Randomness and Fairness in Two-Sided Matching with Limited Interviews

Hedyeh Beyhaghi
Toyota Technological Institute at Chicago, IL, USA
hedyeh@ttic.edu

Éva Tardos
Department of Computer Science, Cornell University, Ithaca, NY, USA
eva.tardos@cornell.edu

Abstract

We study the outcome in a matching market where both sides have limited ability to consider options. For example, in the national residency matching program, doctors are limited to apply to a small set of hospitals, and hospitals are limited by the time required to interview candidates.

Our main findings are the following: (1) In markets where jobs can only consider a limited number of candidates for interview, it increases the size of the resulting matching if the system has a limit on the number of applications a candidate can send. (2) The fair system of all applicants being allowed to apply to the exact same number of positions maximizes the expected size of the matching. More particularly, starting from an integer k as the number of applications, the matching size decreases as a few applicants are allowed to apply to one additional position (and then increases again as they are all allowed to apply to $k+1$). Although it seems natural to expect that the size of the matching would be a monotone increasing and concave function in the number of applications, our results show that neither is true. These results hold even in a market where a-priori all jobs and all candidates are equally likely to be good, and the judgments of different employers and candidates are independent.

Our main technical contribution is computing the expected size of the matching found via the deferred acceptance algorithm as a function of the number of interviews and applications in a market where preferences are uniform and independent. Through simulations we confirm that these findings extend to markets where rankings become correlated after the interviews.

2012 ACM Subject Classification Theory of computation → Algorithmic mechanism design

Keywords and phrases Matching with Short Lists, Stable Matching, Balls in Bins Problem

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.74

Funding *Hedyeh Beyhaghi*: The author was supported in part by NSF grant CCF-1563714.
Éva Tardos: Supported in part by NSF grants CCF-1408673, CCF-1563714 and AFOSR grant FA9550-19-1-0183.

Acknowledgements The authors are thankful to Mark Braverman and Robert Kleinberg for fruitful discussions.

1 Introduction

Matching is a fundamental paradigm in a variety of real-world situations and arises in various domains, e.g., students applying to attend schools, or colleges, applicants get matched with jobs at various job markets, and medical residents get matched with hospitals, just to name a few. In some of these domains, matchings are found through a centralized algorithm such as the deferred acceptance algorithm by Gale and Shapley [5]. Prominent examples include: national residency matching program (NRMP) for assigning medical students to hospital residency programs; schools assignments in some cities the US; and college admissions in

12th Innovations in Theoretical Computer Science Conference (ITCS 2021).
Editor: James R. Lee; Article No. 74; pp. 74:1–74:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

some countries [10]. Although the deferred acceptance algorithm is designed for centralized markets, it can also be thought of as a rough model of how uncoordinated matching markets as applicants collect offers and as they turn down offers, new offers may be made.

An important aspect that is not modeled under the classical deferred acceptance algorithm is the fact that both sides of the market can only consider a limited set of options. This is the issue we will focus on in the current paper. The deferred acceptance algorithm is typically studied assuming that participants express their full list of ordered preferences. However, typically both sides of the match are limited in the number of options they can consider. In some markets, such as the national residency matching program, there are application limits imposed by the system. But even without such explicit limits, preparing applications and interviewing applicants is costly and time consuming; and hence typically extremely limited. Applications, e.g. for college admission, often require writing specialized essays. Residency programs, as well as some of the colleges, and almost all jobs, interview their applicants, which requires significant time and effort. In large matching markets it is not feasible for either side to consider a large set. For instance, in NRMP, there are nearly 5000 residency programs and applicants have to limit their choices. Similarly, interviewing consumes significant time of the hospitals, and so hospitals can only grant interviews to a limited set of doctors. An alternative view of the limitations of the employer side is a limit on the offers they can make for one position: based on their interviews with all or a subset of the applicants, they may consider only a limited number of their favorite applicants. Such selective offer making is observed in the academic job markets, e.g., when some high ranked departments prefer to wait till next year, rather than make offers to applicants they liked less.

The goal of this paper is to study the effect of the number of applications and the number of interviews on the resulting matching. We think of our procedure as a simplified version of the matching process of the national residency matching program. In this program, the matching mechanism has two stages. In the first stage, doctors express their interest to a set of hospitals, and hospitals choose to interview a subset of the applicants. In the second stage, both doctors and hospitals submit their ordered preference list over the set they interviewed with. The centralized matching market uses these preference lists to perform the deferred acceptance algorithm and outputs the final matching.

We will primarily consider a *symmetric* market, where doctors and hospitals are similar in terms of popularity, and as a primary metric of the social welfare of the system, we will mainly consider the size of the matching found. In asymmetric markets, where some applicants and some jobs are a-priori better and preferred by most applicants, the limit on the number of choices gives rise to strategic issues resulting in top candidates getting too many offers, and a number of jobs remaining unfilled – see the discussion in related literature section. In this paper we isolate the effect of limitation from strategicness. We focus on large symmetric markets, where it will be equilibrium for both sides to report their true preferences. We find a number of surprising effects of this limited choice, resulting in two policy recommendations.

- With limited ability granting interviews, social welfare may be maximized by also severely limiting the number of applications one applicant can send.
- The best social welfare is achieved by a fair system, where all applicants can send the same number of applications. Allowing a small subset to send even just one additional application decreases the overall welfare of the system.

1.1 Main Results

The main contribution of the paper is studying the effect of limitation on the number of applications and interviews in matching markets. In contrast to the majority of the previous work that focused on correlated preferences, we completely abstract away the role of correlation, consider a purely random model, and study the role of randomness extensively. The main results of the paper are as following.

- We compute the expected size of matching resulted from the deferred acceptance algorithm as a function of the limit on the number of applications and interviews. See Proposition 6, Theorem 10, and Proposition 20.
- We find the optimal way for distributing the applications among applicants, when the total number of allowed applications is fixed. The maximum size of matching is achieved when all doctors apply to the same number of positions or two consecutive integers. See Proposition 11.
- With a fixed number of interviews per position, the size of the matching as a function of the number of applications is a scallop-shape figure (see Figures 2 to 5), where the function between consecutive integers is U-shaped.

Contribution to Balls in Bins Literature

Numerous findings in the matching literature utilize the results and techniques form *balls in bins* problems. Although this paper is written in the matching language, it has analogues in balls in bins language and contributes to that literature as well. For example consider the following problem: There is a set of n balls and n bins. Each ball d has k_d copies. Each bin has capacity one. All copies of the balls are thrown independently and uniformly at random to the bins. Each bin only accepts one ball at random among those that have been thrown to it. We are interested in $\mathcal{S}$ which is the expected number of unique balls (counting only a single copy from each ball) that land in bins. Various questions can be asked in this scenario. For example, what is the optimal value of S? What distribution for k_d achieves this value? Section 3, which is a warm-up for our main result, completely analyses this problem and surprisingly shows the optimal value of S is ≈ 0.68 for large enough n and it is achieved when there are 3 copies from each ball, $k_d = 3$.

1.2 Related Work

Papers studying matching with short lists have different viewpoints on how the short lists are selected. For example, Immorlica and Mahdian [6], Kojima and Pathak [8], and Arnosti [1] study matchings when preference lists are inherently short, i.e. the participants prefer to stay unmatched rather than being matched outside their preference lists. In contrast, Avery and Levin [2], Kadam [7], Drummond et al. [4], and Beyhaghi et al. [3] study how applicants make the strategic choice to select a limited number of positions for their short preference lists. In this paper, because of the uniform preference of participants we do not focus on their strategic behavior.

Arnosti [1], Lee and Schwarz [9], and Kadam [7] study efficiency of matching either as matching size or social welfare in presence of short lists. Arnosti evaluates social welfare under different preference models. However, unlike our paper, in [1] the preference lists are limited only for one side of the market. Similar to us, Lee and Schwarz study matching size in a setting with an interview stage, where the ex-ante preferences are i.i.d. However, they solve the problem of some doctors not receiving any offer while others receive many, by coordinating the set of doctors that each hospital interviews. In contrast, in this paper, we assume that the interviews are selected in a decentralized way without imposing coordination; and our solution is to limit the number of applications and let the applicants have the same number of applications. Similar to our work, Kadam studies a model with limit on both sides of the market and shows that limiting the length of lists can have positive effects of the size of matching. However, in [7] these effects are due to almost common preferences: When some doctors are more preferred, with a stricter limit on the interviews for doctors, the more preferred doctors do not accept interviews with less preferred hospitals. In this paper, we show that limiting the length can be helpful in the exactly opposite case where all participants are identical in terms of popularity.

Roadmap

The rest of the paper is organized as follows. Section 2 discusses the model and preliminaries. In Section 3, as a warm-up we analytically compute the size of the matching when the number of interviews is limited to one. In Section 4, we compute the size of the matching for arbitrary number of interviews. Section 5 discusses two extensions to the main result in Section 4. In Section 5.1, we show that the same phenomena remain true in unbalanced markets (when the two sides are of different size). In Section 5.2, we study a model where the hospitals and doctors preferences are a-priori uniform, but become correlated after interviewing.

2 Preliminaries

We explain our model in Section 2.1, and discuss the unlimited interview case in Section 2.2.

2.1 Model

There is a finite set of doctors $\mathcal{D}$ and a finite set of hospitals $\mathcal{H}$, with $|\mathcal{D}| = n$ and $|\mathcal{H}| = rn$. We are interested in one-to-one matchings between doctors and hospitals.

Description of the Game

We consider a two-stage matching mechanism. In the first stage, each doctor applies to a set of hospitals and requests interviews. Hospitals then choose a subset from their applications to conduct interviews. In the interview process, doctors and hospitals refine their preferences. In the second stage, at the end of interviews, both doctors and hospitals order the list that they interviewed with, based on their preferences. They submit their ordered preferences to the system which performs a doctor-proposing deferred acceptance algorithm to determine the final assignment of doctors to hospitals. The algorithm starts with all doctors unmatched. In each step, the algorithm simulates all unmatched doctors, who have not yet exhausted their options, propose to their most preferred hospital among those to which he/she has not yet proposed. Now the algorithm simulates that each hospital tentatively accepts their most preferred doctor from the doctors now proposing and the one who has been tentatively assigned to this hospital, and rejects all other doctors. The procedure is repeated until all unmatched doctors have been rejected from every hospital in their lists.

We use the following terminology throughout the whole paper.

Application: The procedure in the first stage where the doctors submit an application and ask for interview.

Grant interview or reject for interview: The procedure in the first stage where the hospitals grant interviews to a number of doctors that have applied for the position and reject the rest.

Valid application: Any application that is granted an interview is a valid application. Other applications are *invalid*.

Proposal: The procedure in the second stage that simulates doctors proposing to hospitals as steps of the deferred acceptance algorithm.

Offer: In the special case where the hospitals limit their number of interviews to one, conducting interviews is of no use because the hospital offers the position to the selected doctor in any case. Therefore, we use offer instead of interview in this case.

Limits on Applications and Interviews

Each doctor d is allowed to apply to k_d positions. In the simpler case, there is a universal limit on the number of applications, $k_d = k$, inspired by NRMP granting all doctors 10 initial applications. There is a universal limit k' on the number of interviews each hospital can conduct.

Random Preferences

For the most of the paper, we assume that everybody's preference comes independently from uniform distribution over the other side. Both doctors and hospitals refine their preference order for the list they interviewed. However, the overall distribution of doctors preferences and hospitals preferences stays uniform and independent. For analytic purposes we can assume that preferences stay unchanged after the interviews.

▶ **Observation 1.** *Suppose prior to interviews preferences are drawn independently from uniform distributions, and after interviews, the overall distribution of preferences are uniform and independent. In this case, assuming that the preferences remained unchanged leads to the same analytic results including the same matching size.*

▶ **Observation 2.** *Being truthful is a Bayes Nash equilibrium.*[1] *We assume both doctors and hospitals follow this strategy and are truthful during the matching procedure: In the first stage, doctors apply to their favorite hospitals; and hospitals grant interviews to their favorite applicants. In the second stage, both doctors and hospitals submit the list that they interviewed, ordered from the most preferred to the least preferred.*

Efficiency Measure

We focus on the size of the matching as our notion of efficiency. We define social welfare as the ratio of size of the matching outcome to the size of the maximum matching. Since we assume all doctors/hospitals prefer to be matched to any hospital/doctor rather than being unmatched, in a maximum-size matching everybody on the less populated side of the market is matched.

▶ **Definition 3** (Social Welfare). *We define social welfare of the matching outcome as the ratio of the size of the matching outcome compared to the size of the maximum-size matching. In a two sided market, maximum-size matching will be the size of the smaller side of the market.*

[1] For more discussion, see the appendix.

Large Markets

The results consider outcome in "large markets" as introduced in [1]. Formally, a sequence of markets indexed by n are considered. In this sequence, the number of applications k, the number of interviews k', and the ratio of number of hospitals to doctors r are held constant. The n^{th} market is characterized by a set of doctors $\mathcal{D}^n$ and a set of hospitals $\mathcal{H}^n$. Consider $n = |\mathcal{D}^n| = |\mathcal{H}^n|/r$. The results in this paper study the properties of the matching for $n \to \infty$.

2.2 No Limit for Granting Interviews

In this section, we recall an analysis from [3] with n doctors and n hospitals in the market. In this case there is no limit on the number of interviews by hospitals and doctors are allowed to list only up to k hospitals, for $k << n$. When doctors apply to k hospitals, Proposition 4 finds the probability p of a single random proposal resulting to a permanent match as n grows to infinity. The main idea is that in the limit as n goes to infinity, the probability of a proposal resulting in a match is independent of the previous proposals of the applicant being rejected.

▶ **Proposition 4** ([1]). *When each doctor applies to k hospitals, the probability p of a single random proposal resulting in a permanent match in the deferred acceptance procedure, satisfies the following equation.*

$$(1-p)^k = e^{-(1-(1-p)^k)/p} \tag{1}$$

Proof sketch. Since the outcome of the deferred acceptance algorithm does not depend on the order in which doctors propose, we may hold out a single doctor d and run the deferred acceptance algorithm on the remainder of the market. Now consider doctor d proposing to her favorite position. Her first few proposals may get rejected. Once a hospital accepts her proposal, it may reject a different doctor, who may propose for her next position, etc. We call the resulting sequence of rejections a *rejection chain*. The probability that a proposal of doctor d causes a rejection chain that gets doctor d rejected from the hospital that first accepted her vanishes as the market grows, therefore we may assume that d's first tentatively accepted proposal will lead to a permanent match. Also, in a large market, the rejection of d's first m proposals does not affect the probability of acceptance of other proposals. Thus, from d's perspective, each hospital that she applied to in the first stage, should be available to her with some probability p, and their availability should be independent. With this argument, the probability that d matches is $1-(1-p)^k$, and the expected number of hospitals d proposes to is

$$1 + (1-p) + (1-p)^2 + ... + (1-p)^{k-1} = \frac{1}{p}(1-(1-p)^k).$$

From the point of view of each hospital, each of these proposals is sent to them roughly with probability $1/n$; thus the probability that a hospital receives at least one proposal is

$$1 - (1-1/n)^{\frac{n}{p}(1-(1-p)^k)};$$

which approaches $1 - e^{-(1-(1-p)^k)/p}$ when $n \to \infty$. Since doctors match with probability $1-(1-p)^k$, and the number of doctors and hospitals that match must be equal, we have that $(1-p)^k = e^{-(1-(1-p)^k)/p}$. ◀

▶ **Proposition 5.** *Suppose that either all the doctors apply to the same number of hospitals or a fraction of them apply to k while others apply to $k+1$ hospitals. In a setting where there is no limit on the number of interviews by hospitals and the preferences are uniformly random, the social welfare of the matching is increasing in the expected number of applications.*

Proof. See the appendix. ◀

3 Warm-Up: Single Offer

In this section, we find the efficiency of matching as a function of the expected number of applications by doctors, when hospitals only select one applicant in the first stage of the mechanism. We begin with the case where all doctors apply to the same number of hospitals, $k_d = k$, and then we move to the case where doctors send different number of applications. We find allocating the number of applications equally results in the most efficient matching and allowing a small subset to apply to an extra position or restricting a small set to apply to one less position, results in a smaller matching.

In this part we assume that hospitals only select one of their applicants in the first stage of the matching. As discussed in Observation 2, they choose their favorite applicant. Since interviews are conducted to compare the selected doctors, in this case that only one doctor is selected, interviews are of no use from the hospitals' point of view. Therefore the hospital immediately offers the position to the selected doctor. Consequently, the outcome of the deferred acceptance algorithm can be easily found in this case; each doctor will be matched to his/her favorite hospital among those who gave an offer. Therefore without going through the complication of the deferred acceptance algorithm, each doctor accepts the most preferred offer. This discussion implies that the size of matching in this case, is the number of doctors who receive an offer.

3.1 Same Number of Applications

First, we study the social welfare of the matching with respect to the number of applications by doctors, where all doctors apply to the same number of hospitals, k.

▶ **Proposition 6.** *Suppose for all d, $k_d = k$ and $k' = 1$. The social welfare equals* $1 - (1 - \frac{(1-e^{-k})}{k})^k$.

To illustrate some of the main points in the proof of the proposition we start with a simple example in which doctors are allowed to send one application.

▶ **Example 7.** When $k = k' = 1$, the social welfare of the matching approaches $(1-1/e) \approx 0.63$ in a large market. In a matching between hospitals and doctors the size of matching equals the number of matched hospitals. Each hospital that receives an application is matched. The reason is when a hospital accepts a doctor, that doctor does not have any other offers and accepts the match. The probability of a hospital receiving at least one application is $1 - (1 - 1/n)^n$ which tends to $1 - 1/e$ with n approaching ∞.

Now, we consider the general case of arbitrary k. The following definition is central to our proof.

▶ **Definition 8** (covered hospital). *A hospital is* covered *if it receives at least one application. The number of covered hospitals equals the number of offers.*

Proof of Proposition 6. First, we find the expected number of covered hospitals. By Observation 2, doctors apply to their favorite positions. Since doctors preferences are independent, the set of hospitals that a doctor applies to is independent from other doctors. Although a doctor applies to k *different* hospitals and these applications are not technically independent, as n grows large, the dependence between different applications of a doctor becomes negligible tends to 0. In other words, the probability that a fresh random choice of a doctor is identical to a previous choice approaches to 0. Therefore, we assume that different applications of a doctor are independent. The independence among applications of the same doctor and the set of applications of different doctors, implies the nk hospitals chosen by applications are selected uniformly and independently at random. The probability of a hospital receiving any application is $\approx 1-(1-1/n)^{nk}$, which tends to $(1-e^{-k})$ in the limit.

Now, we formulate the probability of a doctor being matched. A doctor is matched if one of its applications turns to an offer. The probability of application (d, h) turning into an offer depends on the number of applications that hospital h has received. (Each hospital gives offer to one of its applications.) However, doctors have no knowledge about the number applications a hospital receives and to them all hospitals look identical in terms of the number of applications they have. Thus, from a doctor's perspective, each of their applications has the same probability of turning into an offer. Let p be the probability of an application turning to an offer. The probability of a doctor being matched equals $1-(1-p)^k$.

So far, we found the probability of a hospital being covered and the probability of a doctor being matched as a function of p. Now, we can relate these two terms. The main observation here is that the probability of an application turning to an offer, p, equals the ratio between the number of offers and all applications. Therefore, $p = \frac{1-e^{-k}}{k}$.

The expected social welfare of matching is equal to the probability of a doctor being matched, which is $(1-(1-\frac{(1-e^{-k})}{k})^k)$. ◀

Figure 1 shows the size of matching with respect to the number of applications.

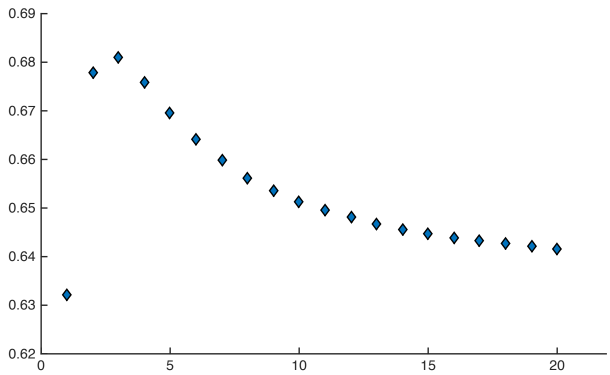

Figure 1 Size of matching with respect to the number of applications, when hospitals make only one offer.

Positive and negative effects of having more applications

As shown in Figure 1, having more applications can have both positive and negative effects: As a positive effect, with more applications the number of hospitals who receive an application increases. This can potentially increase the size of matching. On the other hand, as the number of applications increases, the hospitals become more congested. Since he total number of offers is limited by the number of hospitals (each hospital makes at most one offer), this causes an increase in the probability of rejection of an application. This may increase the number of doctors with no offer, leading to a negative effect on the size of matching. As seen in the picture, The biggest jump in the size of matching occurs when moving from one application to two applications. When increasing the number of applications from one to two, the fraction of covered hospitals changes from $1 - 1/e \approx 0.63$ to $1 - 1/e^2 \approx 0.86$, which is the highest increase when adding more applications. This causes the highest increase in the matching size. The increase stops with only three number of applications, with which the fraction of covered hospitals is $1 - 1/e^3 \approx 0.95$. From this point forward the negative effect takes over and because of the random allocation of hospitals some doctors receive multiple offers while others receive none.

The following example considers the extreme case where doctors apply to all hospitals. Interestingly, the size of the matching in this case equals to the case where doctors were allowed to apply to a single hospital.

▶ **Example 9** (One application equals n applications.). With n applications for each doctor and hospitals accepting one interview, the social welfare of the matching approaches $(1 - 1/e) \approx 0.63$ in a large market.

Proof. Since hospital preferences are uniformly random, each hospital selects an applicant to make an offer to, uniformly at random. Because doctors have n applications, the set of doctors applying to each hospital is the set of all doctors. Therefore the number of doctors who receive an offer is:

$$lim_{n\to\infty} 1 - (1 - \frac{1}{n})^n = 1 - \frac{1}{e}$$ ◀

3.2 Fractional Expected Number of Applications: A Scallop-Shape Function

Unlike Section 3.1, where there was a universal limit on the number of applications by doctors, in this subsection we study the social welfare when doctors are allowed to send different number of applications. First we show that for any expected number of applications x, the optimal social welfare occurs when doctors send either $\lceil x \rceil$ or $\lfloor x \rfloor$ applications. Then we study the social welfare as a function of the expected number of applications. We observe that granting extra applications to a small set of doctors and also retracting an application from a small set both hurt the market; suggesting that unfair treatment is not efficient in terms of social welfare.

▶ **Theorem 10.** *The social welfare of the matching with expected number of applications $k < x < k + 1$, when applicants either apply to k or $k + 1$ positions is:*

$$(\lceil x \rceil - x)(1 - (1 - \frac{(1 - e^{-x})}{x})^k) + (x - \lfloor x \rfloor)(1 - (1 - \frac{(1 - e^{-x})}{x})^{k+1});$$

this function is illustrated in Figure 2.

The following proposition shows that for any expected number of applications x, the optimal social welfare occurs when doctors send either $\lceil x \rceil$ or $\lfloor x \rfloor$ applications.

▶ **Proposition 11.** *With any expected number of applications, x, the distribution of number of applications that achieves the highest social welfare, is one that allocates $\lfloor x \rfloor$ applications to some doctors and $\lceil x \rceil$ to the others, such that the expected number of applications equals x.*

Proof. Suppose the applications are distributed in a different way. Therefore, there are two doctors with l and k applications such that $k - l \geq 2$. We show that the size of matching is improved if we allocate $f = \lfloor (k+l)/2 \rfloor$ and $c = \lceil (k+l)/2 \rceil$ applications to those doctors. This alteration does not change the probability of receiving offers by other doctors as the applications are independent from doctors' perspective. So the only difference is the probability of receiving offers by these two doctors. Let p be the probability of a random application to lead to an offer. We claim

$$1 - (1-p)^k + 1 - (1-p)^l \leq 1 - (1-p)^c + 1 - (1-p)^f.$$

Since $(1-p)^c$ and $(1-p)^f$ have the same product as $(1-p)^k$ and $(1-p)^l$, their sum is larger when the two factors are far apart, therefore:

$$(1-p)^c + (1-p)^f \leq (1-p)^k + (1-p)^l$$

which implies the conclusion. ◀

▶ Remark. This argument shows that in order to find the optimal social welfare for different expected number of applications, we only need to study the case where doctors apply to the same number of hospitals –as studied in Section 3.1– or two consecutive numbers.

Proof of Theorem 10. This proof is similar to that of Proposition 6. From a doctor's perspective, the probability of a random application, leading to an offer is $c/\mathcal{A}$, where c is the number of covered hospitals, and $\mathcal{A}$ is the number of applications. With nx applications, the fraction of covered hospitals equals $(1 - e^{-x})$. Thus, the probability of an application turning to an offer equals $(1 - e^{-x})/x$. Therefore, the expected social welfare of the matching which is the same as the probability of a random doctor receiving an offer is:

$$(\lceil x \rceil - x)(1 - (1 - \frac{(1-e^{-x})}{x})^k) + (x - \lfloor x \rfloor)(1 - (1 - \frac{(1-e^{-x})}{x})^{k+1}).$$ ◀

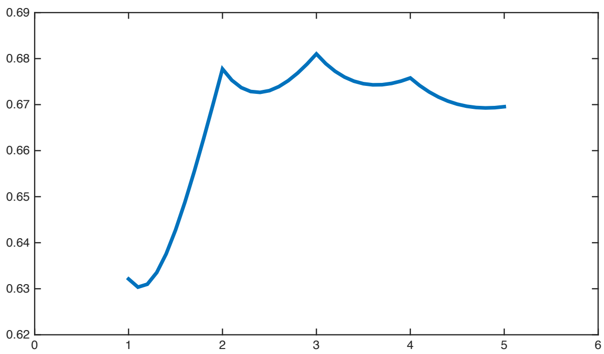

Figure 2 Size of matching with respect to expected number of applications, when hospitals make only one offer.

The scallop-shape Figure 2 illustrates the size of matching as a function of expected number of applications. The unusual behavior of the function at integer points shows that allowing a small group to apply to one more position, or limiting the number of applications of a small group to one less application, has a negative effect on size of matching.

▶ **Observation 12** (3 applications per doctor is the most efficient). *Theorem 10, depicted in Figure 6, and Proposition 11 imply that for any expected number of applications per doctor, and any independent distribution of those applications to doctors, the efficiency of the matching never exceeds* ≈ 0.68*. This efficiency is achieved when all doctors apply to* 3 *hospitals. This is in sharp contrast with unlimited number of interviews (Proposition 5).*

4 Granting multiple interviews

In this section we describe our approach for the general case. The key simplifying feature of the single-offer case was that each hospital immediately made an offer to their top applicant and each doctor was matched to their most preferred hospital that they received an offer from. However, with multiple interviews, finding the final matching requires actually running the usual procedure of the deferred acceptance algorithm.

The main result of this section is the following.

▶ **Theorem 13.** *Suppose each doctor has* k *applications and each hospital grants interviews to at most* k' *doctors. The social welfare of the matching in this case is* $1-(1-p)^k$ *where* $0<p<1$ *is the solution to*

$$1-(1-p)^k = \sum_{i=1}^{k'-1}[g(i,k,p)f(i;k)] + g(k',k,p)(1-F(k'-1;k));$$

where $g(i,k,p) = 1-(1-\frac{1-(1-p)^k}{pk})^i$, $f(i;k) = \frac{k^i e^{-k}}{i!}$ *is the pmf of Poisson distribution with mean* k*, and* $F(\cdot;k)$ *is the CDF of that distribution. Also when the expected number of applications is* x *such that* $x-\lfloor x\rfloor = z$*, the social welfare is upperbounded by* $(1-z)(1-(1-p)^{\lfloor x\rfloor}) + z(1-(1-p)^{\lceil x\rceil})$*, where* $0<p<1$ *is the solution to*

$$(1-z)(1-(1-p)^k) + z(1-(1-p)^{k+1}) = \sum_{i=1}^{k'-1}[h(i,x,p)f(i;x)] + h(k',x,p)(1-F(k'-1;x));$$

and $h(i,x,p) = (1-(1-\frac{(1-(x-\lfloor x\rfloor))(1-(1-p)^{\lfloor x\rfloor})+(x-\lfloor x\rfloor)(1-(1-p)^{\lceil x\rceil})}{px})^i)$*. The upperbound is tight when all doctors apply to either* k *or* $k+1$ *positions.*

Before proceeding with the proof we give a high-level overview of the main steps. In a market of two equal sides the probability of a doctor being matched is equal to the probability of a hospital being matched. We use this equation to find the size of matching. The formula for probability of a doctor being matched is similar to the previous section; however, formulating the probability of a hospital being matched is more complicated. The main technical portion of this section is devoted to formulating this probability. First, we define a modified implementation of the matching procedure which results in the same outcome as the original implementation, but is easier to deal with for analytic purposes. The new implementation introduces a concept called "semi-proposal". Later, we find the relationship between the number of applications, valid applications, proposals, and semi-proposals. The probability of a hospital being matched can be formulated in terms of the number of proposals. Finally, by finding the three other quantities, we find the number of proposals, and formulate the probability of a hospital being matched.

Modified Implementation of Matching Procedure. For the sake of analysis, it is useful to define a *modified implementation* of the matching procedure. Similar to the original implementation, in the first stage doctors apply to hospitals using all their applications and hospitals conduct interviews with a subset of their applicants. The difference between the two implementations arises in the second stage. In the modified implementation, doctors submit their ordered preferences including *both* the hospitals they interviewed and the hospitals that rejected them for interview. Therefore, when the system simulates doctor-proposing deferred acceptance algorithm, doctors are allowed to propose using any of their applications; both valid and invalid ones. Since the invalid applications do not exist in hospitals' lists they get rejected immediately and the outcomes of both implementations are the same. In contrast to the modified implementation, we refer to the main procedure previously defined as the *original* implementation.

Semi-proposal. The procedure in the second stage of the modified implementation that simulates doctors proposing to hospitals as steps of the deferred acceptance algorithm. Note that this includes both the proposals that do not really count (because the hospital didn't extend an interview) and those that led to an interview.

▶ **Lemma 14.** *The expected number of semi-proposals made by a doctor is* $\frac{1-(1-p)^k}{p}$*, where* p *is the probability of a random application turning to a permanent match.*

Proof. By Proposition 4 from a doctor's perspective each hospital they propose to is available to them with equal probability, and their availabilities are independent[2]. Consider the modified implementation defined above. In the second stage of the game doctors send semi-proposals to hospitals. Let p be the probability of a random semi-proposal becoming a permanent match. By independence the probability that a doctor becomes matched is $1-(1-p)^k$. Also the expected number of semi-proposals made by a doctor is $1+(1-p)+\ldots+(1-p)^{k-1} = \frac{1-(1-p)^k}{p}$. ◀

▶ **Lemma 15.** *For any application* $\mathcal{A} = (d, h)$*, consider* $\mathcal{E}_{semi}(\mathcal{A})$ *as the event that* $\mathcal{A}$ *turns to a semi-proposal, i.e.,* d *sends a semi-proposal to* h*, and* $\mathcal{E}_{valid}(\mathcal{A})$ *as the event where* $\mathcal{A}$ *is valid.* $\mathcal{E}_{semi}(\mathcal{A})$ *and* $\mathcal{E}_{valid}(\mathcal{A})$ *are independent.*

Proof. Application $\mathcal{A} = (d, h)$ is turned into a semi-proposal if the previous semi-proposals of d are rejected. Similarly, a valid application $\mathcal{V}$ is turned into a proposal if the previous semi-proposals are rejected. Since preferences are distributed uniformly at random for both doctors and hospitals, the fact that $\mathcal{A}$ is valid or invalid is independent of the rank of h in the preference list of d. Therefore, the probability of $\mathcal{A}$ turning into a semi-proposal conditioned on it being valid is equal to the unconditional probability. ◀

▶ **Corollary 16.** *The probability of a random valid application turning into a proposal,* $p_{V\to P}$*, equals the probability of a random application into a semi-proposal,* $p_{A\to S}$*. This probability is* $\frac{\mathcal{P}}{\mathcal{A}} = \frac{\mathcal{S}}{\mathcal{V}}$*, where* $\mathcal{P}$*,* $\mathcal{A}$*,* $\mathcal{S}$*, and* $\mathcal{V}$ *represent the number of proposals, applications, semi-proposals and valid applications, respectively.*

Proof. Lemma 15 implies $p_{V\to P} = p_{A\to S}$. The first probability equals the ratio of total number of proposals to valid applications, and the second equals the total number of semi-proposals to applications. ◀

A key observation for finding the number of valid applications is the following.

[2] Note that considering proposals *with replacement* is accurate up to lower order terms; with constant proposals, a doctor will make a repeat only with vanishingly small probability.

▶ **Observation 17.** *The number of applications that a hospital receives in the first stage forms a binomial distribution which converges to a Poisson distribution with mean k in the limit. This is due to the independent and uniform preferences of doctors.*

▶ **Lemma 18.** *The expected number of proposals that a hospital receives is*

$$\frac{1-(1-p)^k}{p} \cdot \frac{\sum_{i=1}^{k'-1}\left[i \cdot \frac{k^i e^{-k}}{i!}\right] + k'(1 - e^{-k}\sum_{i=0}^{k'-1}\frac{k^i}{i!})}{k}.$$

Proof. By Corollary 16, $\frac{\mathcal{P}}{\mathcal{A}} = \frac{\mathcal{S}}{\mathcal{V}}$. In order to find $\mathcal{P}$, we need $\mathcal{A}$, $\mathcal{S}$, and $\mathcal{V}$. By definition, $\mathcal{A} = k$ and by Lemma 14, $\mathcal{S} = \frac{1-(1-p)^k}{p}$. Therefore, we only need to find $\mathcal{V}$. Recall that an application (d, h) is valid if hospital h does not reject it in the first stage. A hospital rejects an application only if it receives more than k'. Therefore, by Observation 17, the expected number of valid applications per hospital is,

$$\mathcal{V} = \sum_{i=1}^{k'-1} i \cdot \frac{k^i e^{-k}}{i!} + k'(1 - e^{-k}\sum_{j=0}^{k'-1}\frac{k^j}{j!}).$$

Substituting $\mathcal{A}$, $\mathcal{S}$, and $\mathcal{P}$ in Corollary 16, the expected number of proposals a hospital receives is

$$\frac{1-(1-p)^k}{p} \cdot \frac{\sum_{i=1}^{k'-1} i \cdot \frac{k^i e^{-k}}{i!} + k'(1 - e^{-k}\sum_{i=0}^{k'-1}\frac{k^i}{i!})}{k}.$$ ◀

Now we find the probability that a random hospital is matched.

▶ **Lemma 19.** *The probability of a random hospital being matched is*

$$\sum_{i=1}^{k'-1}[g(i,k,p)f(i;k)] + g(k',k,p)(1 - F(k'-1;k));$$

where $g(i,k,p) = 1 - (1 - \frac{1-(1-p)^k}{pk})^i$, $f(i;k) = \frac{k^i e^{-k}}{i!}$ is the pmf of Poisson distribution with mean k, $F(\cdot;k)$ is the CDF of that distribution, and p is the probability of a random application turning to a permanent match.

Proof. From the hospitals' point of view, each of their valid applications has the same probability of becoming a proposal. Let q be the probability of a random valid application turning into a proposal. In a doctors-proposing deferred-acceptance algorithm, if a hospital is once tentatively matched it will remain matched forever. A hospital becomes tentatively matched if it receives a proposal, in other words if at least one of its valid applications becomes a proposal. The probability that a random hospital with j valid applications receives a proposal is $1-(1-q)^j$. Therefore, the probability that a random hospital is matched equals,

$$\sum_{i=1}^{k'-1}\left[(1-(1-q)^i)\frac{k^i e^{-k}}{i!}\right] + (1-(1-q)^{k'})(1 - e^{-k}\sum_{i=0}^{k'-1}\frac{k^i}{i!}).$$

The probability of a random valid application turning into a proposal, q, is equal to the ratio of the total number of proposals to the total number of valid applications. By Lemma 15, this ratio is equal to the ratio of the total number of semi-proposals to the total number of applications. Therefore, $q = \frac{1-(1-p)^k}{pk}$. ◀

Proof of Theorem 13. Lemma 14, Lemma 19 and the fact that in a matching of two equal sides, the probability of a random doctor being matched is equal to a hospital being matched implies:

$$1-(1-p)^k = \sum_{i=1}^{k'-1} [g(i,k,p)f(i;k)] + g(k',k,p)(1-F(k'-1;k)).$$

Similar reasoning holds when the expected number of applications is an arbitrary real number x such that each doctor has either $\lfloor x \rfloor$ or $\lceil x \rceil$ applications, deriving the formula for general x. By Proposition 11, the size of the matching with expected $k \leq x < k+1$ applications is maximized when all doctors apply to either k or $k+1$ positions. This concludes the proof. ◀

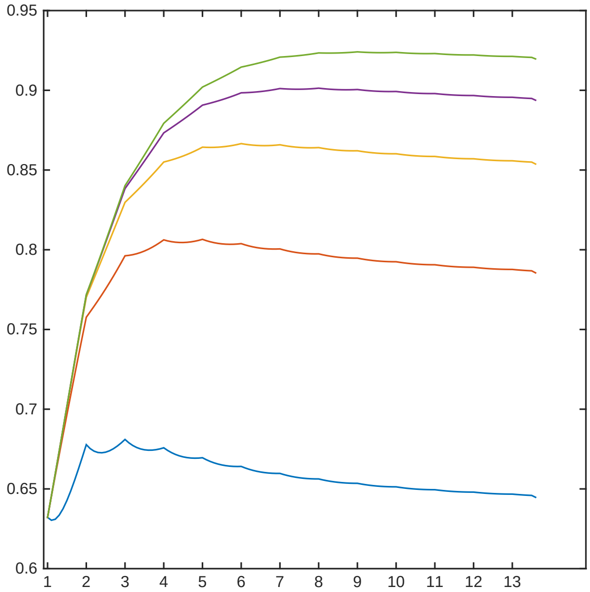

Figure 3 Size of matching with respect to expected number of application. The curves represents the settings where hospitals allow one to five interviews. The lowest curve belongs to one interview and the highest curve to five interviews.

5 Extensions

In this section we discuss two extensions to our main results in Section 4. Section 5.1 discusses the extension to unbalanced markets where the number of doctors and positions is different. Section 5.2 discusses the extension to correlated preferences.

5.1 Beyond Balanced Markets

In this part, we show that the phenomenon of the positive effect of setting the same limit for all doctors is not limited to a balanced market – where the number of applicants and positions are the same – but extends to unbalanced markets with different number of applicants and positions. Although the same phenomenon exits more generally, the exact function is not preserved and the optimum number of applications depends on the ratio between the sizes of the two sides.

We study the case where hospitals make one offer (similar to Section 3) and the ratio of the number of hospitals to the number of doctors is r. The results from Section 4 with multiple interviews also generalize to this setting. However, for simplicity and similarity of the outcomes, we just present the results for the model with one offer. Similar to the previous section, we can compute the size of the matching as a function of the expected number of applications, when doctors send k or $k+1$ applications.

▶ **Proposition 20.** *The expected fraction of matched doctors with expected number of applications $k < x < k+1$, when applicants either apply to k or $k+1$ positions and when the ratio of number of hospitals to number of doctors is r is:*

$$(\lceil x\rceil - x)(1-(1-\frac{r(1-e^{-x/r})}{x})^k) + (x - \lfloor x\rfloor)(1-(1-\frac{r(1-e^{-x/r})}{x})^{k+1}).$$

Proof. As shown in the proof of Proposition 6, from the doctors perspective, the probability of a random application, leading to an offer is $\frac{\mathcal{C}}{\mathcal{A}}$, where $\mathcal{C}$ is the number of covered hospitals and $\mathcal{A}$ is the number of application. From the hospitals perspective, each application is equally likely to be sent to each hospital. Therefore the number of applications received is distributed as a Poisson distribution with $\lambda = \frac{x}{r}$. So the number of expected covered hospitals is $rn(1-e^{-x/r})$ and the probability of a random application, leading to an offer is $\frac{r(1-e^{-x/r})}{x}$. Therefore the expected fraction of doctors who are matched which is the same as the probability of a random doctor receiving an offer is:

$$(\lceil x\rceil - x)(1-(1-\frac{r(1-e^{-x/r})}{x})^k) + (x - \lfloor x\rfloor)(1-(1-\frac{r(1-e^{-x/r})}{x})^{k+1}).$$ ◀

Based on Definition 3, the social welfare of a matching is the ratio of the size of the matching to the size of maximum matching; and in unbalanced markets, the maximum size is the size of the smaller side of the matching. For $r \geq 1$, the doctors make the smaller side therefore the social welfare is equal to expected fraction of doctors who are matched as computed in Proposition 20. If $r < 1$, the social welfare is the expected fraction of doctors who are matched as computed in Proposition 20 divided by r.

Section 5.1 shows the social welfare of the matching as a function of expected number of applications for $r = \frac{1}{2}, 1, 2$. In the figure, the red function refers to $r = 2$, the blue function to $r = 1/2$ and the green function to $r = 1$.

As seen in the figure a similar structure holds for unbalanced networks, but the optimal number of applications depends on the factor of balancedness r. When the number of hospitals is half of the number of doctors, the market achieves its maximum size when each doctor applies to just one position. With more applications, hospitals become more congested; and the number of covered hospitals and therefore the number of total offers does not increase significantly. Therefore, the rejection probability of applications increases and with higher probability a doctor remains unmatched. In contrast, when the number of hospitals is more than the number of doctors, allowing more applications has a positive effect. With more applications –while the number of applications is still small– more hospitals receive at least one application; therefore, the number of offers increases considerably.

Also, note that in the figure, the social welfare of the balanced market is generally lower than the social welfare of markets with $r = 1/2, 2$. This is not surprising since by definition the social welfare is the fraction of matched individuals on the smaller side of the market:

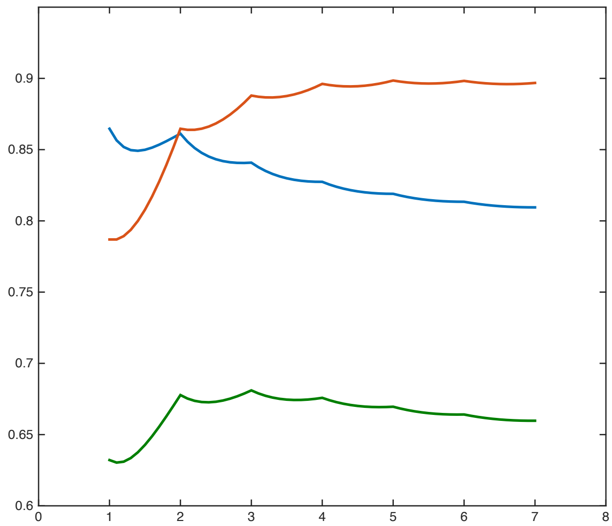

Figure 4 Social welfare of the matching with respect to the expected number of applications. The red curve refers to $r = 2$, blue curve to $r = 1/2$ and the green curve to $r = 1$, where r is the ratio of the number of hospitals to doctors.

doctors for $r \leq 1$, and hospitals for $r > 1$. For $r = 2$, with the same number of applications per doctor, more hospitals are covered that leads to more offers and a higher fraction of matched doctors. For $r = 1/2$, with the same number of applications, a higher fraction of hospitals is covered which leads to a higher fraction of matched hospitals.

5.2 Beyond Uniform Preferences

In this part, we go beyond the uniform and independent assumptions and use simulations to show that the properties of limited interviews exist more generally. The model studied in the previous sections assumes that doctors and hospitals preferences remain uniform and independent after the interview stage. We relax this assumption. We study a model where the two sides have no information prior to the matching procedure. However, after the interview stage, all participants refine their lists with information learned in the interviews. Therefore, preference lists may become correlated.

Without prior knowledge about the other side, the behaviors in the first stage is similar to what stated previously: there is a Bayes Nash Equilibrium such that doctors and hospitals pick the top of their lists. In the second stage, doctors and hospitals are not anymore identical in terms of popularity. We will be assuming that preferences among the interviews are independent samples from a distribution that reflects popularity. Immorlica and Mahdian [6] show that in a large market short preference lists sampled from a distribution, all participants are likely to prefer to be truthful. While this is not exactly the same as our model, analogous to this case, we will make the assumption in our simulation that participants do act truthfully.

To show how generally the properties hold, for simulations we study the other extreme in preferences: completely aligned preferences after interviews. We use a market of 10,000 doctors and hospitals. Similar to previous sections we start with uniform and independent preference in the first stage. For the second stage we simulate running the deferred acceptance algorithm when hospitals agree on a random preference over doctors. Figure 5 shows the size of matching as a function of number of applications when the number of interviews for hospitals is limited to two and four interviews.

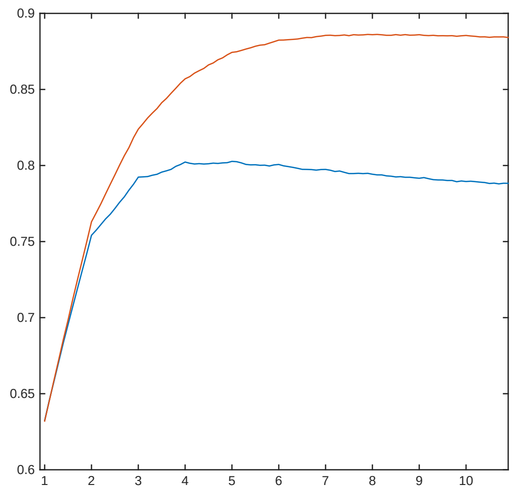

Figure 5 Size of matching with respect to expected number of application. In the first stage the preferences are uniform and independent. In the second stage hospitals preferences are aligned. The bottom curve belongs to two interviews and the top one to four interviews.

As observed in Figure 5, the properties of limited interviews are not restricted to uniform preferences. They hold even in the other extreme case where the preferences are completely correlated after interviews.

References

1 Nick Arnosti. Centralized clearinghouse design: A quantity-quality tradeoff. *Working Paper*, 2016.

2 Christopher Avery and Jonathan Levin. Early admissions at selective colleges. *The American Economic Review*, 100(5):2125–2156, 2010.

3 Hedyeh Beyhaghi, Daniela Saban, and Eva Tardos. Effect of selfish choices in deferred acceptance with short lists. *Match-Up*, 2017.

4 Joanna Drummond, Allan Borodin, and Kate Larson. Natural interviewing equilibria for stable matching. *working paper*, 2016.

5 D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.

6 Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 53–62 (electronic). ACM, New York, 2005.

7 Sangram Vilasrao Kadam. Interviewing in matching markets. *Working Paper*, 2015.

8 Fuhito Kojima and Parag A. Pathak. Incentives and stability in large two-sided matching markets. *American Economic Review*, 99(3):608–27, 2009. doi:10.1257/aer.99.3.608.

9 Robin S Lee and Michael Schwarz. Interviewing in two-sided matching markets. Working Paper 14922, National Bureau of Economic Research, April 2009. doi:10.3386/w14922.

10 Alvin E Roth and Marilda Sotomayor. Two-sided matching. *Handbook of game theory with economic applications*, 1:485–541, 1992.

A Missing Proofs

Discussion of Observation 2. The truthful equilibrium is due to the independence and symmetry and the private information model (the doctors and hospitals are unaware of the number of applications and preference lists of others). At all times, no hospital or doctor is more popular than the rest. Therefore from a doctor's point of view the probability of acceptance in all hospitals is the same and there is no benefit in not applying to a favorite hospital or be non-truthful about the preference order. The same holds for hospitals. ◀

Proof of Proposition 5. We show that if one of the doctors who previously had k applications, now has $k+1$ applications and the number of applications of other doctors remain the same, the size of the matching can only increase. By Observation 2, both doctors and hospitals are truthful, therefore we are comparing the size of matching as the result of deferred acceptance algorithm when doctor d has an extra application and all other doctors have the same number of applications. Since the deferred acceptance algorithm is oblivious to the order in which doctors proposes, we may hold out the last application of doctor d and find the outcome when the doctor d does not have this application in his/her list. The result of the deferred acceptance algorithm, without this application is the same as the case where doctor d had k applications. We show that this last application can only increase the number of matching. If doctor d is matched with one of his/her first k proposals, the last application does not change. If doctor d proposes to $k+1_{st}$ hospital, it can only increase the number of hospitals who have received any proposal. ◀

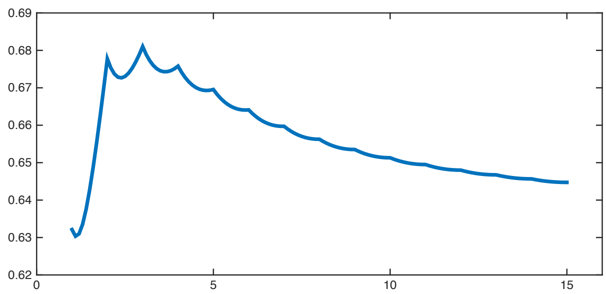

Figure 6 Size of matching with respect to expected number of applications, when hospitals make only one offer. (Extended version of Figure 2.)