006). A generalised model of monopsony. *The Economic Journal 116*, 84–100.

McNeal, M. (2001, October 11). Competition for your internal candidates. *ERE Recruiting Intelligence*.

Mercer (2017). *HR Structures Today.* New York: Mercer LLC.

Niederle, M., D. D. Proctor, and A. E. Roth (2006). What will be needed for the new gastroenterology fellowship match to succeed? *Gastroenterology 130*(1), 218–224.

Polasky, S. (1992). Divide and conquer: On the profitability of forming independent rival divisions. *Economics Letters 40*, 365–371.

Prager, E. and M. Schmitt (2020). Employer consolidation and wages: Evidence from hospitals. *mimeo*.

Ransom, M. R. (1993). Seniority and monopsony in the academic labor market. *The American Economic Review 83*, 221–233.

Ransom, M. R. and D. P. Sims (2010). Estimating the firm's labor supply curve in a "new monopsony" framework: Schoolteachers in Missouri. *Journal of Labor Economics 28*, 331–355.

Salant, S. W., S. Switzer, and R. J. Reynolds (1983). Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium. *The Quarterly Journal of Economics 98*, 185–199.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge: Harvard University Press.

Schwartz, M. and E. A. Thompson (1986). Divisionalization and entry deterrence. *The Quarterly Journal of Economics 101*, 307–321.

Schwenen, S. (2015). Strategic bidding in multi-unit auctions with capacity constrained bidders: the New York capacity market. *RAND Journal of Economics 46*, 730–750.

Sengul, M., J. Gimeno, and J. Dial (2012). Strategic delegation: A review, theoretical integration, and research agenda. *Journal of Management 38*, 375–414.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. von Wachter (2019). Firming up inequality. *The Quarterly Journal of Economics 134*, 1–50.

Sonmez, T. (1997). Manipulation via capacities in two-sided matching markets. *Journal of Economic Theory 77*, 197–204.

Staiger, D. O., J. Spetz, and C. S. Phibbs (2010). Is there monopsony in the labor market? Evidence from a natural experiment. *Journal of Labor Economics 28*, 211–236.

Steakley, D. (2013). *Big corporations are just slow to act.* https://berkonomics.com/?p=1861 (accessed February 27, 2020).

van den Berg, G. J. and G. Ridder (1998). An empirical equilibrium search model of the labor market. *Econometrica 66*, 1183–1221.

Veendorp, E. C. H. (1991). Entry deterrence, divisionalization, and investment decisions. *The Quarterly Journal of Economics 106*, 297–307.

Yello (2019). *How to Execute Campus Recruitment Events with a Decentralized Team.* https://yello.co/blog/centralize-event-recruiting-decentralized-teams (accessed February 12, 2020).