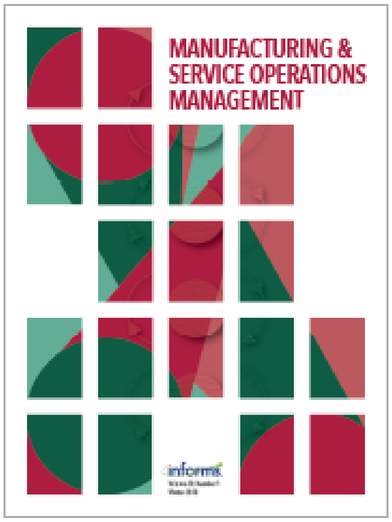


Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Managing Congestion in Matching Markets

Nick Arnosti, Ramesh Johari, Yash Kanoria

informs
http://pubsonline.informs.org/journal/msom
MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 23, No. 3, May-June 2021, pp. 620–636
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Managing Congestion in Matching Markets

Nick Arnosti,[a] Ramesh Johari,[b] Yash Kanoria[a]

[a] Columbia Business School, Columbia University, New York, New York 10027; [b] Department of Management Science and Engineering, Stanford University, Stanford, California 94305

Contact: n.arnosti@columbia.edu, https://orcid.org/0000-0002-6685-1428 (NA); rjohari@stanford.edu, https://orcid.org/0000-0002-3960-0770 (RJ); ykanoria@columbia.edu, https://orcid.org/0000-0002-7221-357X (YK)

Received: October 5, 2018
Revised: June 3, 2019; November 5, 2019
Accepted: November 10, 2019
Published Online in Articles in Advance: January 19, 2021

https://doi.org/10.1287/msom.2020.0927

Abstract. ***Problem definition***: Participants in matching markets face search and screening costs when seeking a match. We study how platform design can reduce the effort required to find a suitable partner. ***Practical/academic relevance***: The success of matching platforms requires designs that minimize search effort and facilitate efficient market clearing. ***Methodology***: We study a game-theoretic model in which "applicants" and "employers" pay costs to search and screen. An important feature of our model is that both sides may waste effort: Some applications are never screened, and employers screen applicants who may have already matched. We prove existence and uniqueness of equilibrium and characterize welfare for participants on both sides of the market. ***Results***: We identify that the market operates in one of two regimes: It is either *screening-limited* or *application-limited*. In screening-limited markets, employer welfare is low, and some employers choose not to participate. This occurs when application costs are low and there are enough employers that most applicants match, implying that many screened applicants are unavailable. In application-limited markets, applicants face a "tragedy of the commons" and send many applications that are never read. The resulting inefficiency is worst when there is a shortage of employers. We show that simple interventions—such as limiting the number of applications that an individual can send, making it more costly to apply, or setting an appropriate market-wide wage—can significantly improve the welfare of agents on one or both sides of the market. ***Managerial implications***: Our results suggest that platforms cannot focus exclusively on attracting participants and making it easy to contact potential match partners. A good user experience requires that participants not waste effort considering possibilities that are unlikely to be available. The operational interventions we study alleviate congestion by ensuring that potential match partners are likely to be available.

History: This paper has been accepted for the *Manufacturing & Service Operations Management* Special Issue on Sharing Economy and Innovative Marketplaces.
Funding: Financial support from the National Science Foundation [Grants CMMI-1234955, CMMI-1653477, and CNS-1343253] is gratefully acknowledged.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2020.0927.

Keywords: matching markets • decentralized • market design • operational interventions • stochastic model • mean field limit • search frictions • equilibrium • dynamics • contraction

1. Introduction

Driven by advances in information technology, online matching market platforms are revolutionizing how trading partners find each other, learn about each other, and ultimately consummate matches. For example, sites such as LinkedIn, Upwork, and TaskRabbit help match employers and employees; sites such as Match .com and OkCupid allow those seeking a romantic partner to browse profiles of nearby users; and travelers looking for short-term housing options can turn to Airbnb and VRBO.

Despite their prevalence, searching through hundreds or thousands of listings on these platforms and contacting promising options remains a costly and time-consuming process. The challenge of finding a partner is exacerbated by the fact that each user's demand for matches typically fluctuates over time. A worker on an online labor market may find themselves temporarily overwhelmed with other projects, or suddenly in need of extra work to pay the month's expenses. An OkCupid user might fail to respond to a message while on vacation or due to a newly formed relationship. A listing on Airbnb may be unavailable for a particular date because of a visit from the host's in-laws.

Typically, these changes in a user's status are not immediately reflected on their profile, or even known to the platform. As a result, a great deal of effort may be wasted screening profiles of users who are currently unavailable. Recent empirical work suggests that lack of information about user availability is prevalent and that this fact has significant welfare consequences.[1]

Motivated by this observation, we develop a stochastic game-theoretic model of an asynchronous, dynamic, two-sided matching market and use this model to investigate operational interventions that can relieve congestion. We refer to the two sides as "applicants" and "employers." Our model exhibits several key features: (i) It is *costly to apply* for positions; (ii) it is *costly to screen* applicants; (iii) employers are *uncertain* about which applicants are *available*; and (iv) some employers make *multiple offers* to attain a successful match.

We note that our model does not include the process of endogenous price or wage formation.[2] Thus, the results of our work are most directly well-suited to markets where such price-formation processes do not play a pivotal role—for example, online heterosexual dating markets (where men are typically expected to "apply"). Despite the lack of endogenous price formation in our formal model, we believe the qualitative insights of our paper are relevant to a wide range of online markets, as we discuss further below.

Our results reveal that equilibrium outcomes in our model are inefficient for both applicants and employers. Our paper is principally concerned with (1) identifying regimes where this inefficiency is especially acute, and (2) quantifying the benefits of simple operational interventions. Below, we summarize our main contributions.

1.1. Analysis of Equilibrium Outcomes

In Section 5, we give a complete characterization of the equilibrium of our model. We demonstrate that the market falls in one of two regimes: It is either *screening-limited* or *application-limited*, as described below.

Screening-limited markets feature low application costs (so applicants send many applications) and enough employers that most applicants can eventually match. However, employers find that many offers are rejected by applicants who have matched elsewhere; thus, employers must screen many candidates in order to hire successfully, and these screening costs offset the benefits of matching. When employers are ex ante homogeneous, the costs are high enough to drop employer welfare to zero, and some employers choose to exit the market rather than continuing their search. If employers are heterogeneous, then the effects on their welfare are heterogeneous as well, with the most selective employers obtaining zero welfare.[3]

Application-limited markets feature low screening costs. Although employers often do well in such markets, applicants face a "tragedy of the commons," as they send many applications competing for positions. This wasteful competition has severe welfare consequences when there is a shortage of employers (implying that not all applicants can match).

In short, equilibrium welfare is low for one or both sides of the market unless the market satisfies two conditions: Screening costs must be low (relative to application costs), and there must be enough employers for nearly all applicants to match.

1.2. Intervening to Improve Welfare

We show that simple platform interventions can significantly improve upon equilibrium outcomes. First, we show that *restricting the number of applications* that participants are allowed to send can provide substantial welfare benefits to employers in screening-limited markets and to applicants in application-limited markets with a shortage of employers. Furthermore, gains by one side typically come at little to no cost to the other side. In fact, in screening-limited markets, an application limit can simultaneously benefit both sides of the market.

We also consider two alternatives to explicitly limiting applications: (i) raising application costs, or (ii) setting a market-wide price (i.e., a wage). First, the platform could make it more costly to apply—for example, by charging a fee for each application or requiring the applicant to complete an extensive application form for each job. Naively, making applications more costly would seem to increase the burden on applicants, and, indeed, we find that (unless payments are redistributed or application forms provide information that reduces screening costs) higher application costs lower applicant welfare. However, this approach can often improve employer welfare—in fact, Theorem 6 states that, from an employer's perspective, raising application costs is equivalent to imposing an application limit. Therefore, this may be a desirable intervention if limiting applications is infeasible or as a way for the platform to monetize its services.

Alternatively, the platform could set prices in the market. Some platforms recommend prices to their sellers (e.g., Airbnb provides price recommendations to hosts (Ye et al. 2018); as these services gain adoption, the platform is effectively setting prices through such recommendations. In our formal analysis, we suppose that on any successful match, the employer is required to pay this price to the matched applicant. We study the potential congestion-alleviating effects of such an intervention. We find that a platform-controlled wage is, in some ways, at least as powerful as an application limit: For any application limit, an appropriately set wage can recover the same aggregate welfare in equilibrium. This suggests that price-setting by the platform may be a viable alternative to setting application limits, although the latter may lead to better outcomes for applicants (a point that we discuss further in the conclusion).

1.3. Mean-Field Analysis

A model capturing the features described above can quickly become intractable, due to complex agent strategies and stochastic fluctuations. A common technique in the literature (see Section 2) is to study such models in a large-market ("mean field") regime, where agents respond to others' *average* behavior, which is itself predictable due to the law of large numbers. Whereas typically this approximation is made without rigorous mathematical justification, we formally prove that our mean-field assumptions hold asymptotically in large markets, via a novel stochastic contraction argument.

We conclude by returning to the broader managerial consequences of our work. As noted above, our model does not include endogenous price negotiation by market participants, an important feature of a wide variety of online markets. Nevertheless, we feel the main qualitative insights of our work are relevant more broadly: The presence of congestion inefficiency for one or both sides of the market in equilibrium induced by a lack of availability information, and the ability to address this inefficiency through operational interventions, such as application limits, application costs, and centrally set market-wide prices. Some evidence of the importance of this issue is that online labor markets have implemented features to overcome precisely the type of congestion effect described here.[4] A formal model including endogenous price formation leads to substantial technical challenges beyond the scope of this work; understanding the operational interventions studied here within such a model remains an important direction for future research.

The remainder of the paper is organized as follows. In Section 2, we discuss related work. In Section 3, we present our model, and in Section 4, we discuss a formal mean-field model that represents a large market limit of our original model. All our equilibrium analysis proceeds within this mean-field model and is presented in Section 5. We give analytical expressions for applicant and employer welfare in equilibrium and demonstrate that simple operational interventions may significantly improve welfare. In Section 6, we present formal justification for using our mean-field model to study markets of finite size. We conclude in Section 7.

2. Related Work

Our work relates to several strands of literature. We discuss each in turn.

2.1. Frictions in Labor Markets

Our model closely resembles those used to study labor markets with costly search. This line of work was initiated by Diamond (1982a,b), Mortensen (1982a,b), and Pissarides (1984a,b) and has since received a great deal of attention. See Rogerson et al. (2005) and Wright et al. (2020) for helpful surveys. Much of this literature focuses on the question of whether equilibrium outcomes are (constrained) efficient: that is, whether agents make socially optimal decisions. Broadly speaking, the conclusion is that outcomes are inefficient unless employers compete by publicly posting wages (this is called "directed search") *and* there are no congestion externalities among employers.

In contrast to the search literature, our focus is not on whether equilibrium outcomes are precisely optimal: The fact that equilibrium is inefficient in our model is unsurprising, given the absence of wages and presence of congestion externalities in our model. Instead, we take a more operational perspective, by identifying conditions under which inefficiency is large and studying simple interventions that can improve welfare.

One of our key findings—that imposing an application limit on workers may dramatically increase employer welfare—does not, to our knowledge, have any analogue in the search literature; the search literature has traditionally not focused on such operational questions. Search models in prior work would yield very different conclusions about the effects of such an intervention, as we now detail. Our finding is driven by three of the key modeling features mentioned in our introduction: Screening is costly, employers are uncertain about availability, and employers may make multiple offers if their first offer is rejected.[5] Existing models in the search literature each lack one or more of these features. For example, Moen (1997), Julien et al. (2000), and Burdett et al. (2001) assume that applicants contact a single employer and, thus, are certain to be available; Albrecht et al. (2006) and Galenianos and Kircher (2009) allow employers to make only one offer, so that the competition among employers is never too severe;[6] and Kircher (2009) assumes that it is costless for employers to evaluate and make offers. As a result, the models in these papers would predict that an application limit offers little to no benefit to employers. We elaborate on this point, and especially the comparison with Albrecht et al. (2006) and Galenianos and Kircher (2009), in Online Appendix B.

2.2. Signaling in Matching Markets

The phenomenon of congestion—that it is impossible or costly for employers to make offers to many applicants and that employers compete for the same applicants—is discussed in Roth and Xing (1997). In such cases, a reasonable design choice is to enable applicants to signal their interest to employers; see, for example, Lee et al. (2011) and Coles et al. (2010) for applications in online data markets and the academic job market, respectively, and Coles et al. (2013) and Halaburda et al. (2018) for examples of how signaling can improve equilibrium outcomes.

Within the context of these papers, our applications can be seen as costly signals of interest. When application costs are small, applicants "signal" (apply to) many employers, and employers' offers are often rejected. Like Coles et al. (2013) and Halaburda et al. (2018), we find that offer acceptance rates rise if the platform limits applicants' ability to signal. However, whereas those papers find that higher acceptance rates result in more matches, we find that limiting applications results in (slightly) *fewer* matches. The reason for the difference is that we assume that employers can make multiple offers, each of which is costly. Thus, low acceptance rates do not reduce the number of matches formed (as they do in prior work), but do result in high search costs (which can be reduced by limiting applications).[7]

Kushnir (2013) finds that signaling may have a detrimental effect, including a reduction in the number of matches, and possibly in welfare. However, the source of welfare losses in that work is distinct from that in our paper: There, the inefficiency is driven by information asymmetry regarding preferences of workers, and signaling can be miscoordinated, reducing the number of matches formed. In our work, agents do not have preference information beforehand; the only way to match is through the (costly) application and screening process, and increasing the number of applications sent weakly increases the number of matches that form.

We remark that the literature on search frictions in labor economics (Section 2.1) generally relies on large-market assumptions that are not formally stated or justified. One exception is the recent work of Galenianos and Kircher (2012), which does provide a convergence proof. However, they consider a static setting in which each applicant contacts a single firm and acknowledge that their analysis does not easily generalize to the case where applicants contact multiple firms.

2.3. Operation of Two-Sided Markets

There is an emergent literature on the design of two-sided platform markets in the operations literature, which includes the current paper. The overall goal may be stated as that of optimally managing the inventory of a marketplace, where the market participants themselves constitute the inventory and are also the ones whose needs the market must meet. The work of Allon et al. (2012) is similar in spirit to our own, showing that, in service marketplaces, operational efficiency may hurt market efficiency due to the involvement of strategic agents. Iyer et al. (2014) and Balseiro et al. (2015) also apply mean-field analysis to answer market-design questions. (For an introduction to mean-field games, see Lasry and Lions 2007 and Weintraub et al. 2008.) In other contexts, operational optimization of dynamic matching markets has been carried out without a strategic model; see, for example, Hu and Zhou (2018), Ashlagi et al. (2013), Anderson et al. (2017), and Gurvich and Ward (2014).

2.4. Empirical Evidence from Platform Markets

Recent empirical work by Fradkin (2017) and Horton (2019) concludes that on both Airbnb and oDesk, it is common for users to contact unavailable partners and that this fact has significant welfare consequences. On Airbnb, guests are uncertain about host availability because transactions take time to complete, and hosts may not reliably update the calendars. Fradkin (2017) notes that only 15% of inquiries eventually lead to a transaction and that an initial rejection decreases the probability that the guest eventually books *any* listing by 50%. On oDesk, potential employers can browse profiles and invite workers to apply, but, in practice, the best workers are often too busy to accept the invitation. Horton (2019) concludes that employers are often unable to discern which workers are busy, that this fact causes many invitations to be declined, and that spurned employers are much less likely to eventually match.

3. The Model

In the market we consider, employers and applicants arrive, interact with each other, and eventually depart. Informally, we aim to capture the following behavior:

1. Employers arrive to the market, and each posts an *opening*.
2. When applicants arrive, each *applies* to a subset of the employers currently in the market.
3. Upon exit from the system, employers may screen candidates who have applied to learn whether they are *compatible* with the job. They may make offers to compatible applicants.
4. Whenever an applicant receives an offer, he chooses whether to accept it.

We model a dynamic market, in which participants are posting and applying for jobs, screening, and making offers asynchronously. The timing of this market is described in more detail below.

3.1. Arrivals

Our market starts empty at $t = 0$. Our dynamic markets are parameterized by $n > 0$, which describes the market size. Individual employers arrive at intervals of $1/n$, and applicants arrive at intervals of $1/(rn)$. Here, $r > 0$ is a parameter that controls the relative magnitude of the two sides of the market. Employers remain in the system for a unit lifetime. Applicants depart the system according to a process that we describe below. We endow employers and applicants with unique IDs, which convey no other information about the agent.

Upon arrival, employers post an opening. They do not make any decisions at this time. Upon arrival, each applicant selects a "search intensity" $m_a \in [0, n]$ and applies to each employer currently in the system with probability m_a/n. Note that for all $t \geq 1$, there are exactly n employers in the system, and thus the expected number of applications sent by an applicant who arrives after time $t \geq 1$ is[8] m_a.

3.2. Applicant Departure

Applicants remain in the system for a maximum of one time unit. Whenever the applicant receives an offer from any employer (as described below), the applicant must choose immediately whether to accept it. If they accept, they immediately and automatically depart from the market at that time. Excluding application costs, an applicant earns a payoff of one as long as she is matched and zero otherwise. Thus, *it is a dominant strategy for applicants to accept the first offer they receive.* We assume henceforth that applicants follow this strategy.

3.3. Employer Departure

Each employer stays in the system for one time unit, and then departs. At the time of departure, the employer sees the set of applicants to her opening. Initially, she does not know which of these applicants are compatible for her job, nor does she know which would accept the job if offered it (i.e., which applicants remain in the system). The employer takes a sequence of "screening" and "offer" actions, instantaneously learning the result of each, until an offer she makes is accepted (causing her to exit) or she chooses to exit the market.

At each stage of the employer's sequential decision process, she may screen any unscreened applicant, thereby learning whether this applicant is compatible for the job. If the employer has yet to match, she may also make an offer to any applicant whom she has found to be compatible (the applicant responds immediately to the offer; as mentioned above, we assume that they accept it if and only if it is the first offer that they receive). Employers are also allowed to skip some or all applications and exit the market at any point.

We require that before making an offer to any applicant, employers must screen this applicant and find him compatible (this assumption is justified if, for example, the cost of hiring an unqualified worker is sufficiently high). We assume that that each employer–applicant pair is compatible with probability β (independently across all such pairs) and that this is common knowledge.[9]

3.4. Utility

If a compatible pair matches to each other, the employer earns v and the applicant earns w. We normalize $v = w = 1$, which is without loss of generality because we do not compare absolute welfare of agents on opposite sides of the market throughout most of the paper.[10] Applicants pay a cost c_a for each application that they send, and employers pay a cost c_s for each applicant that they screen. The net utility to an agent will be the difference between value obtained from any match, and costs incurred, and agents act so as to maximize their expected utility.

Assumption 1. $\max(c_a, c_s) < \beta$.

This assumption rules out the uninteresting case where costs are so high that no activity occurs in the marketplace. If $c_s \geq \beta$, then, regardless of applicant behavior, it would be optimal for employers to exit the market rather than screening. Similarly, if $c_a \geq \beta$, then, because employers hire only compatible applicants, no applicant strategy can earn positive surplus.

For later reference, it will be useful to consider *normalized* versions of the screening and application costs, given by

$$c'_s = c_s/\beta, \qquad c'_a = c_a/\beta. \tag{1}$$

4. The Large Market: A Stationary Mean-Field Model

In principle, the strategic choices facing an agent in the model described above may be quite complicated. Consider the case of an employer who knows that he has only one competitor. If he finds that one applicant has already accepted another offer, he learns that every other applicant is still looking for a job. Similar logic suggests that, in thin markets, information revealed during screening may induce significant shifts in the employer's beliefs. This could conceivably cause optimal employer behavior to be quite complex.

As the market thickens, however, one might expect that the correlations between agents on the same side of the market become weak. In particular, employers screening applicants from among a large number of workers might reasonably assume that learning that one applicant has already accepted an offer does not inform them about the availability of other applicants. Further, if the employer cannot distinguish individual applicants, each one should appear to be available with equal probability. Similarly, applicants who know nothing about individual employers may be justified in assuming that each of their applications convert to offers independently and with equal probability p.

In this section, we develop a formal stationary *mean-field* model for our dynamic matching market and introduce a notion of game-theoretic equilibrium for this model; in particular, we study a model that arises from a limiting regime where the market thickens.

In our formal model, agents make the following assumptions.

Mean-Field Assumption 1 (Employer Mean-Field Assumption). *Each applicant in an employer's applicant set is available with probability q, and availability in the applicant set is independent across applicants.*

Mean-Field Assumption 2 (Applicant Mean-Field Assumption). *Each application yields an offer with probability p, independently across applications to different openings.*

Mean-Field Assumption 3 (Large Market Assumption). *The number of applications sent by an applicant who chooses $m_a = m$ is Poisson distributed with mean m. If all applicants select $m_a = m$, the number of applications received by each employer is Poisson distributed with mean rm.*[11]

Under these assumptions, optimal agent behavior simplifies greatly. We describe optimal employer and applicant responses in Section 4.1. For applicants, we show that there exists a unique optimal choice of m, given p; and for employers, we show that given q, their optimal response is either to employ a simple sequential screening strategy or to exit immediately (they may also randomize when indifferent between these options).

Of course p and q are not given exogenously, but, rather, arise endogenously from the choices made by agents. In Section 4.2, we derive "consistency checks" that p and q should satisfy, if they indeed arise from the conjectured employer and applicant strategies.

The work in Sections 4.1 and 4.2 allows us to define a *mean-field equilibrium* (MFE) in Section 4.3. Informally, a mean-field equilibrium consists of strategies for employers and applicants that are best responses to the stationary market dynamics that they induce. We prove that there exists a *unique* MFE. We conclude with Section 4.4, which discusses a simple intervention available to the market operator: placing a limit ℓ on the value m_a chosen by each applicant. We show that in this setting, too, there exists a unique MFE.

Importantly, we prove that our mean-field model is, in fact, the correct limit of our dynamic market as the thickness n grows. In particular, Theorems 8 and 9 in Section 6 justify our study of MFE: They state that the mean-field assumptions hold as n approaches infinity and that, as a consequence, any MFE is an approximate equilibrium in the game with finite, but sufficiently large n.

4.1. Optimal Decision Rules

We first study how agents respond when confronted with a world where the mean-field assumptions hold.

4.1.1. Applicants. As discussed in Section 3, it is a dominant strategy for applicants to accept the first offer (if any) that they receive, and we assume applicants follow this rule. Therefore, the only decision an applicant a needs to make on arrival is her choice of m_a, the expected number of applications sent.

An applicant a who chooses $m_a = m$ incurs an expected cost of $c_a \cdot m$. If the applicant applies to Poisson(m) employers, and each application independently yields an offer with probability p, then at least one offer is received—that is, the applicant matches to an employer—with probability $1 - e^{-mp}$. Thus, the expected payout of an applicant in the mean-field environment who selects $m_a = m$ is

$$W_a(m, p) = 1 - e^{-mp} - c_a m. \tag{2}$$

Applicants choose $m \geq 0$ to maximize this payoff. Because their objective is strictly concave and decays to $-\infty$ as $m \to \infty$, this problem possesses a unique optimal solution identified by first-order conditions. If $p \leq c_a$, the optimal choice is $m = 0$. Otherwise, applicants select $m = \frac{1}{p}\log(\frac{p}{c_a})$. We define $\mathcal{M}$ to be the function that maps p to the unique optimal value of m:

$$\mathcal{M}(p) = \begin{cases} 0, & \text{if } p \leq c_a; \\ \frac{1}{p}\log\left(\frac{p}{c_a}\right), & \text{if } p > c_a. \end{cases} \tag{3}$$

4.1.2. Employers. Next, we consider the optimal strategy for employers, when Mean-Field Assumption 1 holds. We consider a simple strategy, which we denote ϕ^1. An employer playing ϕ^1 sequentially screens candidates in her applicant list. When she finds a compatible applicant, she makes an offer to this candidate; otherwise, she considers the next candidate. This process repeats until one applicant accepts or no more applicants remain.

The optimal strategy for the employers is straightforward to characterize. First, suppose an employer has exactly one applicant. The employer will prefer to screen the applicant if $\beta q - c_s > 0$—that is, if $q > c_s'$; exit if $q < c_s'$; and is indifferent if $q = c_s'$. Now, it is clear that if an employer has more than one applicant in her list, because all applicants are ex ante homogeneous from the perspective of the employer, the same reasoning holds: The employer will screen or exit immediately according to whether q is larger or smaller than c_s', respectively. (Note the essential use of Mean-Field Assumption 1: If there is correlation in the availability of successive applicants in the employer's list, the preceding reasoning no longer holds.) The following proposition (proved in Online Appendix C) summarizes the preceding discussion.

Proposition 1. *Let ϕ^1 be the strategy of sequentially screening applicants, offering them the job if and only if they are qualified, until either an applicant is hired or no more applicants remain. Then, ϕ^1 is uniquely optimal if and only if $q > c_s'$, exiting immediately is uniquely optimal if and only if $q < c_s'$, and any mixture of these strategies is optimal if $q = c_s'$.*

Motivated by this proposition, we define ϕ^α to be the strategy that plays ϕ^1 with probability α and exits immediately otherwise. Define the correspondence $\mathcal{A}(q)$ by:

$$\mathcal{A}(q) = \begin{cases} \{0\} & \text{if } q < c_s' \\ [0,1] & \text{if } q = c_s' \\ \{1\} & \text{if } q > c_s'. \end{cases} \tag{4}$$

This correspondence captures the optimal employer response, as described in Proposition 1, so that $\mathcal{A}(q) = \{\alpha \in [0,1] : \phi^\alpha \text{ is optimal for the employer, given } q\}$.

We define

$$W_e(\alpha, m, q) = \alpha\left(1 - e^{-rm\beta q}\right)\left(1 - c_s'/q\right) \tag{5}$$

to be the expected welfare of an employer who screens with probability α, receives a number of applications that is Poisson with mean rm, and finds each applicant qualified with probability β and (independently) available with probability q.

4.2. Consistency

In the previous section, we discussed the best responses available to employers and applicants when the mean-field assumptions hold; that is, given p and q, we found the strategies that agents would adopt. However, p and q are clearly *determined* by agent strategies. In this section, we identify consistency conditions that p and q must satisfy, given specified agent strategies.

We focus on strategies that could conceivably be optimal, as identified in the preceding section. We assume that all applicants choose the same $m \geq 0$ and that all employers play ϕ^α—that is, they play ϕ^1 with probability α and exit immediately otherwise. From any m and α, we derive a unique prediction for the pair (p, q).

We emphasize at the outset that our analysis aims only to derive the correct consistency conditions under the mean-field assumptions. We provide rigorous justification for these assumptions via the propositions in Section 6. As a consequence, those propositions also justify the consistency conditions described below.

We start by deriving a consistency condition for q, given p and the strategy adopted by applicants. Intuitively, q should be equal to the long-run fraction of offers that are accepted. Fix the value of m chosen by applicants, and let X be the number of offers received by a single applicant. This applicant will accept an offer if and only if $X > 0$, so the expected fraction of offers that are accepted is $P(X > 0)/E[X]$. If Mean-Field Assumptions 2 and 3 hold, then X is Poisson with mean mp, so we should have

$$q = \frac{1 - e^{-mp}}{mp}. \tag{6}$$

To derive a consistency condition for p, note that when employers follow ϕ^α, only compatible applicants receive offers. Thus, p should equal β times the long-run fraction of qualified applications that result in offers. Because an applicant's availability does not influence whether they receive an offer (as it is unobserved by prospective employers), this should be equal to β times the fraction of applications by qualified *available* applicants that result in offers. Fix an employer playing ϕ^α, and let Y be the number of qualified, available applicants received by this employer. This employer successfully hires if and only if $Y > 0$ and she decides to screen. Thus, the fraction of qualified available applicants who receive offers should be $\alpha P(Y > 0)/E[Y]$. Because each applicant is qualified with probability β, and because available and unavailable applicants receive offers with the same probability (as employers do not observe availability), we conclude that the ex ante probability that an applicant receives an offer from a given employer is $\alpha\beta P(Y > 0)/E[Y]$. Mean-Field Assumptions 1 and 3 jointly imply that Y is distributed as a Poisson random variable with mean $rm\beta q$, from which we conclude that

$$p = \alpha\beta \frac{P(Y > 0)}{E[Y]} = \alpha\beta \frac{1 - e^{-rm\beta q}}{rm\beta q}, \tag{7}$$

Equations (6) and (7) are a system for p and q, given the values of m and α (as well as the parameters r and β). The following proposition states that the pair of consistency Equations (6) and (7) have a unique solution (the proof is in the full technical report, Arnosti et al. 2019).

Proposition 2. *For fixed m, α, r, and β, there exists a unique solution (p, q) to (6) and (7).*

We refer to the unique pair (p, q) that solve (6) and (7) as a *mean-field steady state* (MFSS). This pair provides a prediction of how a large market should behave, given specific strategic choices of the agents. For later reference, given strategies m and α (and parameters r and β), let $\mathcal{P}(m, \alpha; r, \beta)$ and $\mathcal{Q}(m, \alpha; r, \beta)$

denote the unique values of p and q guaranteed by Proposition 2, respectively. Because our analysis is conducted with r and β fixed, we will omit the dependence on r and β in favor of the more concise $\mathcal{P}(m,\alpha)$ and $\mathcal{Q}(m,\alpha)$.

4.3. Mean-Field Equilibrium

In this section, we define *mean-field equilibrium* (MFE), a notion of game-theoretic equilibrium for our stationary mean-field model. Informally, an MFE should be a pair of strategies such that (1) agents play optimally given their beliefs about the marketplace—that is, the values p and q in the mean-field assumptions—and (2) agent beliefs are consistent with the strategies being played—that is, (p,q) is an MFSS corresponding to the agents' strategies. Section 4.1 addressed the first point, and Section 4.2 addressed the second. We define a mean-field equilibrium by composing the maps defined in those sections.

Definition 1. A *mean-field equilibrium is* a pair (m^*,α^*) such that $m^* = \mathcal{M}(\mathcal{P}(m^*,\alpha^*))$ and $\alpha^* \in \mathcal{A}(\mathcal{Q}(m^*,\alpha^*))$.

In an MFE, m^* and α^* are optimal responses (under the mean-field assumptions) to the steady-state (p,q) that they induce. For future reference, we define $p^* = \mathcal{P}(m^*,\alpha^*)$ and $q^* = \mathcal{Q}(m^*,\alpha^*)$. Our main theorem in this section establishes existence and uniqueness of MFE (the proof is in the full technical report, Arnosti et al. 2019).

Theorem 1. *Fix any r,β,c_a,c_s such that Assumption 1 holds. Then, there exists a unique mean-field equilibrium (m^*,α^*).*

4.4. A Market Intervention: Application Limits

As noted in the introduction, we are interested in comparing the outcome of the market described above to the outcome when the platform operator intervenes to try to improve the welfare of employers and/or applicants. We consider a particular type of intervention: a limit on the number of applications that can be sent by any individual.

In our model with application limits, agent payoffs are identical to before, as are the strategies available to employers. Applicants, however, are restricted to selecting $m_a \le \ell$. In the corresponding mean-field model, given p, applicants choose m_a to maximize $1 - e^{-m_a p} - c_a m_a$ (their expected payoff), subject to $m_a \in [0,\ell]$. The applicant objective is concave in m_a, so this problem has a unique solution given by

$$\mathcal{M}_\ell(p) = \min(\ell, \mathcal{M}(p)). \tag{8}$$

The consistency conditions are identical to those in Section 4.2. We define a *mean-field equilibrium* of the market with application limit ℓ as a pair $(m^*_\ell, \alpha^*_\ell)$ solving the following pair of equations:

$$m^*_\ell = \mathcal{M}_\ell(\mathcal{P}(m^*_\ell,\alpha^*_\ell)), \quad \alpha^*_\ell \in \mathcal{A}(\mathcal{Q}(m^*_\ell,\alpha^*_\ell)). \tag{9}$$

The following proposition is an analog of Theorem 1 for the market with an application limit (the proof is in the full technical report, Arnosti et al. 2019).

Proposition 3. *Fix r,β,c_a,c_s such that Assumption 1 holds, and let (m^*,α^*) be the corresponding MFE in the market with no application limit. Then, for any $\ell \ge 0$, there exists a unique mean-field equilibrium in the market with application limit ℓ. If $m^* \le \ell$, then $(m^*_\ell,\alpha^*_\ell) = (m^*,\alpha^*)$. Otherwise, $m^*_\ell = \ell$ and α^*_ℓ is the unique solution to $\alpha^*_\ell \in \mathcal{A}(\mathcal{Q}(\ell,\alpha^*_\ell))$.*

For future reference, we define $p^*_\ell = \mathcal{P}(m^*_\ell,\alpha^*_\ell)$, $q^*_\ell = \mathcal{Q}(m^*_\ell,\alpha^*_\ell)$.

5. Welfare Analysis

In this section, we study the expected welfare of applicants and employers, both in equilibrium and under operational interventions such as an application limit. We compare welfare to two simple *frictionless* benchmarks:

Applicant welfare is at most $\min(1, 1/r)$, because if $r > 1$, then not all applicants can match.

Employer welfare is at most $\min(1, r)$, because if $r < 1$, then not all employers can match.

These benchmarks are very optimistic, as they completely ignore application and screening costs. However, when costs are small, one might hope to obtain welfare for both sides that is close to these benchmarks. We make Assumption 1 throughout this section.

In Section 5.1, we explicitly characterize employer and applicant welfare in equilibrium, without any operational intervention by the platform. We show that there are two regimes for welfare: one where the market is *application-limited* (welfare is determined by applicants' willingness to apply) and one where the market is *screening-limited* (welfare is determined by employers' willingness to screen). Equilibrium welfare of applicants and employers can fall well short of the benchmarks above, even if costs are small: When the market is application-limited, applicants' welfare can be an arbitrarily small fraction of the benchmark $\min(1, 1/r)$, and when the market is screening-limited, employers' welfare is identically *zero*.

We demonstrate in Section 5.2 that the platform can substantially improve the situation by imposing an application limit. In particular, we establish that when search costs are low, for any one side of the market, an appropriately chosen application limit can bring the welfare of that side close to the corresponding frictionless benchmark discussed above. We also provide evidence that the tradeoff between optimizing for applicants and employers is never too severe and establish that when the market is screening-limited, then a suitable application limit simultaneously raises welfare for both sides.

In Sections 5.3 and 5.4, we discuss alternative interventions that can be used to mitigate congestion. Section 5.3 studies the effect of imposing additional application costs, while Section 5.4 considers a model in which the platform can set a market-wide price (i.e., a wage) that employers pay to applicants upon a successful match. Raising application costs can significantly improve employers' welfare (exactly as with an application limit), but always weakly reduces applicant welfare. We also show a suitable market-wide wage can achieve identical aggregate welfare to any desirable application limit.

Finally, Section 5.5 explores the generalizability of our findings to models with ex-ante heterogeneity. In it, we consider one extension in which some employers' jobs are harder than others' (lower probability of compatibility β), and another in which some applicants are more skilled (higher probability of compatibility β) than others. Numerical investigation of the solution to these models shows that our key insights from Sections 5.1 and 5.2 continue to apply.

We begin with the following notational preliminaries. Fix parameter values r (the ratio of applicants to employers), β (probability of compatibility), c_a (application cost), and c_s (screening cost). For given applicant strategy m and employer strategy α, we let $\Pi_a(m,\alpha)$ and $\Pi_e(m,\alpha)$ denote the mean-field applicant and employer welfare, respectively. These are given by

$$\Pi_a(m,\alpha) = W_a(m,\mathcal{P}(m,\alpha)) = 1 - e^{-m\mathcal{P}(m,\alpha)} - c_a m; \quad (10)$$

$$\Pi_e(m,\alpha) = W_e(\alpha,m,\mathcal{Q}(m,\alpha)) = \alpha\left(1 - e^{-rm\beta\mathcal{Q}(m,\alpha)}\right)\left(1 - \frac{c'_s}{\mathcal{Q}(m,\alpha)}\right), \quad (11)$$

using Equations (2) and (5), where we recall that $\mathcal{P}(m,\alpha)$ and $\mathcal{Q}(m,\alpha)$ are, respectively, the mean-field acceptance probability and availability; see Proposition 2 and the subsequent discussion. The expression for $\Pi_a(m,\alpha)$ is the probability of at least one application being successful, less application costs. The expression for $\Pi_e(m,\alpha)$ is the probability of successfully hiring, times the net match value for an employer (match value minus expected screening cost per successful match).

In much of what follows, we make use of a reduced parameter set consisting of $r, c'_a = c_a/\beta$ and $c'_s = c_s/\beta$, as defined in (1). We do so because it turns out that welfare in equilibrium (with or without application limits) depends only on the value of these three parameters.

5.1. Quantifying Equilibrium Welfare

We begin by providing analytical expressions for equilibrium welfare of both sides. Define Π^*_a, Π^*_e to be the applicant and employer welfare, respectively, in the unique mean-field equilibrium (m^*,α^*) guaranteed by Theorem 1. In other words,

$$\Pi^*_a = \Pi_a(m^*,\alpha^*), \Pi^*_e = \Pi_e(m^*,\alpha^*). \quad (12)$$

Examining (10) and (11) suggests that participant welfare Π^*_a and Π^*_e depend on equilibrium strategies m^*,α^*, as well as equilibrium outcomes p^* and q^*. However, the relationships between the values m^*,α^*,p^*,q^* allow us to express Π^*_a and Π^*_e in terms of only model primitives and the equilibrium acceptance probability p^*.[12] Theorem 2 provides these expressions, as well as equations that define p^*.

Theorem 2. *There are unique solutions $\hat{p},\overline{p} > c_a$ to the following equations:*

$$r\left(1 - \frac{c_a}{\hat{p}}\right) = 1 - e^{-\frac{r\beta}{\hat{p}}\left(1-\frac{c_a}{\hat{p}}\right)}, \quad (13)$$

$$1 - \frac{c_a}{\overline{p}} + \frac{c_s}{\beta}\log\left(\frac{c_a}{\overline{p}}\right) = 0. \quad (14)$$

In the unique MFE, $p^* = \min(\hat{p},\overline{p})$, and

$$\Pi^*_a = 1 - \frac{c_a}{p^*} + \frac{c_a}{p^*}\log\left(\frac{c_a}{p^*}\right), \quad (15)$$

$$\Pi^*_e = r\left(1 - \frac{c_a}{p^*} + \frac{c_s}{\beta}\log\left(\frac{c_a}{p^*}\right)\right). \quad (16)$$

Theorem 2 identifies that the parameter space can be divided into two regions: In one, welfare is determined by $\hat{p}$, and in the other, it is determined by $\overline{p}$. The interpretation for this is that $\hat{p}$ is the acceptance probability that arises when all employers choose to screen, and applicants behave according to a corresponding partial equilibrium.[13] If the screening cost c_s is sufficiently small, this outcome will be an equilibrium: We will have $p^* = \hat{p}$. On the other hand, if screening costs are large, then not all employers will choose to screen. In this case, the value p^* will be such that when applicants respond optimally, employers are exactly indifferent about whether to screen—this is the value $\overline{p}$. Motivated by this observation, we introduce the following definition.

Definition 2. The market is *application-limited* if $\overline{p} > \hat{p}$ (in which case $p^* = \hat{p}$) and *screening-limited* if $\overline{p} < \hat{p}$ (in which case $p^* = \overline{p}$).

As we will see, the distinction between application-limited and screening-limited markets plays an important role in determining which side(s) have low welfare in equilibrium. Accordingly, we briefly discuss the parameter regimes that are characteristic of application-limited and screening-limited markets. Figure 1 depicts the boundary between the application-limited and screening-limited regions, as a function of the model primitives r, c_s and c_a, with $\beta = 0.5$. In order for the market to be screening-limited, applicant availability must be a concern to employers. This occurs when there are more employers than applicants and application costs are small. More precisely, if $r < 1$, then for any screening cost $c_s > 0$, the market is screening-limited for all sufficiently small application costs c_a. Conversely, if either of these conditions fail—that is, there are more applicants than employers or application costs are moderate—then the market is application-limited unless screening costs are very high.[14]

Although Theorem 2 gives analytical expressions for the welfare of both sides of the market, these expressions can be difficult to interpret, as they depend on the equilibrium quantity p^*. As we now demonstrate, employer and applicant welfare can fall far short of the frictionless benchmarks, even when search and application costs are small.

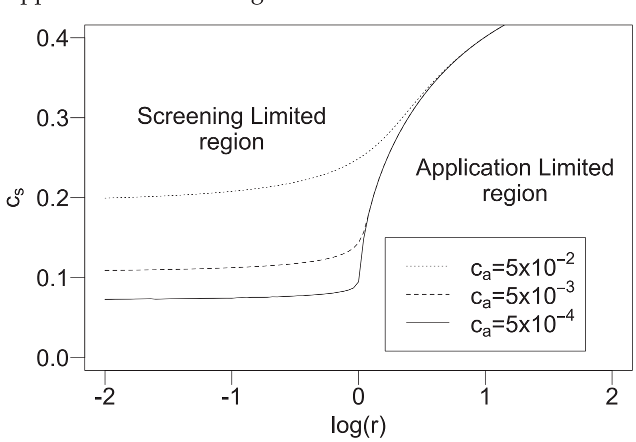
Theorem 3.

1. *If the market is screening-limited, then equilibrium employer welfare* $\Pi_e^* = 0$.
2. *If the market is application-limited and* $r > 1$, *then equilibrium applicant welfare*

$$\Pi_a^* \leq \frac{1}{r}\left(1 - (r-1)\log\left(\frac{r}{r-1}\right)\right).$$

Theorem 3 states that when the market is screening-limited, employer welfare is zero (indeed, this is a direct consequence of the definition of screening-limited: If some employers choose not to participate, they must anticipate zero welfare from the market). Meanwhile, when $r > 1$ and the market is application-limited, applicant welfare in equilibrium is substantially below the frictionless benchmark level $1/r$: the bound in Theorem 3 implies that applicant welfare is at most half of the upper bound of $1/r$ when $r = 1.4$, at most one-third of $1/r$ when $r = 1.9$, and becomes a vanishingly small fraction of $1/r$ as r increases—regardless of application costs! In other words, at least one side has low welfare in equilibrium unless the market is application-limited and there is a shortage of applicants ($r < 1$).[15]

Figure 1. The Boundary Between Screening-Limited and Application-Limited Regimes



Notes. We take $\beta = 0.5$ and vary c_a. The y axis gives the screening cost c_s, and the x axis gives the log of the ratio of applicants to employers. Smaller application costs lead to a larger screening-limited region. For fixed c_s and $r < 1$, the market is screening-limited for all sufficiently small c_a.

5.2. Improving Welfare by Limiting Applications

As we now show, an application limit is a powerful operational tool for improving participants' welfare.

We let Π_a^ℓ and Π_e^ℓ be expected applicant and employer payoffs in the unique equilibrium of the market with application limit ℓ. In other words,

$$\Pi_a^\ell = \Pi_a(m_\ell^*, \alpha_\ell^*),\ \Pi_e^\ell = \Pi_e(m_\ell^*, \alpha_\ell^*). \tag{17}$$

The intuition for the benefits of an application limit is as follows. When the market is screening-limited, employers are choosing not to screen because the expected cost of screening completely offsets any benefits from matching. By imposing an application limit, the platform ensures that applicants will be more likely to accept offers. This reduces the screening cost required to find an available applicant and increases employer welfare. Applicants, in turn, always suffer from a "tragedy of the commons" in the no-intervention equilibrium: They do not internalize the externality that their applications incur upon others. This has especially severe consequences when the market is application-limited and $r > 1$, implying that a significant fraction of applicants remain unmatched. An application limit can reduce this costly competition. Theorem 4 formalizes these observations.

Theorem 4.

1. *Suppose the market is screening-limited. Then*:

$$\sup_\ell \Pi_e^\ell = r(1 - c_s' + c_s' \log c_s').$$

2. *Suppose the market is application-limited and* $r > 1$. *Then,*

$$\sup_\ell \Pi_a^\ell \geq \frac{1}{r} - \frac{c_a'}{r} + c_a' \log(c_a') \log\left(\frac{r}{r-1}\right).$$

Note that if $r \leq 1$, c_s' is small and c_a' is small enough that the market is screening-limited, then the first part of Theorem 4 implies $\sup_\ell \Pi_e^\ell \approx r$, the frictionless benchmark for employers. Similarly, if $r > 1$ and c_a' is small, then the market is application-limited and the second part of Theorem 4 implies that $\sup_\ell \Pi_a^\ell \approx 1/r$, the frictionless benchmark for applicants.

Figure 2 complements the insights from Theorems 3 and 4. In it, we plot the ratio of equilibrium welfare to the welfare attainable with an application limit, for both sides of the market (and for varying parameter values). The figure demonstrates that an application limit can substantially help one or both sides of the market unless the market is application-limited and $r < 1$. The figure displays the screening-limited region where employers get zero welfare and also shows that an application limit significantly benefits applicants when r is large.

The figure also reveals that an application limit substantially improves applicant welfare when c_s is high (so that the market is screening-limited). The intuition is that, in this case, many employers choose not to screen, leaving applicants competing for a small number of positions: The effective market imbalance exceeds one. As a result, the tragedy of the commons is severe. Although an application limit cannot increase the number of matches, it can dramatically reduce the wasteful competition among applicants.

While there may exist choices of ℓ that result in high employer welfare and choices of ℓ that result in high applicant welfare, these choices of ℓ might not coincide. We now address this concern, both theoretically and numerically. Theorem 5 states that whenever the market is screening-limited, it is possible to choose a single application limit such that both employers and applicants are better off than they would be without any intervention.

Theorem 5. *If the market is screening-limited, there exists ℓ such that $\Pi_e^\ell > \Pi_e^*$ and $\Pi_a^\ell > \Pi_a^*$.*

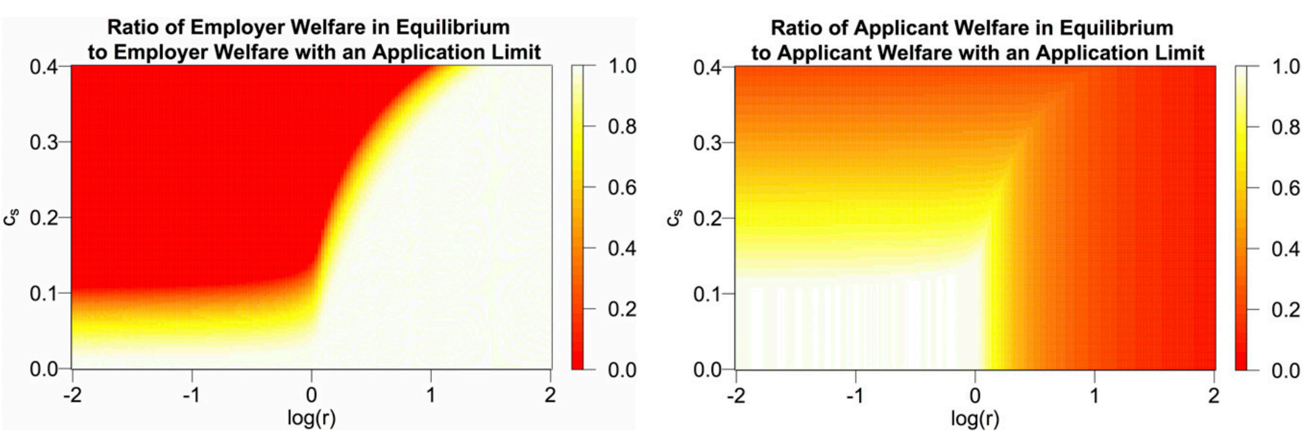
One might wonder whether Pareto improvements are possible when the market is application-limited. The answer turns out to be, "not always." Indeed, if either r or c_a' is large enough, then availability never becomes a pressing concern to employers, and *any* binding application limit lowers employer welfare (this holds, for example, if $c_s' < \max\{1 - 1/r, c_a'\}$).

Even when Pareto improvements are not possible, however, an application limit may substantially improve applicant welfare at little cost to employers. Figure 3 shows that the tension between optimizing for applicants and employers is not too severe. It displays the largest fraction δ such that there exists a single limit ℓ' where employers earn an expected surplus of $\Pi_e^{\ell'} \geq \delta \sup_\ell \Pi_e^\ell$ and applicants (simultaneously) earn an expected surplus of $\Pi_a^{\ell'} \geq \delta \sup_\ell \Pi_a^\ell$. Observe that δ is never below 3/4, and is often much higher. We conjecture that it is always possible to choose a limit such that each side attains at least 3/4 of the welfare that would be obtained by optimizing solely for that side.[16]

5.3. Increasing Application Costs

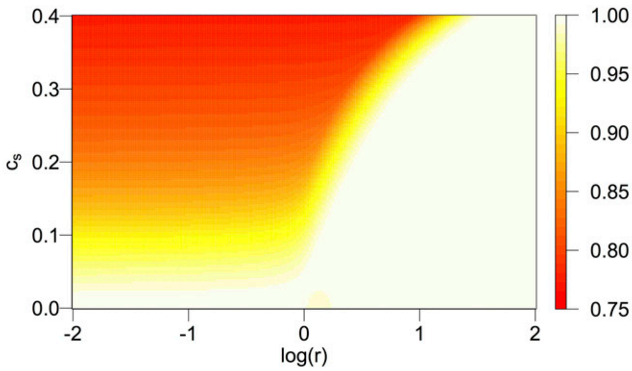
In this section, we study a second intervention that reduces the number of applications sent: raising the application cost. From the perspective of the employer, limiting applications and raising application costs have similar effects. To applicants, however, they look different. Although both interventions reduce the amount of competition for each opening, raising the application cost also directly harms applicants. A priori, it is not obvious which effect dominates. Theorem 6 shows that within our model, raising application costs can only hurt applicants.[17] Additionally, it formalizes the idea that the two interventions are equivalent from the perspective of employers.

Figure 2. (Color online) Ratio of Equilibrium Welfare to Welfare Attainable with an Application Limit



Notes. Probability of compatibility $\beta = 0.5$ and application cost $c_a = 0.005$, and we vary the screening cost c_s (y axis) and ratio of applicants to employers r (x axis). An application limit can substantially help one or both sides of the market, unless screening costs are low and there are fewer applicants than employers. See Theorems 3 and 4.

Figure 3. (Color online) An Application Limit Can Ensure Good Welfare for Applicants and Employers Simultaneously



Notes. For $\beta = 0.5$, $c_a = 0.005$ and varying values of c_s (y axis) and $\log(r)$ (x axis), this figure shows the largest fraction δ such that there exists a single limit ℓ for which employers earn an expected eurplus of $\delta \sup_\ell \Pi_a^\ell$ and applicants (simultaneously) earn an expected surplus of $\delta \sup_\ell \Pi_a^\ell$. Note that the shading scale is different from that in Figure 2 and that $\delta \geq 3/4$ for all depicted parameter values. This suggests that the tension between optimizing for applicants and employers is not too severe.

Theorem 6. *Fix values of r, c_s, β, such that $c_s < \beta$, and fix the application cost $c_a < \beta$ under no intervention. Let $m^*(c)$ denote the equilibrium value of m selected by applicants when application costs are raised to $c \in [c_a, \beta]$, and let $\Pi_a^*(c)$ and $\Pi_e^*(c)$ denote the applicant welfare and employer welfare, respectively. Then, (i) $\Pi_a^*(\cdot)$ is weakly decreasing; and (ii) for each $c_a \in (0, \beta)$ and each $\ell \in [0, m^*(c_a)]$, there exists a unique $c > 0$, such that $m^*(c) = \ell$. For this value of c, it holds that $c \in [c_a, \beta]$ and $\Pi_e^\ell = \Pi_e^*(c)$. (Recall that Π_e^ℓ is the employer welfare with an application limit ℓ and no change to the application cost c_a.)*

5.4. Price Setting by the Platform

The model we have studied so far does not include transfers. In this section, we study an extension where the platform can set a marketwide wage w that is paid on each job performed, from the employer to the worker. The wage offers another "lever" for the platform: As wages fall, the value of an application decreases. We investigate how this can be used to manage congestion.

As before, each successful match generates a total surplus of two units. We allow the platform to set any wage $w \in (0, 2)$, which causes a match utility of w to accrue to the applicant and a match utility of $v = 2 - w$ to accrue to the employer. The model discussed previously corresponds to the special case where $w = 1$ is fixed. For any wage w, the expected utility of an applicant in the mean-field environment with success probability p who applies with intensity m is given by the following generalization of (2):

$$W_a(m, p) = w(1 - e^{-mp}) - c_a m. \tag{18}$$

The applicant best response is given by the following generalization of (3):

$$\mathcal{M}_w(p) = \begin{cases} 0, & \text{if } p \leq c_a/w; \\ \frac{1}{p}\log\left(\frac{pw}{c_a}\right), & \text{if } p > c_a/w. \end{cases} \tag{19}$$

Let $c_s' = c_s/\beta$ as before. Proposition 1 remains true with the modification that the threshold on q is now $c_s'/(2 - w)$. Hence (generalizing the definition of $\mathcal{A}(q)$ in (4)), we define the correspondence

$$\mathcal{A}_w(q) = \begin{cases} \{0\} & \text{if } q < c_s'/(2 - w) \\ [0, 1] & \text{if } q = c_s'/(2 - w) \\ \{1\} & \text{if } q > c_s'/(2 - w). \end{cases} \tag{20}$$

As before, the strategy ϕ^α that plays ϕ^1 with probability α and exits immediately otherwise is an employer best response for $\alpha \in \mathcal{A}_w(q)$. The expected welfare of an employer who screens with probability α in the mean-field environment is given by the following generalization of (5):

$$W_e(\alpha, m, q) = \alpha(1 - e^{-rm\beta q})(2 - w - c_s'/q). \tag{21}$$

The definition of MFE is as before, except that for any fixed wage, the applicant and employer best responses are given by (19) and (20). In contrast to the remainder of the paper, we assume that employer and applicant welfare are quasilinear in the wage and, thus, measured in monetary units. We define the aggregate welfare as the sum of employer and applicant welfare, given by

$$\Pi_{\text{tot}} = r\Pi_a + \Pi_e. \tag{22}$$

Our main result in this section is the following.

Theorem 7. *Consider any primitives r, β, c_a, c_s and application limit $\ell > 0$ (with $w = 1$ fixed), such that the resulting equilibrium $(m_\ell^*, \alpha_\ell^*, p_\ell^*, q_\ell^*)$ satisfies $m_\ell^* = \ell$ and $\alpha_\ell^* = 1$. If the platform sets a fixed wage $w_\ell = (c_a/p_\ell^*)\exp(\ell p_\ell^*) \in (c_a/p_\ell^*, 1]$ and no application limit, the resulting equilibrium is again $(m_\ell^* = \ell, \alpha_\ell^* = 1, p_\ell^*, q_\ell^*)$. In particular, aggregate welfare is identical in the two settings.*

The intuition for Theorem 7 is that, in order to achieve the desired application intensity ℓ via a wage, we simply set w_ℓ in a manner that $\mathcal{M}_{w_\ell}(p)$ for $p = p_\ell^*$ is equal to ℓ. We show that $w_\ell \leq 1$ (as one might expect). Because the employers found it worthwhile to screen (that is, $\alpha_\ell^* = 1$) at q_ℓ^* with a wage $w = 1$, they continue to screen when the wage is $w_\ell \leq 1$. The total welfare Π_{tot} remains the same as if there were an application limit of ℓ because the wage is merely a transfer and does not affect the total welfare.

We close with two remarks. First, the *distribution* of welfare is not identical in the two equilibria: Applicant welfare is lower (and employer welfare higher) when wages are w_ℓ than they are with an application limit of ℓ and a wage of $w = 1$. We return to this point in our closing discussion. Second, by increasing the wage, the platform can increase application intensity: an effect that an application limit cannot achieve.[18]

5.5. Heterogeneity

Our model to this point has consisted of *homogeneous* employers and applicants. This section relaxes this assumption through two extensions: one with heterogeneous employers, and another with heterogeneous applicants. In our first extension, employers have a "type" β drawn according to F_e. Any applicant to a job posted by an employer of type β is compatible with probability β. In our second extension, applicants vary in their "skill" β, which is drawn according to F_a. An applicant with skill β is compatible with probability β for each job they apply to. In both extensions, as in the base model, each match partner receives utility one from a compatible match. We treat an employer's type or an applicant's skill as private information: known to the individual, but not directly observable to others. Online Appendix A provides more details, including the definition of a mean-field equilibrium for each setting, as well as expressions for the welfare of each side.

We numerically investigate the effect of imposing an application limit that holds uniformly across applicants. We find that, compared with our baseline model, the effects of this limit can be complex, as we now discuss.

With *employer heterogeneity*, an application limit may make the market attractive for employers with difficult jobs (low β), who would otherwise choose not to participate. Although a limit unambiguously benefits these employers, it may reduce the match rate for employers with easier jobs, who were already choosing to screen.

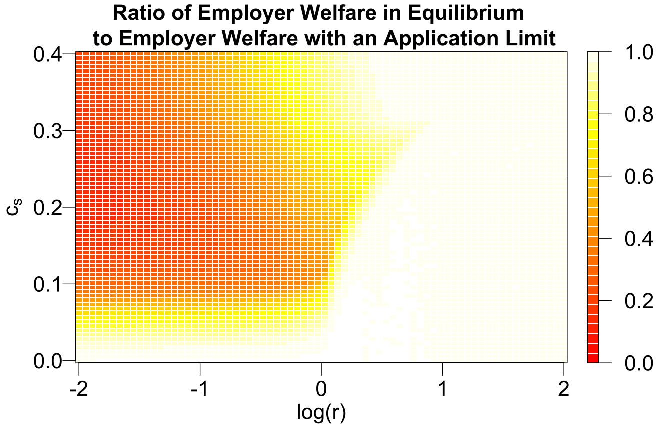
With *applicant heterogeneity*, an application limit may be binding for some applicants and not others. In some cases, the limit binds only for low-skill applicants, thereby unambiguously helping high-skill applicants and increasing the average skill perceived by employers. In others, the limit binds only for high-skill applicants. The resulting reduction in competition may attract low-skill applicants who had previously abstained.

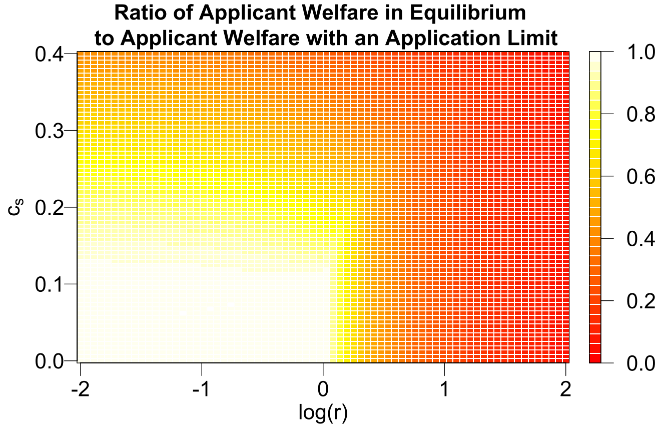
A complete analysis of these effects is beyond the scope of this work.[19] We focus on setting an application limit to maximize the *aggregate* welfare on one side. Figures 4 and 5 show the potential benefit of an application limit for each side. Comparing to Figure 2, there are some differences: For example, with heterogeneous employers, employer welfare is never identically zero, and an application limit has little benefit when $r > 1$ (even when screening costs are high). However, the bottom line is that *our key insights remain valid*: We observe that an application limit significantly benefits employers when there is a shortage of applicants ($r < 1$) and c_s is not too small, and significantly benefits applicants when either r or c_s is large.

6. Mean-Field Approximation

We show that our mean-field model is asymptotically valid for our finite system when the market grows large. First, we show in Theorems 8 and 9 that the mean-field assumptions hold as $n \to \infty$, as long as all applicants a choose $m_a = m$, and all employers e choose $\alpha_e = \alpha$.[20] Second, we use the preceding results to show Theorem 10, which states that any MFE is an approximate equilibrium in sufficiently large finite markets. Technical details of our approach are in Online

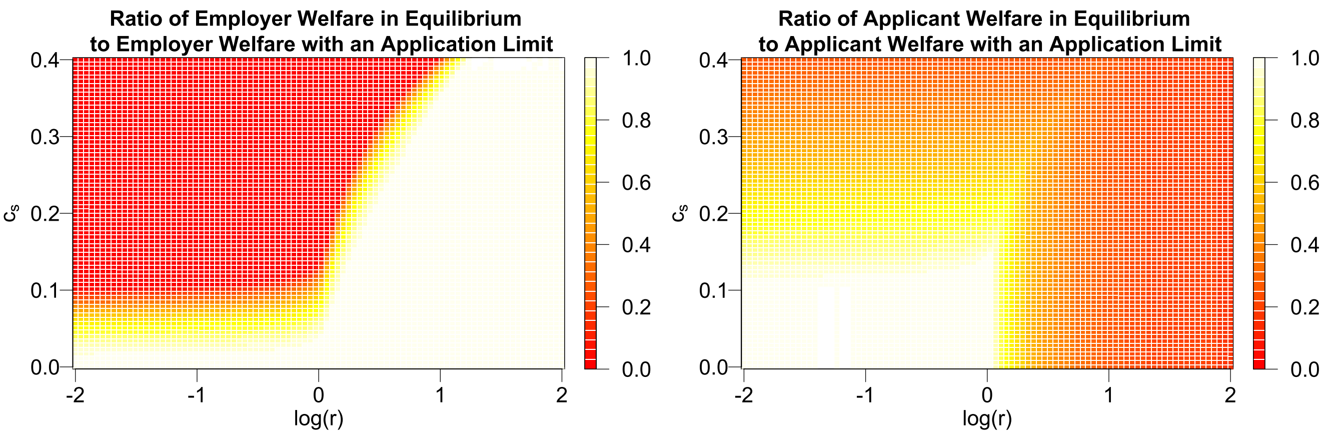
Figure 4. (Color online) A Variant of Figure 2 when β Is Employer-Specific and Uniformly Distributed on the Interval $[0.4, 0.6]$





Note. We fix $c_a = 0.005$ and vary $\log(r)$ (x-axis) and the screening cost c_s (y-axis).

Figure 5. (Color online) A Variant of Figure 2 when β Is Applicant-Specific and Uniformly Distributed on the Binary Set $\{0.4, 0.6\}$



Note. We fix $c_a = 0.005$ and vary $\log(r)$ (x-axis) and the screening cost c_s (y-axis).

Appendix D. Proofs are in the full technical report (Arnosti et al. 2019).

We let Binomial(n, p) denote the binomial distribution with n trials and probability of success p, and let Poisson(a) denote the Poisson distribution with mean a.

In the following two theorems, we show that Mean-Field Assumptions 1 and 2 hold as $n \to \infty$. In the process, we also show Mean-Field Assumption 3, that the number of applications received by an employer and sent by an applicant are each Poisson distributed.

Theorem 8. *Fix r, β, m, and α. Suppose that the n-th system is initialized in its steady-state distribution. Consider any employer e that arrives at $t_e \geq 0$. Let $R_e^{(n)}$ denote the number of applications received by employer e in the n-th system, and let $A_e^{(n)}$ be the number of these applicants that are still available when the employer screens.*

Then, as $n \to \infty$, the pair $(R_e^{(n)}, A_e^{(n)})$ converges in total variation distance to (R, A), where $R \sim$ Poisson(rm), and conditional on R, we let $A \sim$ Binomial(R, q).

Theorem 9. *Fix r, β, m, and α and any $m_0 < \infty$. Suppose that the n-th system is initialized in its steady-state distribution. Consider any applicant a arriving at time $t_a \geq 0$, denote the value chosen by applicant a by m_a (all other applicants are assumed to choose m, and all employers are assumed to follow ϕ^α). Let $T_a^{(n)}$ denote the number of applications sent by applicant a in the n-th system, and let $Q_a^{(n)}$ be the number of these applications that generate offers. Let $T \sim$ Poisson(m_a), and conditional on T, we let $Q \sim$ Binomial(T, p). Then,*

$$\lim_{n\to\infty} \left\{ \max_{m_a \in [0, m_0]} d_{\mathrm{TV}}\left(\left(T_a^{(n)}, Q_a^{(n)}\right), (T, Q) \right) \right\} = 0, \tag{23}$$

where $d_{\mathrm{TV}}(X, Y)$ denotes the total variation distance between the distributions of random variables X and Y that take values in the same countable set.

We use the preceding results on mean-field approximation to establish that any MFE is an approximate equilibrium in sufficiently large finite markets.

Theorem 10. *Fix r, β, c_s and c_a. Let the MFE be (m^*, α^*). For any $\varepsilon > 0$, and any nonnegative integer R_0, there exists n_0, such that for all $n \geq n_0$ the following hold:*

1. *For any applicant a, if all other agents follow their prescribed mean-field strategies, then applicant a can increase her expected payoff by no more than ε by deviating to any $m \neq m^*$.*
2. *For any employer e, if e receives no more than R_0 applications, and if all other agents follow their prescribed mean-field strategies, then in any state in the corresponding dynamic optimization problem solved by employer e (cf. Online Appendix C), employer e can increase her expected payoff by no more than ε by deviating to any strategy other than ϕ^{α^*}.*

The first statement establishes that applicants cannot appreciably gain by changing the number of applications they send. The second statement makes an analogous claim for employers. In particular, if employers follow ϕ^{α^*}, they will either exit immediately or sequentially screen and make offers to compatible candidates until such an offer is accepted or the applicant pool is exhausted. In doing so, they will obtain information about the compatibility and availability of the subset of applicants they have already screened. Our result states that at any stage in this dynamic optimization problem, the employer cannot appreciably increase their payoff by deviating.[21]

7. Conclusion

Our paper presents a benchmark dynamic matching model to study welfare in markets with costly screening and uncertain availability. We characterize welfare in equilibrium for both applicants and employers, and we show how introducing an appropriately chosen application limit can significantly improve welfare for one or both sides.

We also compare the effects of an application limit to those of other available levers: either raising application costs or lowering the wage paid to applicants. Although these interventions can lead to the same aggregate welfare as an application limit, they differ in how they *distribute* this welfare. Charging fees and lowering wages both increase aggregate welfare at the expense of applicants. Although these interventions may be appropriate for a platform looking to monetize its services or attract more employers, an application limit can yield Pareto improvements in welfare and may be more suitable if the platform is primarily concerned with applicant welfare. These considerations might explain why the tutoring platform TutorZ charges tutors for each potential client that they contact, whereas the dating platforms Coffee Meets Bagel and Tinder limit the number of likes/right swipes permitted in a certain period.

Of course, the choice between these interventions is also influenced by technological or cultural constraints not modeled in this paper. For example, an application limit may be unenforceable if applicants can easily create multiple accounts. In a setting with heterogeneity, raising the application cost might be more suitable than a one-size-fits-all application limit. Conversely, charging fees to apply or intervening in the wage-setting process may be less palatable than a simple application limit. We do not assert that any one approach is universally preferable to the others: Each can improve efficiency by reducing the effort required to find a suitable match. This work helps to show when (and for whom) the potential improvements are greatest.

Acknowledgments

Y. Kanoria thanks Microsoft Research New England. The full technical report corresponding to this paper is available on SSRN (Arnosti et al. 2019).

Endnotes

[1] Fradkin (2017) shows that on Airbnb, 49% of inquiries are rejected or ignored by the host, and only 15% of inquiries eventually lead to a transaction. This work concludes that one of the most common reasons for rejection is a "stale" (or unavailable) listing, and that an initial rejection decreases the probability that the guest eventually books *any* listing by 50%. Horton (2019) finds a similar phenomenon on an online labor market.

[2] As noted above, prior academic work establishes that uncertainty about availability is a significant challenge in markets with endogenous prices, and we believe it to be at least as much of a concern in platforms where prices cannot be adjusted to signal availability.

[3] This may help to explain why participants in many online marketplaces leave the platform before finding a match. For example, Fradkin (2017) finds that on Airbnb, hosts rejecting proposals from searchers (potential guests) leads searchers to leave the Airbnb platform; in particular, it causally reduces the likelihood the searcher will eventually book a listing. Horton (2019) finds that in a large online labor market, when an employer pursues a worker but is rejected, this causally reduces the likelihood that the job opening will ultimately be filled.

[4] As an example, the online labor markets Elance and oDesk, which merged to form Upwork, used to have limits on the number of applications freelancers could send in a fixed time period. More recently, Upwork has moved to a system with paid Connects (Field 2019) to address congestion inefficiency.

[5] Costly screening and uncertain availability imply that there is a "congestion externality" among employers. As we argue in Online Appendix B, when employers make multiple offers, this externality is much more severe than if employers made, at most, one offer. Correspondingly, the benefit of an application limit (which reduces this congestion externality) is small or nonexistent in models where employers make only a single offer.

[6] Although Albrecht et al. (2006) numerically study an extension in which employers can make a second offer if their first is declined, they assume that second-round offers come after first-round offers. This implies that even an employer who is constrained to make a single offer is never harmed by the fact that others make more: The presence of multiple rounds of offers *alleviates*, rather than *exacerbates*, competition among employers. Thus, their model (with or without the extension) would predict that an application limit does not significantly benefit employers.

[7] An additional consequence of modeling screening as costly and endogenous is our finding that firm welfare may fall severely—for example, all the way to the value of the outside option when firms are homogeneous—a conclusion that does not hold when screening is costless.

[8] We could instead allow applicants to directly select the number of employers to whom they apply. We chose a probabilistic specification primarily for technical convenience; we further discuss this point in Section 4.

[9] In fact, we can relax this assumption: All of our results hold if we only assume that *conditional on applicant a having applied to an employer e*, e finds a acceptable with probability β. In particular, this allows the possibility that the parameter β is not necessarily exogenous to the platform: Instead, it can be interpreted as the quality of the platform's recommendation and/or search algorithms—that is, improved search and recommendation algorithms simply increase the value of β.

[10] The exception is Section 5.4, where we allow the platform to set the wage w. There, we define $v = 2 - w$ (meaning that the employer utility is the match surplus minus the wage), and we *do* consider the total welfare across the two sides of the platform.

[11] Recall that our applicants apply to each employer independently with probability m/n. Allowing applicants to directly choose how many applications to send complicates our analysis of the finite system. However, in the mean field, this alternative model remains quite tractable: This simply changes the probability of receiving an

offer from $1 - e^{-mp}$ to $1 - (1 - p)^m$. As we focus on settings in which application costs are small (and, thus, m is large and p is small), this does not substantively change our results.

[12] The expression for applicant welfare (15) comes from substituting the applicant best-response function (3) into (10). The expression for employer welfare (16) comes from applying the consistency conditions (6) and (7) to (11) and then making use of the applicant best-response function (3).

[13] We can interpret (13) as follows. Recall that by (3), an applicant facing acceptance probability $p > c_a$ chooses $m = \frac{1}{p}\log(p/c_a)$ and matches with probability $1 - c_a/p$. Thus, the left side of (13) represents the mass of matched applicants. Meanwhile, if employers always screen, then they match whenever they have a qualified available applicant, which occurs with probability of $1 - e^{-rm\beta q^*}$, which equals the right side of (13) using Equation (6).

[14] In particular, when $r > 1$ a necessary condition for the market to be screening-limited is that $c_s' \geq (r\log(\frac{r}{r-1}))^{-1}$. Even for $r = 1.05$, this requires c_s' to be above 0.31; for $r = 1.25$, c_s' must be approximately 0.5. A second necessary condition for the market to be screening-limited is that $c_s' \geq (1 - c_a')/\log(1/c_a')$. For $c_a' = 0.05$, this implies that c_s' must be above 0.31.

[15] In this case, both sides do fairly well, and an application limit cannot significantly help, as shown in Figure 2.

[16] Both numerical results and intuition suggest that the wishes of applicants and employers are most at odds when c_a' is small and c_s' is large, or vice versa. We can prove that as $c_a' \to 0$ and $c_s' \to 1$, or as $c_a' \to 1$ with any fixed c_s', we have $\delta \geq 3/4$.

[17] We assume that the costs paid by applicants cannot be redistributed. This is reasonable if the increased cost is not a monetary transfer, but instead an additional barrier to application (such as additional questions that each applicant must answer). Even when the additional cost is monetary, redistribution may be logistically challenging, may create incentives for individuals to create multiple accounts, and may be undesirable from the point of view of the platform operator.

[18] Numerical results suggest that, within our model, this is unlikely to be beneficial. The benefit of raising wages is highest in a setting with small screening cost and large application cost. Setting (small) $c_s' = c_s/\beta = 0.01$ and $c_a = 0.01$, we found numerically that across all values of r, the total welfare improvement from raising w is no more than about 0.3%. Even with an implausibly large $c_a = 0.1$ and the same $c_s' = 0.01$, across all values of r, the total welfare improvement from raising w is no more than about 6%.

[19] In principle, the mean-field models described in Online Appendix A could accommodate heterogeneity in both the compatibility parameter β and the application or screening cost. Heterogeneous costs would cause participants to make different choices and prefer different application limits. Because these features are present in the model with heterogeneity in β alone, our numerical results focus on this case.

[20] Note that once we fix m and α, we have removed any strategic element from the evolution of the n-th system, and so our results are limit theorems about a certain sequence of stochastic processes.

[21] An apparent limitation of the second statement is that it applies only to employers who receive no more than R_0 applications. The fraction of employers who receive more than R_0 applications scales as $\exp(-\Omega(R_0))$; thus, by choosing R large enough, this fraction can be made as small as desired. Because we show that the n-th system satisfies an appropriate stochastic contraction condition, we expect that, in fact, for sufficiently large R_0, both statements in Theorem 10 hold even under arbitrary behavior by employers who receive more than R_0 applications. In the interest of brevity, we choose to omit this slightly stronger result.

We make the additional simplifying assumption that applicants cannot observe the number of other applicants to each job. If this assumption were relaxed, then in the mean-field analysis each employer would receive rm applications. The probability that an employer matches in this model would become $1 - (1 - \beta q)^{rm}$ (as opposed to $1 - e^{-rm\beta q}$ in our model). Again, this does not substantively change our results.

References

Albrecht J, Gautier PA, Vroman S (2006) Equilibrium directed search with multiple applications. *Rev. Econom. Stud.* 73(4):869–891.

Allon G, Bassamboo A, Cil E (2012) Large-scale service marketplaces: The role of the moderating firm. *Management Sci.* 58:1854–1872.

Anderson R, Ashlagi I, Gamarnik D, Kanoria Y (2017) Efficient dynamic barter exchange. *Oper. Res.* 65(6):1446–1459.

Arnosti N, Johari R, Kanoria Y (2019) Managing congestion in matching markets. Preprint, submitted November 12, http://dx.doi.org/10.2139/ssrn.2427960.

Ashlagi I, Jaillet P, Manshadi VH (2013) Kidney exchange in dynamic sparse heterogenous pools. Preprint, submitted January 15, https://arxiv.org/abs/1301.3509.

Balseiro SR, Besbes O, Weintraub GY (2015) Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Sci.* 61(4):864–884.

Burdett K, Shi S, Wright R (2001) Pricing and matching with frictions. *J. Political Econom.* 109(5):1060–1085.

Coles P, Cawley J, Levine PB, Niederle M, Roth AE, Siegfried JJ (2010) The job market for new economists: A market design perspective. *J. Econom. Perspect.* 24(4):187–206.

Coles P, Kushnir A, Niederle M (2013) Preference signaling in matching markets. *Amer. Econom. J. Microeconom.* 5(2):99–134.

Diamond PA (1982a) Aggregate demand management in search equilibrium. *J. Political Econom.* 90(5):881–894.

Diamond PA (1982b) Wage determination and efficiency in search equilibrium. *Rev. Econom. Stud.* 49(2):217–227.

Field H (2019) Upwork to charge freelancers to bid on jobs, further squeezing those who rely on the 'gig economy.' *Entrepreneur* (April 3), https://www.entrepreneur.com/article/331700.

Fradkin A (2017) Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb. Working paper, Boston University, Boston.

Galenianos M, Kircher P (2009) Directed search with multiple job applications. *J. Econom. Theory* 144(2):445–471.

Galenianos M, Kircher P (2012) On the game-theoretic foundations of competitive search equilibrium. *Internat. Econom. Rev.* 53(1):1–21.

Gurvich I, Ward A (2014) On the dynamic control of matching queues. *Stochastic Systems* 4(2):479–523.

Halaburda H, Jan Piskorski M, Yıldırım P (2018) Competing by restricting choice: The case of matching platforms. *Management Sci.* 64(8):3574–3594.

Horton JJ (2019) Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution *Management Sci.* 65(8):3518–3540.

Hu M, Zhou Y (2018) Dynamic type matching. Rotman School of Management Working Paper 2592622, University of Toronto, Toronto.

Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. *Management Sci.* 60(12):2949–2970.

Julien B, Kennes J, King IP (2000) Bidding for labor. *Rev. Econom. Dynam.* 3(4):619–649.

Kircher P (2009) Efficiency of simultaneous search. *J. Political Econom.* 117(5):861–913.

Kushnir A (2013) Harmful signaling in matching markets. *Games Econom. Behav.* 80:209–218.

Lasry JM, Lions PL (2007) Mean field games. *Japanese J. Math.* 2(1):229–260.

Lee S, Niederle M, Kim HR, Kim WK (2011) Propose with a rose? Signaling in internet dating markets. *Experiment. Econom*. 18(4): 731–755.

Moen ER (1997) Competitive search equilibrium. *J. Political Econom.* 105(2):385–411.

Mortensen DT (1982a) The matching process as a noncooperative bargaining game. McCall JJ, ed. *The Economics of Information and Uncertainty* (National Bureau of Economic Research, Cambridge, MA), 233–258.

Mortensen DT (1982b) Property rights and efficiency in mating, racing, and related games. *Amer. Econom. Rev.* 72(5):968–979.

Pissarides CA (1984a) Efficient job rejection. *Econom. J.* 94:97–108.

Pissarides CA (1984b) Search intensity, job advertising, and efficiency. *J. Labor Econom.* 2(1):128–143.

Rogerson R, Shimer R, Wright R (2005) Search-theoretic models of the labor market: A survey. *J. Econom. Literature* 43(4):959–988.

Roth AE, Xing X (1997) Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *J. Political Econom.* 105(2):284–329.

Weintraub GY, Benkard CL, Van Roy B (2008) Markov perfect industry dynamics with many firms. *Econometrica* 76(6):1375–1411.

Wright R, Kircher P, Julîen B, Guerrieri V (2020) Directed search: A guided tour. *J. Econom. Literature.* Forthcoming.

Ye P, Qian J, Chen J, Wu Ch, Zhou Y, De Mars S, Yang F, Zhang L (2018) Customized regression model for Airbnb dynamic pricing. *Proc. 24th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery, New York), 932–940.