Top of the Batch: Interviews and the Match

Federico Echenique[*] Ruy González[†] Alistair Wilson[‡] Leeat Yariv[§¶]

February 12, 2020

Abstract

Most doctors in the NRMP are matched to one of their most-preferred internship programs. Since various surveys indicate similarities across doctors' preferences, this suggests a puzzle. How can nearly everyone get a position in a highly-desirable program when positions in each program are scarce? We provide one possible explanation for this puzzle. We show that the patterns observed in the NRMP data may be an artifact of the interview process that precedes the match. Our analysis highlights the importance of interactions occurring outside of a matching clearinghouse for resulting outcomes, and casts doubts on analysis of clearinghouses that take reported preferences at face value.

Key words: NRMP, Deferred acceptance, Interviews, First-rank matches

JEL: C78, D47, J44

---

[*]Division of the Humanities and Social Sciences, Caltech.

[†]Caltech.

[‡]Department of Economics, University of Pittsburgh.

[§]Department of Economics, Princeton University, CEPR, and NBER.

[¶]Echenique gratefully acknowledges the support of NSF SES-1558757 and CNS-1518941. Yariv gratefully acknowledges the support of NSF grant SES-1629613.

1 Introduction

The National Resident Matching Program (NRMP; or "The Match") has matched millions of doctors to residency programs across the United States. In 2019 alone, there were over 38,000 active applicants matched to over 35,000 positions. At face value, match results reported by the NRMP suggest comforting news to newly-minted doctors: In 2019, 47.1% of applicants were matched to their first-ranked choice, and 72.5% were matched to one of their three top-ranked choices. The most-recent year's figures are by no means an aberration. In fact, the fraction of applicants matched to their first-ranked choice in 2019 is the lowest it has been over the past two decades. Our paper offers one explanation for this pattern through the interview process preceding the match. We show that interactions outside of a matching clearinghouse may be at least as important as the matching protocol prescribed by the clearinghouse itself.

Why is the prevalence of doctors' matches to their top-ranked residencies surprising? The algorithm governing the NRMP match is designed to implement a stable matching. If residents submit their true preferences to the NRMP clearinghouse, a predominance of matches to top-ranked employers implies that prospective residents' preferences are highly negatively correlated—as if pre-sorted. Indeed, if residents have very similar preferences (for example through a common ranking of residency positions), with limited positions at each program, only a handful can conceivably get their top-ranked choice. More specifically, consider a simple market with a hundred jobs and a hundred doctors. Common preferences on both sides (a fully assortative market) would yield an outcome where only *one percent* of doctors are matched to their top-ranked program. As we will show below, even a relatively small common component in doctors' preferences yields a low volume of matches to top-ranked hospitals.

Certainly, one explanation might be that prospective residents' preferences are diametrically opposed to one another and only a handful of candidates prefer each position. This explanation stands in the face of various surveys and preference estimations (see Rees-Jones, 2018; Agarwal, 2015). We propose another story. Prior to the centralized match, prospective residents interview with hospitals around the nation. The determination of who interviews with whom is decentralized in nature, resembling the process underlying many academic job markets. There are two important features of the interview stage. First, hospitals have interview capacities and consider only a small number of candidates out of the entire pool. Second, hospitals, and consequently newly-minted doctors, by and large only submit rankings to the centralized match *for those that they interviewed with*.[1]

For illustration, consider a setting where hospitals and prospective residents' preferences can be decomposed into two components: common and idiosyncratic. The common component for prospective residents may reflect public rankings of hospitals, quality of life in the hospitals' geographical area, etc. For hospitals, it may reflect doctors' academic performance and test scores (Agarwal, 2015). The idiosyncratic component reflects any personal or match-specific preferences—proximity to family for doctors, particular research interests for hospitals, and so on. At the interview stage, it is the common component of preferences that governs the interviews that take place.[2] Namely, each hospital has a fixed quota of interviews and the unique stable matching with respect to everyone's common preference com-

[1]The Results of the 2019 NRMP Applicant Survey (available at `http://www.nrmp.org/main-residency-match-data/`), reports on the median respondent by four types (matched/unmatched US senior/independent applicant) in each of 21 specialties (anesthesiology, pediatrics, surgery, etc.). Sixty-three of the 84 medians reported have the number of interviews attended and programs ranked perfectly coincide, where for 81 the two figures are ±1. For example, the median matched/unmatched US senior (independent applicant) respondent in emergency medicine applied to 46/56 programs (62/77). This process led to 14/6 interviews attended (10/3), and a rank-order list of length 14/6 (10/3).

A similar picture emerges in the NRMP's 2018 Program Director survey of the hospitals (here, across 23 specialties as it includes thoracic and vascular surgery). Looking at the per position numbers and averaging across specialties, we find 104.3 applications for each open position, leading to 14.5 interviews, and then 12.7 ranked candidates. Across all 23 specialties, the number of ranked candidates is always lower than the number of interviews.

[2]Returning to the 2018 Program Director survey, across the 1,233 respondents, the four most frequently cited factors in choosing applicants for interviews are primarily commonly observed test-scores and letters: (i) USMLE Step 1/COMLEX 1 test scores (94 percent of respondents); (ii) letters of recommendation (86 percent); (iii) medical student performance evaluations (81 percent); and (iv) USMLE Step 2/COMLEX Level 2 scores (80 percent). On average, program directors report that 48 percent of applicants are rejected out-of-hand based on standardized screening procedures, while 47 percent receive an in-depth review. In contrast, the most frequently cited factors for ranking applicants after interviews are more idiosyncratic: (i) interactions with faculty during interview and visit (96 percent); (ii) interpersonal skills (95 percent); (iii) interactions with house staff during interview and visit (91 percent); and (iv) feedback from current residents (78 percent).

ponent determines who interviews with whom. At the interview stage, the idiosyncratic components are realized, either through information discovery during the interviews, or through priorities that take greater precedence once the actual match is considered. At the centralized match stage, the "full" preferences, accounting for both common and idiosyncratic components, are used to determine everyone's submitted rankings. Importantly, only interview partners are ranked. We refer to these reported preferences as "interview-truncated."

The truncation induced by the interview process necessarily narrows agents' original preferences. Furthermore, capacity constraints at the interview stage imply that some prospective residents do not interview with programs they originally rank very highly. As a result, their matched programs' *stated* rank is higher than they really are.

The presence of a common component in prospective residents' preferences is crucial for this conclusion. In fact, we show that interview truncation will *not always* lead to a matching with inflated rankings. With sufficient disagreement in doctors' preferences, we show that the introduction of interviews may cause matched partners' stated rankings to go down, not up. Nonetheless, if all doctors have a common ranking for hospitals, we show that interviews necessarily lead to inflated rankings of prospective residents' matched programs.[3]

The assumption that all doctors are in perfect agreement over their ranking of hospitals is, of course, extreme. However, our main theoretical finding illustrates that in large markets, an *arbitrarily weak* common component is sufficient for interviews to generate the patterns observed in the data. We show that a small common-value component on both sides' preferences assures that a large fraction of doctors is very likely to be matched to programs that are highly ranked in the interview-truncated preferences, but these matched programs are far lower-ranked in the underlying preferences over all programs.

While our most-general results are asymptotic, they are also supported by an array of simulation exercises pertaining to realistic market sizes. These

[3]A related idea is pursued in Beyhaghi and Tardos (2018), who show that interviews may increase the size of a match, not decrease it. Rees-Jones (2018) suggests the scope of doctors misreporting.

simulations fit well the aggregate statistics reported by the NRMP, the fraction of unmatched agents, and the distribution of outcomes in the submitted rankings. [4]

Our results have important implications for the NRMP, and the matching literature more broadly. Doctors participating in the deferred-acceptance algorithm that underlies the match have incentives to truthfully report their preferences (Roth and Peranson, 1999). Traditionally, economists view the NRMP as an ideal case study in strategy-proof design. Our findings suggest that reported preferences in the NRMP should be interpreted with caution since they are filtered through an interview stage that pre-sorts the participants. In particular, reported high-rank matches should not be read literally. More importantly, any conclusions drawn about welfare using estimated preferences from the match itself are suspect. This message is particularly stark given that our paper ignores the strategic effects of interviews; we assume that agents' preferences are known, or reported truthfully.[5] Moving beyond the NRMP, our results highlight the important role interactions occurring outside of a matching clearinghouse can have on outcomes. Zooming in on behavior within a clearinghouse may yield misleading conclusions when done in isolation.

2 Setting Up the Puzzle

The introduction sets out the broad pieces to a puzzle presented by the NRMP match outcome data. In this section, we briefly outline and discuss why standard assumptions on preferences and the process are at odds with observed outcomes.

To begin, consider Figure 1. Here we illustrate the fraction of *matched* NRMP residents whose final outcome is their top-ranked hospital/program. Conditional on matching, approximately one half of all applicants get their top-ranked outcome across each year of the past two decades. This is true

[4] Ashlagi, Kanoria, and Leshno (2017) show that imbalanced markets lead to high match ranks for the short side of the market. In the NRMP, it is the hospitals who are on the short side, at least on the aggregate. Furthermore, Ashlagi et al. (2017) assume random independent preferences with no common-value component.

[5] See Beyhaghi, Saban, and Tardos (2017) for an analysis of the strategic implications of interviews.

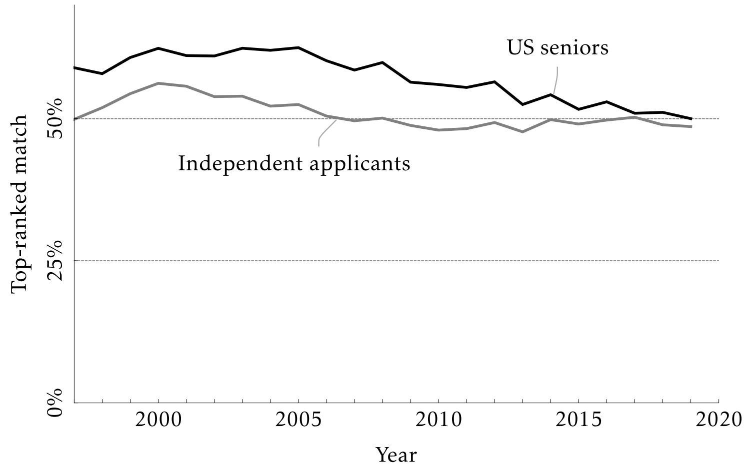

Figure 1: NRMP residents matched to top-ranked program (conditional on matching)

both for US seniors (residents graduating from a MD-granting medical school in the US) and independent applicants (those from DO-granting medical schools or based outside of the US).[6]

In order to get a sense for the types of preferences that could generate these patterns, consider a two-sided matching market with $N = 100$ participants on each side of the market. Suppose that for each doctor d and each hospital h we draw cardinal match utilities for the doctor, $u_d(h)$, and for the hospital, $u_h(d)$, as:

$$u_d(h) = \lambda^D \cdot c_h + (1 - \lambda^D) \cdot \eta_{d,h}$$

$$u_h(d) = \lambda^H \cdot c_d + (1 - \lambda^H) \cdot \eta_{h,d},$$

where c_h and c_d are each i.i.d. $\mathcal{N}(0,1)$ random variables representing the common utility components for each hospital and doctor, respectively, while the η terms are i.i.d. idiosyncratic terms, again distributed as $\mathcal{N}(0,1)$ random variables. Simulating the deferred-acceptance algorithm (DA) on the resultant ordinal rankings, and repeating the process across 100 simulations, we calculate the fraction of doctors matched to their top-ranked hospital.

[6] Across the 23 years reported, the fraction of unmatched US seniors is 5.2 percent, with a range of 4.2–6.2 percent and no clear trend. In contrast, for independent applicants, on average 45.4 percent go unmatched, where this figure shows a significant downward trend from 57.9 percent unmatched in 1997 to 31.9 percent in 2019.

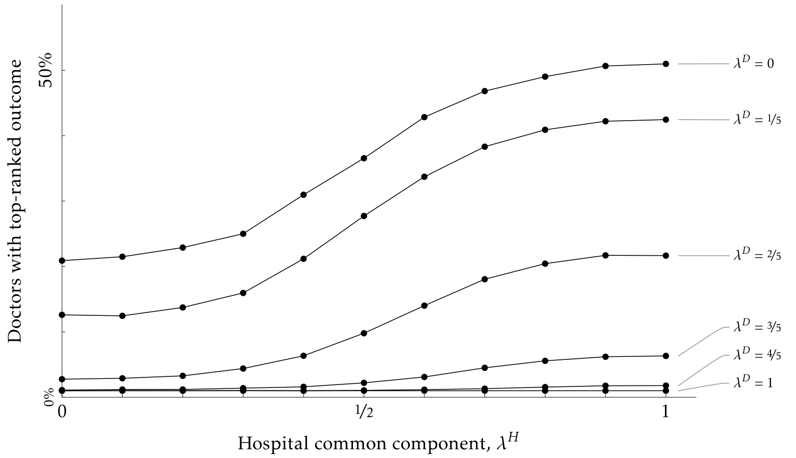

FIGURE 2: Simulated outcomes

Figure 2 indicates the fraction of doctors matched to their top-ranked program in the simulations, as we vary the weight hospitals place on the common utility component λ^H on the horizontal axis (simulations vary this weight from 0 to 1 in 0.1 increments). Each separate curve indicates a different value for λ^D, the weight on the doctors' common component (λ^D illustrated from 0 to 1 in 0.2 increments).

The simulation shows that we only achieve 50 percent matched to their top-ranked program when hospitals' preferences heavily weight the common component (λ^H close to one) and doctors' preferences heavily weight the idiosyncratic component (λ^D close to zero). While hospitals having a strong common component makes sense, and is consistent with NRMP survey figures that place a lot of weight on common components (test scores, recommendation letters, etc.) the requirement that doctors' preferences are almost completely idiosyncratic is in tension with what we know from the process.

First, examining the NRMP's 2019 post-match survey of residents, while potentially idiosyncratic components are highly cited as determining applications to and rankings of programs ("perceived goodness of fit" and "geographic location,"), common-value components appear at similar frequencies ("reputation of program," having an "academic medical center program," as well as quality of the residents, faculty, and educational cur-

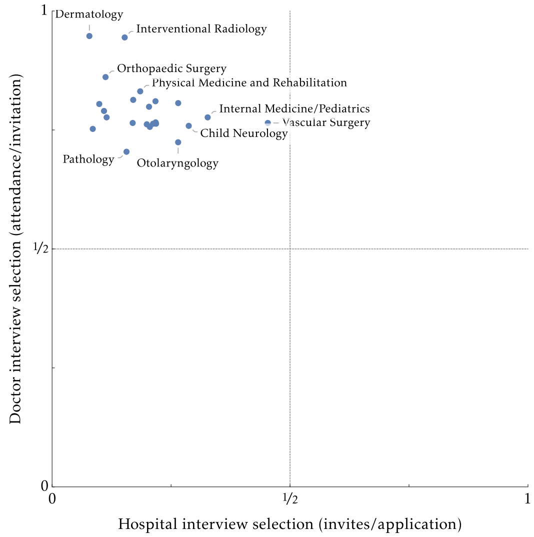

FIGURE 3: Selection of interviews

riculum).[7]

Second, in terms of applications, doctors apply to many more programs than they rank. Some elimination of programs may be due to idiosyncratic preferences on the part of doctors that get realized through the process. However, the main reduction in the number of ranked programs stems from hospitals *choosing* a subset of the applicants for interviews, rather than applicants choosing not to interview once given an invitation. In Figure 3 we use data reported in the NRMP's 2018 program director survey to illustrate where selection is occurring. On the horizontal axis we indicate the number of interviews invited by hospitals per application made, while on the vertical axis we indicate the interviews attended by applicants per invitation.

Averaging across the 23 medical specialties reported on, only one in every five applications is extended an interview invitation, ranging from a 7.8 percent invitation rate in dermatology to a 45.3 percent rate for vascular surgery. In contrast, candidates on average attend four out of every five interviews offered, ranging from a low of 70.4 percent attendance for pathology, to a high of 94.7 percent in dermatology. As the median matched US senior (averaging across specialties) reports applications to 46.6 programs but ranks just 12.9, extensive selection takes place. The survey indicates

[7] Geographic location also has an arguably large common-value component.

that the majority of this selection appears to occur through hospitals' preferences.[8]

Finally, the NRMP's reported match outcome data indicates that across the years very similar proportions of matched US seniors and independent applicants obtain their top-ranked outcome (see Figure 1). In general, US seniors tend to be the more-desirable matches from hospitals' point of view. This preference shows up in the overall match rates where US seniors entering the NRMP are matched at very high rates (∼95 percent) while independent applicants remain unmatched approximately 40 percent of the time.[9] However, if hospitals do have common preferences, and rank US seniors as more desirable than independent applicants, even if doctors preferences were fully idiosyncratic we would expect to see large differences in the top-rank matching rate across the two groups.

Turning back to our DA simulation with 100 doctors and 100 hospitals, if we set $\lambda^H = 1$ and $\lambda^D = 0$ we find that 50.6 percent of the simulated doctors match to their first-ranked program. However, if we break the sample up into above- and below-median doctors (using the hospital's common ranking c_d) the simulation predicts starker results. For the above-median doctors our simulations suggest 75.8 percent are matched to their top-ranked program, while this figure is only 25.6 percent for the below median group. In contrast, while the data does suggest substantial differences in the match rate for US seniors and independent applicants, conditional on matching there are no substantial difference in the rate of top-ranked outcomes.

In the next section, we outline our theoretical model for how the interviews affect outcomes. After outlining our main theoretical results, we return to a number of simulation exercises at the end of the paper.

[8]In contrast, the median unmatched US senior applies to *more* programs (an average of 55.8 across specialties) but ranks only 7 programs on average. The majority of the reduction again stems comes about through hospital selection over the interview invitations.

[9]This preference is also exhibited in the number of interviews extended to each type of applicant.

3 The Model

Our model is a simple variant of the standard two-sided matching model, as described in, for example, Roth and Sotomayor (1990). We introduce an initial interview stage, which is captured through a many-to-many stability notion. After interviews are conducted, agents' preferences are "pruned," or "truncated," so as to only rank interview partners. These preferences are then used in the standard deferred acceptance (DA) algorithm that governs the NRMP match.

3.1 Basic Definitions

A *market* is a triple (H, D, U) where:

- H is a finite set of *hospitals*;
- D is a finite set of *doctors*;
- $U = ((u_d)_{d\in D}, (u_h)_{h\in H})$ is a profile of *utility functions*, with $u_d : H \cup \{d\} \to \mathbf{R}$ and $u_h : D \cup \{h\} \to \mathbf{R}$ for each d and h.

A utility u_a induces an ordinal preference $\geq_a$ over the relevant set of alternatives. We shall assume throughout that $u_a(b) = u_a(b')$ implies that $b = b'$, so that the resulting preferences $\geq_a$ are *strict*. The *rank* of b in u_a is 1 + the number of b' with $u_a(b') > u_a(b)$. In particular, agent a's most preferred match partner is her ranked-1 partner. An agent b is *unacceptable* for a if $u_a(a) > u_a(b)$.

A *matching* is a function $\mu : H \cup D \to H \cup D$, with the properties that $\mu(h) \in D \cup \{h\}$, $\mu(d) \in H \cup \{d\}$, and $\mu(d) = h$ if and only if $\mu(h) = d$. A matching μ is *stable* for a market (H, D, U) if $u_a(\mu(a)) \geq u_a(a)$ for all $a \in D \cup H$, and there is no $(d, h) \in D \times H$ with $u_d(h) > u_d(\mu(d))$ and $u_h(d) > u_h(\mu(h))$.

A *many-to-many* matching is a function $\mu : H \cup D \to 2^{H\cup D}$ with the properties that $\mu(d) \subseteq H$, $\mu(h) \subseteq D$, and $h \in \mu(d)$ if and only if $d \in \mu(h)$. When an agent a is unassigned, we have that $\mu(a) = \emptyset$. Given the pair of positive integers (k, k'), we say that a many-to-many matching μ is *pairwise stable* for (k, k') if

- $|\mu(d)| \leq k$ and there is no $h \in \mu(d)$ with $u_d(h) < u_d(d)$

- $\left|\mu(h)\right| \leq k'$ and there is no $d \in \mu(h)$ with $u_h(d) < u_h(h)$
- There is no pair (h,d) such that $d \notin \mu(h)$ while h would want to add d, and d would like to add h, to its list of partners in μ. Formally, there is no (h,d) such that $d \notin \mu(h)$ and one of the following applies
    1. $u_d(h) > u_d(h')$ and $u_h(d) > u_h(d')$ for some $(h',d') \in \mu(d) \times \mu(h)$; or
    2. $u_d(h) > u_d(h')$ and $u_h(d) > u_h(h)$ for some $h' \in \mu(d)$ while $\left|\mu(h)\right| < k'$; or
    3. $u_d(h) > u_d(d)$ and $u_h(d) > u_h(d')$ for some $d' \in \mu(h)$; while $\left|\mu(d)\right| < k$;

3.2 Interview Schedules

In our model, doctors and hospitals first schedule interviews and then participate in the match.

An *interview schedule* is a many-to-many matching. Given a pair of integers (k,k'), a (k,k')-*constrained interview schedule* is a many-to-many matching μ with $\left|\mu(d)\right| \leq k$ and $\left|\mu(h)\right| \leq k'$ for all d and h. In words, each doctor is constrained to interview with at most k hospitals; and each hospital is constrained to interview at most k' doctors.

Given an interview schedule μ, the agents' *interview-truncated* preferences are determined by setting $u_a(b) < u_a(a)$ for all $b \notin \mu(a)$. That is, interview-truncated preferences rank all interviewed agents the same as in the original preferences, and sets all non-interviewed agents as unacceptable.

The timing in our model is then:

1. An interview schedule is determined as the doctor-optimal may-to-many (k,k')-stable matching.
2. Doctors and hospitals participate in the match. They report interview-truncated preferences, which are used as inputs in the doctor-proposing DA. The outcome is the doctor-optimal stable matching in the market with interview-truncated preferences.

A doctor-optimal interview schedule can be found algorithmically by the "T-algorithm" (see Blair (1988), Fleiner (2003) and Echenique and Oviedo

(2006)), but we assume that it results from a decentralized process of interview scheduling. One may imagine several reasons why an interview schedule might be unstable, but our focus is on the tension between a "pure" application of the DA, and one that is preceded by interviews. As such, assuming a stable outcome at the interview stage seems reasonable.[10]

The two steps, 1 and 2, in the above timeline are termed "TADA," as the T-algorithm is followed by Deferred Acceptance. We denote by μ^T the outcome of TADA.

For comparison with μ^T, the doctor-proposing DA algorithm starting from agents' original preferences, not the interview-truncated preferences, determines a matching μ^{DA}.

4 Preliminaries: Interviews Need not Increase Rankings

In general, interviews need not yield outcomes corresponding to higher-ranked partners in submitted preferences. In this section, we describe a particularly simple counter-example, which will shed light on the conditions that ensure interviews increase reported rankings of partners (and used in our main theoretical results).

Consider a matching market with three doctors, $\{d_1, d_2, d_3\}$, and three hospitals, $\{h_1, h_2, h_3\}$. Hospitals' preferences are common: they all prefer d_1 to d_2, d_2 to d_3, and d_3 to staying unmatched. Doctors' preferences are described in the following table, where higher-ranked hospitals for each doctor appear higher in the corresponding column.

[10] In one-to-one matching markets, experimental evidence suggests that decentralized interactions in matching markets yield stable outcomes at high rates, see Echenique and Yariv (2013). For more on the theory of many-to-many matchings, see also Sotomayor (1999) and Konishi and Ünver (2006). Finally, by focusing on the doctor-optimal stable matching we are rigging the model against our conclusion. A worse stable matching for doctors will result in an even larger difference between the reported match ranks in TADA, and the corresponding ranks in the DA.

d_1	d_2	d_3
h_1	h_2	h_1
h_3	h_3	h_2
h_2	h_1	h_3
d_1	d_2	d_3

If we run the regular deferred acceptance algorithm (DA), or any stable mechanism, it is easy to see that d_i matches to h_i, so the rank of d_i's match, h_i, is i.

Suppose that the number of interviews is constrained by $k = k' = 2$. Each doctor would then interview with their top two choices. That is, the resulting interviews are:

d_1 : h_1 and h_3;
d_2 : h_2 and h_3;
d_3 : h_1 and h_2.

To complete the outcome of TADA, consider the post-interview DA procedure. Given the interview truncated preferences, d_i matches with h_i for $i = 1, 2$. But h_3 and d_3 are unmatched. So the rank of d_3's match (being single) is four. Thus we have made one doctor match to a worse outcome than if there were no interviews at all.

In fact, with some disagreement between hospitals' preferences, we can construct examples that shift a doctor's match rank from one in the DA without interviews to four in TADA, with the interview stage. For example, take agents' preferences as given in the following two tables, where, again, higher-ranked partners appear higher in the corresponding column.

d_1	d_2	d_3
h_1	h_2	h_3
h_2	h_1	h_1
h_3	h_3	h_2
d_1	d_2	d_3

h_1	h_2	h_3
d_2	d_2	d_3
d_3	d_1	d_2
d_1	d_3	d_1
h_1	h_2	h_3

With these preferences, DA absent interviews again implies that d_i matches with h_i. However, with a capacity of two interviews for each participant ($k = k' = 2$) doctor d_1 does not get to interview with h_1. Consequently, d_1

and h_1 remain unmatched and d_1's rank falls from one to four.

Note that in these examples there is substantial disagreement between doctors' preferences. Indeed, there is not a single pairwise comparison of hospitals for which the doctors agree. As we show in the subsequent section, some agreement on the ranking of hospitals rules out such examples.

5 Theoretical Results

In the introduction and our discussion of the NRMP data, we emphasized the role of a common component in doctors' preferences. Our first theoretical result (Proposition 1) confirms that, indeed, if doctors agree on how hospitals are ranked, interviews improve observed match ranks in the succeeding clearinghouse.

The assumption of a common ranking seems strong, and unrealistic. Our second result (Proposition 2) shows that the main force behind our first result kicks in for a large market, as long as there is a common-value component, however small, in agents preferences. Finally, and before we provide a battery of quantitative findings through simulations, we illustrate convergence rates for the large-market result (Proposition 3).

5.1 Aligned Preferences

Proposition 1. *Suppose that $k = k'$ and that all doctors' preferences are the same. Then for any doctor d, the rank of $\mu^T(d)$ in her interview-truncated preference is always less than or equal to the rank of $\mu^{DA}(d)$ in her actual preference $\succeq_d$.*

The result assumes $k = k'$, mainly for expository reasons. As we shall see below, in our main result we allow for the two bounds to differ. In fact, the main result is driven by k, not k'.

5.2 Large Markets

Proposition 1 relies on perfectly aligned preferences. Here we show that even a small common-value component suffices to deliver our result, for a *large market*. We provide a probabilistic statement, when agents' utilities are randomly generated. An arbitrarily large fraction of doctors, with arbitrarily large probability, assign a higher rank to their match in the TADA assignment than their true ranking of their DA match.

The model of Section 3 has to be modified to account for the size of the market, and for randomly generated preferences. For each n, let (D_n, H_n, U_n) be a market where $D_n = \{d_1, \ldots, d_n\}$, $H_n = \{h_1, \ldots, h_n\}$ and each utility function is randomly drawn with a common value component, as well as an idiosyncratic value. Specifically, suppose that

$$\begin{aligned} u_d^n(h) &= \lambda^D c_h + (1-\lambda^D)\eta_{d,h} \\ u_h^n(d) &= \lambda^H c_d + (1-\lambda^H)\eta_{h,d} \end{aligned}$$

for all $d \in D_n$ and $h \in H_n$, where $\lambda^D, \lambda^H \in (0,1)$. Suppose, moreover, that $u_a^n(a) = 0$.

Here, c_h and c_d are common-value components of agents' utilities, while $\eta_{d,h}$ and $\eta_{h,d}$ are idiosyncratic. The existence of a common-value component is crucial to our results, but it does not need to be the dominant component of doctors utilities. All we need is that $\lambda^D, \lambda^H > 0$.

Suppose that each c_h, c_d, $\eta_{d,h}$ and $\eta_{h,d}$ is drawn from an absolutely continuous distribution with support $[0,1]$. (This can be relaxed to any continuous distribution with compact support and strictly positive density.)

Let μ_n^{DA} be the outcome of the doctor-proposing DA in the n-th market; such a matching is random and depends on the realized utilities, but we omit writing it explicitly as a function of (c, η).

The TADA procedure determines a matching μ_n^T by choosing a (k_n, k_n') constrained interview schedule $\hat{\mu}$ as the doctor-optimal many-to-many stable matching, as before, and running the doctor-proposing DA algorithm in the original matching.

We state the next result somewhat informally.

Proposition 2. *Suppose that* $\limsup k_n/n < 1$ *and let* $\varepsilon > 0$ *and* $\theta > 0$. *The probability of the following event converges to 1 as* $n \to \infty$: *For a fraction at least* $1-\theta$ *of doctors* $d \in D_n$, *the rank of* $\mu_n^T(d)$ *in d's interview-truncated preference is less than the rank of* $u_d(\mu_n^{DA}(d)) - \varepsilon$ *in d's actual preference* $\succeq_d$.

A formal statement of Proposition 2 can be found in Section 8.2, together with a proof. The idea behind this proposition is simple. Consider the DA. By Sangmok Lee's results (Lee, 2016), an arbitrarily large fraction of doctors get arbitrarily close to what their "assortative" match utility would be, based on their common-value utilities. If k_n is fixed, or does not grow faster than n, with high probability a large fraction of doctors will have at least k_n hospitals ranked above their common value utility (minus ε). Since the rank of a doctor's matching in TADA is at most k_n, the result follows.

Finally, we provide convergence rates for the large-market result in Proposition 2. The rates are modest, implying (poly-)logarithmic or polynomial growth in the relevant "approximation guarantees" θ and π. Our result on convergence rates is complementary to the simulation results in Section 6, which assume (arguably) realistic market sizes.

Proposition 3. *The statement in Proposition 2 holds for* $n = \Omega((\ln(1/\pi))^4)$ *as* $\pi \to 0$, *and* $n = \Omega((1/\theta)^4)$ *as* $\theta \to 0$.

6 Quantitative Results for the NRMP

Our theoretical results raise two important questions. The first regards the size of the market. Proposition 2 is asymptotic, so it is natural to ask whether interviews matter for smaller, more-realistic, market sizes. The second question regards unmatched agents. One might worry that interview-truncated preferences give rise to large numbers of unmatched market participants. If interviews play the role suggested by our theoretical results, we would want the number of unmatched agents to roughly match the numbers observed in the NRMP.

We address these questions using numerical simulations. In the past year, matched US seniors applied to 46.7 programs on average and got an average of 13.1 interviews.[11] In our simulations, we consider small markets, with 50 doctors and 50 positions. These are roughly within the range of the

[11] Unmatched US seniors applied to a somewhat higher number of programs, averaging at 55.8 and received an average of 7 interviews. Independent applicants applied to a higher number of programs and receive fewer interviews. Matched independent applicants, on average, applied to 69.4 positions and received 8.7 interviews, while unmatched independent applicants applied to an average of 70.1 programs and received 3.1 interviews.

Table 1: Hospital rank and probability of being unmatched in simulated markets ($N = 50$, $k = 5$)

Rank	First	Second	Third	Fourth	Fifth +	Unmatched
NRMP data (2019)						
US seniors	47.1	15.3	10.1	6.5	15.2	5.8
Simulations						
DA, No interviews	12.7	11.3	9.8	8.5	57.6	0
TADA with Reported	41.3	24.8	14.3	8.2	5	5.9
TADA with True	11.2	10.1	8.9	7.8	62	5.9

number of positions in any particular specialty, and corresponding to the figures above. We assume an interview capacity of 5.[12]

In our simulations, we use normally distributed common and idiosyncratic draws to derive the cardinal preferences $u_d(h)$ and $u_h(d)$, as discussed in Section 2, where we apply a weight of $\lambda^D = \lambda^H = \frac{2}{5}$ to the common component. Analysis of different weights generates very similar results (see also our Table 2 below).[13]

Table 1 summarizes the data from the NRMP and the outcomes from our simulations. The first row simply reproduces the NRMP data for US seniors. In the subsequent rows we report on the simulated outcomes under: (i) DA with no interviews; (ii) TADA with ranks calculated according to the interview truncated preference; and (iii) TADA with the ranks calculated with the true preference.

DA without interviews generates substantially fewer higher-ranked matches than observed in the data. In contrast, when looking at the reported preferences under TADA, the figures exhibit great resemblance to those of the NRMP. Our theoretical results speak to the comparison between the DA ranks and the reported TADA ranks in a large market. Here we see the same conclusion borne out in a small market simulation. Focusing on reported preferences is crucial. Indeed, while 41.3% of doctors are matched

[12] For the source of the data reported in this section, see the applicant (2019) and program director (2018) surveys and the annual Results and Data reports for the Main Residency Match on the NRMP website: http://www.nrmp.org/main-residency-match-data/

[13] All reported figures are average outcomes across the 50 doctors, further averaged across 2,500 simulations, each with a new set of draws of $\langle c_h, c_d, \eta_{d,h}, \eta_{d,h} \rangle$.

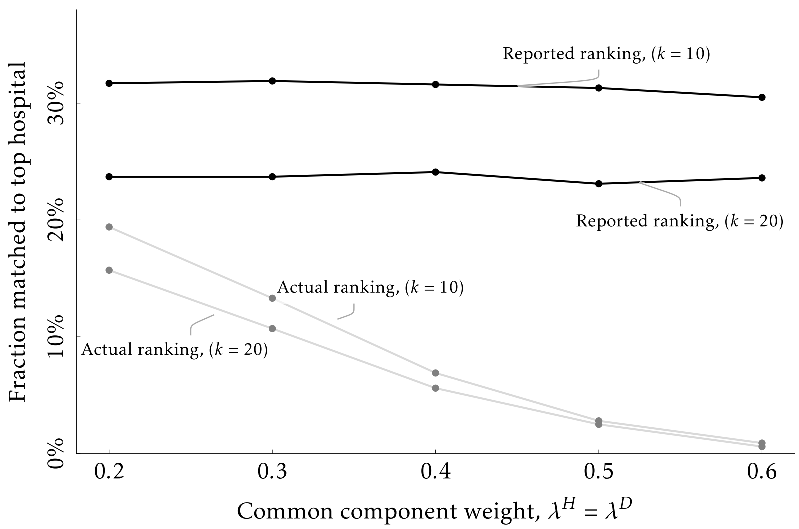

FIGURE 4: Fraction matched to first-ranked hospital (N=1,700)

to their top-ranked program under the reported preferences (compared to 47.1% in the data from the NRMP in 2019 Match), only 11.2% are matched to their top-ranked program in terms of their underlying preferences.

In terms of likelihood of matching, since our preference specification makes all doctors and hospitals acceptable to one another, DA produces a complete matching in the market. Interviews naturally come at a cost of unmatched agents. Nonetheless, our simulations suggest a small fraction of unmatched individuals, and comparable to what is observed in the NRMP data.

The simulations in Table 1 assume a market size of 50, which, despite our large-market theoretical results, might raise questions about the validity of the conclusions in larger markets. To this end, we ran simulations on a medium-sized market of $N = 1,700$ doctors and hospitals.[14] The results are in line with our theoretical results and simulations for small markets. For computational reasons, we were not able to implement TADA for

[14]While the 2019 NRMP had approximately 34,000 positions listed, much of the match breaks down into specialty sub-markets. For the listed specialties in the 2019 NRMP outcome report (Table 13), the median sub-market has 484 positions, ranging in size from 21 positions for Pediatrics/Medical Genetics (the NRMP only provide data for specialties with more than 20 total positions) to 8,512 for Internal Medicine. The 20th and 80th percentiles across the listed sub-markets have 37 and 1,740 positions listed, which are approximated by our chosen simulation sizes of 50 and 1,700.

TABLE 2: Difference between reported and true ranks: medium-market simulations ($N = 1700$, $k = 10$).

Rank difference	First	Second	Third	Fourth	Unmatched
$\lambda^D = \lambda^H = 0.2$	12.3	7.4	3.9	2.2	2.5
$\lambda^D = \lambda^H = 0.3$	18.6	11.6	6.5	3.7	2.5
$\lambda^D = \lambda^H = 0.4$	24.7	15.6	10.2	6.2	2.7
$\lambda^D = \lambda^H = 0.5$	28.5	19.1	12.2	7.3	2.7
$\lambda^D = \lambda^H = 0.6$	29.6	20.5	13.9	9.1	2.6

large markets. Instead, we had interviews scheduled exclusively along the common-value components of agents' preferences. Effectively, then, interviews follow a version of a serial dictatorship.[15]

Figure 4 and Table 2 indicate the results from these larger market simulations. First, in Figure 4 we indicate the fraction of top-ranked outcomes measure according to the actual rankings (the gray lines) and the interview-truncated rankings (the black lines). We report the results across differing weights placed on the common-value component on the horizontal axis (varying $\lambda^D = \lambda^H$ from 0.2 to 0.6 in one-tenth increments) and for the number of interview slots ($k = 10$ and $k = 20$). The figure indicates the growing difference the between the true and stated top-rank outcomes as the common-value component becomes more important.

In Table 2 we focus on the simulations with $k = 10$ interviews. The table reports, for each rank, the difference between the percentage of doctors who report that rank in the interview-truncated match and the percentage of doctors who fall into that rank according to their true simulated preference. So for the *First* column, the table simply reports the distance between the black and gray $k = 10$ lines in Figure 4. The *Second* through *Fourth* columns report the differences for the other top-ranked positions. Finally, the last column in Table 2 reports the percentage of unmatched doctors.[16]

[15]As a verification that this change in procedure does not distort the outcomes much, we compared this simplified version of TAGS to the full computation in smaller markets where computational impediments are not present. Results across the two algorithms are very similar when N is small.

[16]We have conducted larger simulations still with $N = 2,500$ and $k = 20$. The results obtained are similar to those reported for $N = 1,700$.

Overall, the conclusions from the larger markers are in line with our observations from the small-market simulations. Interviews substantially increase the likelihood of a match with a highly ranked partner according to the doctors' reported preferences. Furthermore, as can be seen in Figure 4, our results remove sensitivity to the precise weight put on the common component. Naturally, however, as less weight is put on the common component, the interview process has a weaker impact. Similarly, the larger the interview cap k the smaller the effects.

7 Conclusions

Much of the matching literature has focused on the centralized clearinghouse governing the match of newly-minted doctors and residency positions. In this paper, we illustrate the possibility that *decentralized* interactions preceding the match—namely, interviews—may dramatically impact ultimate outcomes.

Using NRMP data we suggest a puzzle for the standard model without interviews: *too many doctors match to their stated top-ranked hospitals*. Ostensibly, this might indicate the clearinghouse's algorithm is doing its job. If doctors' preferences over hospitals are idiosyncratic, such patterns are to be expected. However, idiosyncratic preferences are difficult to reconcile with surveys of NRMP participants, as well as common intuitions suggesting an important common-value component in preferences, due to public rankings of hospitals, shared quality of life considerations, etc. Our paper illustrates that the decentralized interview phase preceding the match may explain the puzzle.

For the NRMP, our results imply that empirical estimations based on preferences submitted to the centralized clearinghouse should be used with great caution. Stated rankings can reveal information on the relative preferences over ranked programs. However, the application and interview process may naturally truncate the stated preferences to a far smaller set of options. As such, a great deal of information on the complete preference is lost within the process.

More broadly, and beyond the application to the NRMP, our paper suggests that interactions outside of a matching clearinghouse can have dramatic effects on centralized outcomes. In particular, the analysis of matching clearinghouses cannot be reliably performed in isolation from other decentralized features of the market.

8 Proofs

8.1 Proof of Proposition 1

Let $\succeq$ be the common preference that doctors have over hospitals. Note that the DA is the same as serial dictatorship (SD) with the order dictated by hospital rank in $\succeq$.

Consider a doctor d, assigned to $h = \mu^{DA}(d)$ in the rth round of SD. The rank of h in d's preference is therefore r. If $k \leq r$ then we are done, as the rank of $\mu^T(d)$ in d's truncated preference can be at most k.

So suppose that $r < k$. Then two observations follow. Consider the interview stage and a hospital $h = \mu^{DA}(d)$ matched to d in stage $r' < k$ of the DA. When choosing whom to interview, h can choose any doctor, as all of them would have received strictly fewer than k interview requests at the time they get a request from h. So the hospital choosing at stage r' of the DA will interview the highest k doctors in her preference.

The second observation is that $\mu^{DA}(h) = \mu^T(h)$ for the hospital h choosing at round r.[17] This can be shown by induction: The statement is obviously true for the highest ranked hospital. Suppose that $\mu^{DA}(h) = \mu^T(h)$ for all the hospitals choosing at any stage $r' < r$. If h is the hospital of rank r then the set of doctors available to h in the DA stage of TADA is D, by our first observation, minus the choices of the hospitals with rank $r' < r$. By the inductive hypothesis the doctors chosen by the hospitals with rank $r' < r$ is the same as in the DA. So the set of available doctors to hospital h is the same in TADA as in the DA. Thus $\mu^{DA}(h) = \mu^T(h)$.

[17] Incidentally this may not happen for hospitals choosing at round $r' > k$. It is easy to come up with examples.

8.2 Proof of Proposition 2

We start by providing a formal statement of Proposition 2.

Proposition. *Let $k_n \geq 1$ be a sequence of positive integers. Let $\varepsilon, \theta, \pi \in (0,1)$. Suppose that $\limsup k_n/n < 1$. Then there is $N \in \mathbf{N}$ such that for all $n \geq N$ $P(E_n) \geq 1-\pi$, where E_n is the set of c_h, c_d, $\eta_{d,h}$ and $\eta_{h,d}$ such that in the resulting market (D_n, H_n, U_n), for a fraction at least $1-\theta$ of doctors d, the rank of $\mu_n^T(d)$ in her interview-truncated preference is less than the rank of $u_d^{-1}(\cdot)(u_d(\mu_n^{DA}(d))-\varepsilon)$ in her actual preference $\succeq_d$.*

In well-intentioned notational abuse, we write c_d for $\lambda^D c_d$, $\eta_{d,h}$ for $(1-\lambda^D)\eta_{d,h}$, etc. So that the utilities are the sum of the common and private value components: $u_d^n(h) = c_h + \eta_{d,h}$, and $u_h^n(d) = c_d + \eta_{h,d}$. The relevant probability distributions are rescaled correspondingly, but remain absolutely continuous, with support on a compact interval in $\mathbf{R}$. Without loss we assume that this interval is $[0,1]$.

Let $D = \cup_n D_n$, $H = \cup_n H_n$. Consider tuples (c,η), with $c = (c_a)_{a \in H \cup D}$ and

$$\eta = ((\eta_{a,b})_{(a,b)\in H\times D}, (\eta_{a,b})_{(a,b)\in D\times H}).$$

The tuples (c,d) are endowed with the product probability measure from the iid distributions described above. Any probabilistic statement refers to this probability space.

Let G denote the cdf of the distribution from which c_d is drawn and fix $\theta, \varepsilon, \pi > 0$.

To understand how the proof works, note that if agents match assortatively then a doctor d should be able to find a hospital h for which it has idiosyncratic utility close to 1, and this hospital should provide d with (approximately) the same utility $c_d + 1$ as it receives from matching with d. Think of $c_d + 1$ as d's "target utility."

To this end, let

$$A_n(\varepsilon,(c,\eta)) = \{d \in D_n : c_d + 1 - \varepsilon < u_d(\mu_n^{DA}(d)) < c_d + 1 + \varepsilon\}$$

be the set of doctors for which this is achieved (in the DA), up to an ε. We shall prove that, when n is large enough, with large probability a fraction at least $1-\theta/2$ doctors are in $A_n(\varepsilon,(c,\eta))$.

Next, we consider how many hospitals are ranked above a doctor's target utility c_d+1. Let

$$B(c_d,n)=\{\left|h\in H_n: c_h+\eta_{d,h}>c_d+1\right|\leq k_n\}.$$

be the event that fewer than k_n hospitals give d a utility greater than d's target utility. We denote by β_n the probability that a fraction of at least $\theta/2$ doctors have a "small" number (at most k_n) of hospitals above their target utility.

Now, we shall prove below that, for n large enough, $\beta_n<\pi/2$ and $P\left(\frac{1}{n}\left|A_n(\varepsilon,(c,\eta))\right|\geq 1-\theta/2\right)> 1-\pi/2$. Thus, the event that $B(c_d,n)$ is false for a fraction $\geq 1-\theta/2$ of doctors *and* the event $\left(\frac{1}{n}|A_n(\varepsilon,(c,\varepsilon))|\geq 1-\theta/2\right)$ holds, has probability $\geq(1-\pi/2)+(1-\pi/2)-1=1-\pi$. At the intersection of these events, it holds for a fraction $\geq(1-\theta/2)+(1-\theta/2)-1=1-\theta$ of $d\in D_n$ that $B(c_d,n)$ is false and $d\in A_n(\varepsilon)$. Hence, for a fraction $\geq 1-\theta$ of $d\in D_n$ there are more than k_n hospitals above their target utility, and they are within ε of their target utilities. This is the statement of the proposition and we are done.

To finish the proof we carry out the calculations needed in the above paragraph. To this end, let

$$l=\limsup_{n\to\infty}\frac{k_n}{n}$$

and recall that $l\in[0,1)$ by hypothesis. Choose c^* and $\delta>0$ such that $1-G(c^*)+\delta<\theta/4$ and $l<P(c_h+\eta_{d,h}>c^*+1)$. This is possible by absolute continuity of the distributions of c_h and $\eta_{d,h}$. Let $p(c^*)=P(c_h+\eta_{d,h}>c^*+1)$.

Note that, if $c_d\leq c^*$ then

$$\begin{aligned}P(B(c_d,n))&=P(\sum_{h\in H_n}\mathbf{1}_{c_h+\eta_{d,h}>c_d+1}\leq k_n)\\&\leq P\left(\frac{1}{n}\sum_{h\in H_n}\mathbf{1}_{c_h+\eta_{d,h}>c^*+1}\leq p(c^*)-(p(c^*)-\frac{k_n}{n})\right)\\&\leq\exp(-2(p(c^*)-\frac{k_n}{n})^2n)\end{aligned}\tag{1}$$

by Hoeffding's inequality (observe that, eventually, $p(c^*) - \frac{k_n}{n} > 0$).

Let

$$\begin{aligned}
\beta_n &= P\left(|\{d \in D_n : B(c_d, n)\}| > n\theta/2\right) \\
&\leq P\left(\underbrace{|\{d \in D_n : B(d,n) \text{ and } c_d \leq c^*\}|}_{Z_n} + \underbrace{|\{d \in D_n : c_d > c^*\}|}_{Y_n} > n\theta/2\right) \\
&\leq P(\frac{1}{n}Z_n + 1 - G(c^*) + \delta > \theta/2) + P(\frac{1}{n}Y_n > 1 - G(c^*) + \delta)
\end{aligned}$$

The first inequality follows by counting all d with $c_d > c^*$ as if $B(d,n)$ were true. So the random variable Y_n counts all $d \in D_n$ with $c_d > c^*$ as if they were in $B(c_d, n)$.

The second inequality is a truncation exercise, partitioning the probability space into two events. The first event is $\frac{1}{n}Y_n \leq 1 - G(c^*) + \delta$ and the second is $\frac{1}{n}Y_n > 1 - G(c^*) + \delta$. Under the second event, we automatically have that $\frac{1}{n}Z_n + \frac{1}{n}Y_n > \theta/2$ as $1 - G(c^*) + \delta > \theta/2$. Under the first event, the inequality is obtained by "raising" $\frac{1}{n}Y_n$ to $1 - G(c^*) + \delta$.

Applying Hoeffding's inequality again,

$$P(\frac{1}{n}Y_n > 1 - G(c^*) + \delta) \leq \exp(-2\delta^2 n). \tag{2}$$

Now,

$$\begin{aligned}
P(Z_n > n(\theta/2 - [1 - G(c^*) + \delta])) &\leq P(\cup_{d \in D_n} B(d,n)) | c_d = c^*) \\
&\leq \sum_{d \in D_n} P(B(d,n) | c_d = c^*) \\
&\leq n \exp(-2(p(c^*) - \frac{k_n}{n})^2 n), \qquad (3)
\end{aligned}$$

where the first inequality follows as $n(\theta/2 - (1 - G(c^*) + \delta)) \geq 1$, and the probability of $B(d,n)$ is maximized by setting $c_d = c^*$.

Choose n such that

$$n(\theta/2 - [1 - G(c^*) + \delta]) > 1, \tag{4}$$

$$\exp(-2\delta^2 n) < \pi/4, \tag{5}$$

$$n \exp(-2(p(c^*) - \frac{k_n}{n})^2 n) < \pi/4, \tag{6}$$

$$\text{and } P\left(\frac{1}{n}\left|A_n(\varepsilon, (c, \eta))\right| \geq 1 - \theta/2\right) > 1 - \pi/2. \tag{7}$$

Observe that (4) is possible as $\theta/2 - [1 - G(c^*) + \delta] > 0$. Inequality (6) requires that k is $O(n)$, which holds by hypothesis, and our choice of c^* to ensure that $p(c^*) - k_n/n > 0$ is eventually bounded below. Inequality (7) is possible by Theorem 1 of Lee (2016).

By (3) and (2) (5) and (6), we obtain that

$$\beta_n \leq n \exp(-2(p(c^*) - \frac{k}{n})^2 n) + \exp(-2\delta^2 n) < \pi/2 \tag{8}$$

Statements (7) and (8) provide the two bounds needed to complete the proof.

8.3 Proof of Proposition 3

Specifically, we show that there are constants N, K, K', K'' and K''' that do not depend on θ and π, such that for all

$$n \geq \max\{\bar{N}, \frac{\ln(\pi/4)}{K}, \frac{\ln(4/\pi)}{2\delta^2}, (\frac{\theta}{2} + K')^{-1}, (\frac{12}{\theta})^4, \left(\frac{\log(1 - \frac{\pi}{2})}{\log K''} + 3\right)^4 K''', \}$$

the statement in Proposition 2 holds.

The market size in the proof of Proposition 2 is determined from inequalities (4)-(7). These are the starting point of the proof. Using the bounds in Lee (2016), these mean that we need to choose n such that

$$-2[p(c^*) - \frac{k_n}{n}]^2 n \leq \ln(\frac{\pi}{4n}) \tag{9}$$

$$-2\delta^2 n < \ln(\frac{\pi}{4}), \tag{10}$$

$$\frac{1}{\frac{\theta}{2} - (1 - G(c^*) + \delta)} < n \tag{11}$$

$$\frac{2}{n}\left(\frac{1}{n^{1/4}} - 3\right)\sqrt{n}\log(n) + \frac{6}{n^{1/4}} > \frac{\theta}{2} \tag{12}$$

$$(1 - g_n)^{2n^{1/4}-4} \geq 1 - \frac{\pi}{2}, \tag{13}$$

where g_n is $o(e^{-\sqrt{n}\log n})$

For (9), choose N_0 and K_0 such that if $n \geq N_0$ then $(p(c^*) - k_n/n)^2 \leq K_0$. This is possible given the hypothesis that $\limsup k_n/n < 1$. Next, let $N_1 \geq N_0$ and K_1 be such that, for all $n \geq N_1$, $2K_0 n - \ln n \geq K_1 n$. Then we need that

$$K_2 n \geq \ln(\frac{4}{\pi}) \tag{14}$$

For (10) and (11), we have

$$n > \frac{\ln(4/\pi)}{2\delta^2} \tag{15}$$

$$n \geq \frac{1}{\frac{\theta}{2} - (1 - G(c^*)) - \delta} \tag{16}$$

For (12) we need that

$$\frac{2\sqrt{n}\log n}{n^{5/4}} - \frac{6\sqrt{n}\log n}{n} + \frac{6}{n^{1/4}} < \frac{\theta}{2}$$

$$\Longleftrightarrow \frac{2\log n}{n^{1/4}}\left(\frac{1}{\sqrt{n}} - \frac{3}{n^{1/4}}\right) + \frac{6}{n^{1/4}} < \frac{\theta}{2}$$

Let $N_2 \geq N_1$ be such that for all $n \geq N_2$, $\frac{1}{\sqrt{n}} - \frac{3}{n^{1/4}} \leq 0$. Then all we need is that $\frac{6}{n^{1/4}} < \frac{\theta}{2}$, or that

$$n \geq (\frac{12}{\theta})^4. \tag{17}$$

For (13), fix $N_3 \geq N_2$ and K_4 such that for all $n \geq N_3$ $1 - g_n \geq K_4$. So we need to obtain $\log(1 - \frac{\pi}{2})) \leq (2n^{1/4} - 3)\log K_3$. That is,

$$n \geq \left(\frac{\log(1 - \frac{\pi}{2})}{\log K_3} + 3 \right)^4 \frac{1}{16} \tag{18}$$

Set $\bar{N} = N_3$, $K = K_2$ $K' = (1 - G(c^*)) + \delta$, $K'' = K''' = 1/16$. Then the calculations above correspond to (14), (15), (16), (17), and (18).

References

Agarwal, Nikhil (2015), "An empirical model of the medical match." *The American Economic Review*, 105, 1939–78.

Ashlagi, Itai, Yash Kanoria, and Jacob D Leshno (2017), "Unbalanced random matching markets: The stark effect of competition." *Journal of Political Economy*, 125, 69–98.

Beyhaghi, Hedyeh, Daniela Saban, and Eva Tardos (2017), "Effect of selfish choices in deferred acceptance with short lists." MATCH-UP.

Beyhaghi, Hedyeh and Eva Tardos (2018), "Two-sided matching with limited interviews." Mimeo, Cornell University.

Blair, Charles (1988), "The lattice structure of the set of stable matchings with multiple partners." *Mathematics of operations research*, 13, 619–628.

Echenique, Federico and Jorge Oviedo (2006), "A theory of stability in many-to-many matching markets." *Theoretical Economics*, 1, 233–273.

Echenique, Federico and Leeat Yariv (2013), "An experimental study of decentralized matching." California Institute of Technology working paper.

Fleiner, Tamás (2003), "A fixed-point approach to stable matchings and some applications." *Mathematics of Operations Research*, 28, 103–126.

Konishi, Hideo and M Utku Ünver (2006), "Credible group stability in many-to-many matching problems." *Journal of Economic Theory*, 129, 57–80.

Lee, SangMok (2016), "Incentive compatibility of large centralized matching markets." *The Review of Economic Studies*, 84, 444–463.

Rees-Jones, Alex (2018), "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match." *Games and Economic Behavior*, 108, 317–330.

Roth, Alvin E and Elliott Peranson (1999), "The redesign of the matching market for american physicians: Some engineering aspects of economic design." *American economic review*, 89, 748–780.

Roth, Alvin E and Marilda Sotomayor (1990), *Two-sided Matching: A Study in Game-theoretic Modeling and Analysis*. Econometric Society Monographs volume 18, Cambridge University Press.

Sotomayor, Marilda (1999), "Three remarks on the many-to-many stable matching problem." *Mathematical social sciences*, 38, 55–70.