Risk and Conflation in Matching Markets

Aaron L. Bodoh-Creed
Haas School of Business, University of California, Berkeley*

Abstract

Real-world matching markets typically treat each participant as a unique "good," and when these markets grow large it becomes impossible for participants to truthfully describe their preferences over all potential partners. We show that forcing participants to use truncated preference lists, which is a common matching market design, exposes participants to a significant amount of risk even in very large markets (e.g., the New York City school match). Redesigning the set of preferences that can be declared to the mechanism by conflating traits limits this risk, but conflation can have ambiguous welfare consequences.

1 Introduction

Matching markets without transfers have been the subject of intensive mechanism design efforts with the goal of constructing effective, centralized matching mechanisms. Centralized matching mechanisms are used to place medical students into residencies in numerous countries, a process that involves accommodating the preferences of tens of thousands of students and thousands of hospitals (Roth [19]). Large cities such as Boston, Chicago, and New York City (NYC) employ centralized matching procedures to allocate students to public schools (Abdulkadiroğlu and Sönmez [2]).[1] The U.S.

*I would like to thank Ilya Segal, Paul Milgrom, Doug Bernheim, Fuhito Kojima, Jon Levin, C. Lanier Benkard, Matthew Jackson, Eduardo Azevedo, Lisa Seeman, Marcello Miccoli, Juuso Toikka, Alex Hirsch, Carlos Lever, Edward Augenblick, Isabelle Sin, John Lazarev, and Bumin Yenmez. I also thank the seminar participants at and faculty of the NYU Economics department and the Haas School of Business.

[1]Abdulkadiroğlu et al. [1] provides an excellent description of the differences between the NYC school match and canonical matching models.

military uses a centralized matching mechanism to assign cadets graduating from military academies to branches within the military (Sönmez and Switzer [24]).

In theoretical treatments matching markets are (usually) modeled as revelation mechanisms that allow participants to submit arbitrary rankings of potential partners to the mechanism. Since matching markets typically treat each participant as a unique kind of good, preferences become increasingly complex as these markets grow. As a result, a typical implementation of a large matching market requires participants to submit truncated rankings, which means these markets are only superficially revelation mechanisms. For example, the NYC school system matches 80,000 students with over 600 high school programs. Even if preference declarations of arbitrary length could be chosen, it is unlikely that parents would be able to thoughtfully rank such a wide array of high school programs.[2] To address this concern, the NYC school system allows parents to rank 12 programs (slightly less than 2% of the total), and the matching system assumes that the unranked programs are "unacceptable" for the student.

If one assumes (as we do) that it is unreasonable to expect students to describe their true preferences over hundreds of school programs accurately and completely, then at best the market can require students to provide an approximation of their true preferences. The truncated preference lists allowed in the NYC school match are one such approximation and are capable of precisely describing a small portion of the student's true preference relation.[3] Given the absence of compelling criteria for choosing between the many (non-truthful) methods for issuing approximate preferences, prior research has not focused on how market designers ought to build approximate preference declaration schemes.[4] As a result, truncated preference lists are a common feature of real-world matching mechanisms.

Our first contribution is to point out that truncated preference lists expose participants to a great deal of risk and can cause significant welfare losses from risk even in large markets. The welfare losses provide a strong motive to explore other methods for describing approximate student preferences. Our second contribution is to define

[2] We are agnostic regarding why it is difficult for agents to formulate complete preference rankings. Possible reasons include costs of information gathering as well as bounded rationality explanations such as rational inattention or costly information processing.

[3] The incentive issues that arise from trunctation are known (Haeringer and Klijn [10]).

[4] An important recent exception is Pathak and Sönmez [18], which defines a metric for market manipulability and suggests that less manipulable markets endure for longer. To the extent this notion of manipulability applies to our setting, Pathak and Sönmez [18] implies longer preference lists result in less manipulable markets.

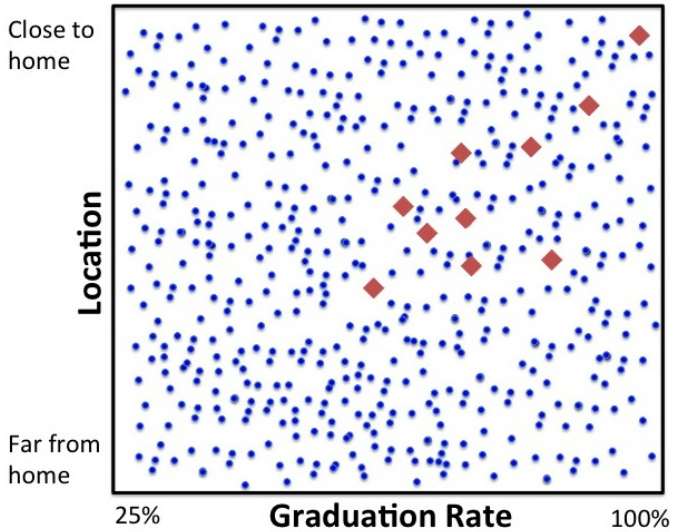

Figure 1: High Schools and Declared Ranking

a practical and intuitive way in which preferences can be approximated that we term *conflation*. Conflation can reduce the risk to which students are exposed and may have beneficial welfare effects even in complete information environments.

When a matching market requires that participants submit truncated preference rankings, the agents are exposed to the risk that they will be assigned to a partner when a superior (but unranked) partner would have accepted the agent. In extreme cases, it may be that an agent is unable to match with any partner ranked in his or her declared preference list when acceptable (but unranked) partners are available. We show that the risk faced by the agents fades as the market grows so long as the number of copies of each type of potential partner grows with the market, but the risk can vanish at a rate much slower than $O(\sqrt{N})$ if the set of types of partners also grows with the market.

We now graphically illustrate in the context of a school choice problem the joint effect of a rich set of partners and the requirement that preference orderings be truncated. Assume each parent's preferences over schools depend only on the schools' qualities and closeness to the parent's home. In figure 1 each dot represents a single school program, and we assume that the utility derived from each program is increasing along each dimension. Since the parents can only submit a truncated preference list, they must select a handful of these programs to rank. The school programs ranked are described as red diamonds in figure 1.

The match outcome is described in figure 2. The yellow diamond denotes the school to which the student is assigned by the mechanism, while the blue diamond indicates a superior, but unranked, school that would have accepted the student if only it had been ranked by the parents. In effect, the student has been matched with a "safety school."[5]

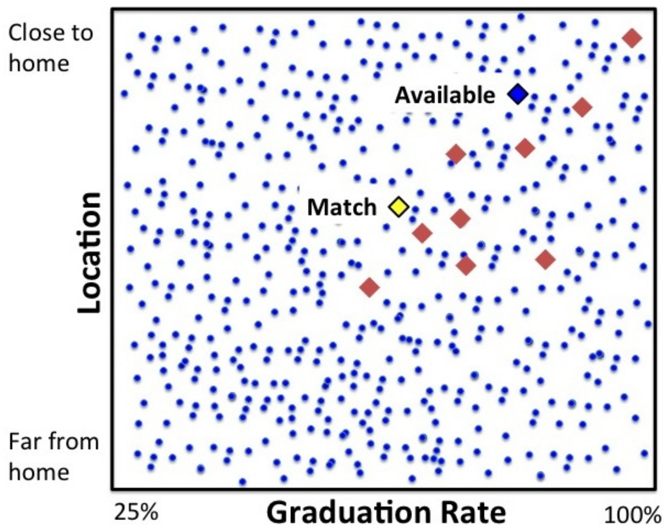

Figure 2: Match Outcome

Any matching market that requires participants to use truncated preference lists in an environment of incomplete information will result in risk of the sort illustrated above. As an alternative to mandating that preference lists be truncated, we propose that matching market designers ought to consider conflating participants that are close substitutes into categories and allowing the participants to include these categories in their preference ranking. For example, if a parent's top choice is a technology magnet program, then the parent could rank this kind of program rather than rank each such high school program separately. Conflating types has the advantage of allowing parents to submit a ranking of many more partners than would be possible under a typical implementation. The downside of conflation is that the ranking is imprecise in that the conflated partners are treated as perfect substitutes for one another.

The idea of conflation has already been implemented within the ERASMUS student exchange program. Students are allowed to conflate schools based on the country in which the schools are located. For example, a student could declare that a particular

[5] Also note that this outcome is unstable.

school in France is her top choice, her second choice is any school in Germany, and her third choice is another particular school in France.[6]

Throughout our work we refer to agents on one side of the market as colleges with the other side of the market referred to as students. Students wish to match with a single college, while colleges wish to match with at least one student. An agent's type is the combination of an observable trait and an unobservable preference relation over partner traits. A trait represents all of the information that is salient for describing the preferences of potential partners. In the school choice example above, the trait describes a school's quality and location (e.g., the neighborhood).

We use our model of matching markets to analyze the level of risk faced by the agents in the market. We define the risk facing a participant as the probability of a mechanism participant receiving an outcome significantly different than the expected outcome. Holding the set of traits fixed, we show that the level of risk each agent faces regarding his match is $O(\sqrt{N})$.

However, in most matching markets new types of agents must be introduced as the market grows. For example larger school systems require more school buildings, and these schools will be imperfect substitutes for each other if for no other reason than their differing locations. We show that even modest rates of growth of the number of traits as the set of participants increases can cause the risk faced by the agents to remain high even in large markets with 100,000 participants. This suggests that current mechanisms, which treat each participant as a unique "good," may subject the participants to high levels of uncertainty and that the associated welfare losses are of first order importance. The risk faced by the agents is particularly important in cases where the preference orders submitted to the mechanism only permit the ranking of a small subset of potential partners.

To make a comparison with data, Abdulkadiroğlu et al. [1], Table 1, provides simulation results based on preferences submitted to the New York City high school match that suggest roughly 7.5% of grade 8 applicants (5,500 out of over 73,000 applicants in total) are unmatched. Our analysis shows that this level of risk is compatible with a moderate rate of increase in the richness of the type-space as the number of agents increases.

On a technical note, our goal is to study the level of risk the participants are exposed to by analyzing the exact asymptotic distribution of match outcomes, whereas

[6] We thank a seminar audience member for this application.

prior works focus on proving asymptotic convergence (e.g., Kojima and Pathak [14], Lee [15]). To do this we introduce a new model of many-to-one matching markets and, in our main technical result, prove that our model is continuous in the distribution of agent preference declarations. Proving continuity in a traditional matching setting is extremely difficult - in fact, since matching mechanisms often have discrete outcomes, the appropriate notion of continuity is not obvious. The continuity of our model allows us to define the expected outcome of the matching mechanism using a limit model with a continuum of agents on each side. We can then apply central limit theorems to compute the asymptotic distribution of outcomes around the expected outcome. We use the limit distribution to describe how the richness of the type-space influences the risk faced by the agents.[7]

Finally, we formally define conflation of agent types by a market designer in the context of our model. Even if agents are defined by an exogenous set of traits, the mechanism designer can allow (or even require) agents to express preferences that conflate the exogenous traits into what we call designed traits. For example, the Stanford Graduate School of Business does not reveal any grade information to recruiters, which conflates students with different levels of academic performance. Kidney donations are based on compatibility between donors and recipients based on blood- and tissue-type. (Roth, Sönmez, and Unver [21]).[8] The New England Program for Kidney Exchange uses a coarser definition of compatibility than that used by the Eurotransplant Kidney Allocation System, which demonstrates that the conflation of degrees of compatibility may be important in practice.

By conflating traits, the designer allows participants to express longer preference lists and reduce the risk facing individual agents. We show that conflation has ambiguous welfare effects even in the large-market limit where risk vanishes. We conclude that an optimal market design requires a balance between the reduced risk facing the agents and the ambiguous welfare effects of the conflation of traits.

We close our paper by illustrating how our model can be extended to many-to-many matches and matching with contracts. We also trace out the limits on achieving goals for class composition (e.g., diverse student bodies) and discuss the marginal welfare costs of the stability and capacity constraints.

[7]It is our hope that by introducing a tractable, nonatomic matching model that our work may be used to address other issues in the matching literature.

[8]The tissue-type match is determined by a comparison of the human leukocyte antigens in the donor and recipient.

2 Related Literature

A number of papers study large matching markets with a focus on whether truthfulness by all agents can be supported in an approximate equilibrium, but none of these papers addresses the issue of the amount of risk facing the agents. Roth and Peranson [20] points out that the set of stable matches in the National Residency Matching Program is small, which implies approximate strategy-proofness. Immorlica and Mahdian [12] and Kojima and Pathak [14] analyze models of one-to-one and many-to-one matching markets (respectively) using the Gale-Shapley algorithm when the students have preference lists of fixed length independent of the market size. Lee [15] analyses truthfulness in environments without a restriction to short preference lists. Azevedo and Leshno [3] is the most closely related to this study in that a matching problem with a continuum of students is studied, although the questions addressed are entirely different.[9]

Levin and Milgrom [16] studies the conflation of goods in an auction environment. The focus of Levin and Milgrom [16] is the trade-off between the sharp delineation of goods that increases match value against the difficulty of monetizing these gains. We make the point that, in the context of matching markets, some conflation may be desirable as a way to thicken the market in order to reduce the risk faced by participants.

3 Model

We adopt the language of the college admissions problem and describe one side of the market as students and the other side of the market as colleges. We denote the set of types of students as $\mathcal{S}$ and the set of types of colleges as $\mathcal{C}$. A matching mechanism in the college admissions problem matches each student to one college, but each college may be matched with multiple students.

In traditional matching models, each individual student and college is defined as a unique "good" in the market, and agent preferences define an ordinal ranking of all possible partners. In our setting we instead define agent preferences over the set of possible partner traits, where a trait represents all of the payoff relevant information about a potential partner. Each type of student $s \in \mathcal{S}$ is characterized

[9]Echenique et al. [5] considers a matching market with a similar structure to ours to study identification.

by an observable trait ($t_s \in \mathcal{T}$) and an ordinal preference ranking over the traits of colleges to which they may be matched ($\succeq_s \in \mathcal{P}$).[10] For notational ease we assume that students and colleges draw their traits from the same set. $\succeq_s$ can denote some traits as unacceptable matches, and in the event that a trait t is unacceptable under $\succeq_s$ we use the notation $\varnothing \succeq_s t$. We assume throughout that $\mathcal{T}$ is finite. The set of student types is an element of $\mathcal{S} \subset \mathcal{T} \times \mathcal{P}$.

In an abuse of notation, for a student of type $s \in \mathcal{S}$ we use the notation $\succeq_s$ to denote a preference ordering over $\mathcal{C}$. The representations are equivalent - when used as a preference ordering over $\mathcal{C}$, $\succeq_s$ ranks colleges with preferred traits over those with less preferred traits, and colleges with the same trait are equally preferred.

We use the notation $\succeq_c^S$ to denote the preference ordering of college $c \in \mathcal{C}$ over sets of types of students comprising a class assigned to the college. A college $c \in \mathcal{C}$ has a capacity $q_c \in \{1, 2, ...\}$ that denotes the maximum number of students that may be assigned to the school. We refer to q_c as the number of seats in a college with each matched student occupying a single seat. We generically denote a set of student types as S, where we let S include multiple copies of each type of student since several students of the same type may be in a class. In addition, we assume that preferences over classes depend only on the traits of the students comprising the class. We use $\varnothing$ to denote a class with no students. We assume that $\succeq_c^S$ is **responsive** (Roth and Sotomayor [22]), where this entails

1. For any $\{s\}, \{s'\}$, and $S \succeq_c^S \varnothing$ with $|S| < q_c$, we have $\{s\} \cup S \succeq_c^S \{s'\} \cup S$ if and only if $\{s\} \succeq_c^S \{s'\}$

2. For any $\{s\}$ and $S \succeq_c^S \varnothing$ with $|S| < q_c$, we have $\{s\} \cup S \succeq_c^S S$ if and only if $\{s\} \succeq_c^S \varnothing$

3. $|S| > q_c$ implies $\varnothing \succeq_c^S S$

(1) and (2) imply that a college's preference ordering between students $s, s' \in \mathcal{S}$ is independent of the other students admitted to the college, while (3) is a feasibility constraint. We assume that $\succeq_c^S$ expresses indifference between singleton classes of students when the students have the same trait. We let $\succeq_c$ denote the preferences of

[10] We can nest the traditional models of matching without transfers (e.g., Roth and Sotomayor [22]) by assigning each individual a unique trait.

college c over the trait of singleton classes of students, which (since $\succeq_c^S$ is responsive) is equivalent to the full preference relation over classes of students.[11]

A college type $c \in \mathcal{C}$ is characterized by a trait ($t_c \in \mathcal{T}$), an ordinal preference ordering over the traits of students to which they may be matched ($\succeq_c \in \mathcal{P}$), and a capacity $q_c \in \{1, 2, ...\}$. The set of colleges is an element of $\mathcal{C} \subset \mathcal{T} \times \mathcal{P} \times \{1, 2, ...\}$. When a college finds a student type unacceptable, we use the notation $\varnothing \succeq_c s$. Let the set of mutually acceptable matches be denoted $\Gamma \subset \mathcal{C} \times \mathcal{S}$.

In the N-agent match, we assume that N students have types drawn independently from the distribution $\pi^{S,N}$ over $\mathcal{S}$. We denote the number of colleges in the N student match as M_N, and the types of the colleges are drawn independently from the distribution $\pi^{C,N}$ over $\mathcal{C}$. We assume throughout that these distributions are common knowledge. We refer to a generic realized empirical distribution of student (college) types in the N student game as $\pi_E^{S,N}$ ($\pi_E^{C,N}$). We denote the fraction of the population of type $s \in \mathcal{S}$ ($c \in \mathcal{C}$) in the N-agent game as $\pi_E^{S,N}(s)$ ($\pi_E^{C,N}(c)$).

We approximate the N-agent model using a limit game with a continuum of measure 0 agents of each type. In this limit game the set of students has measure 1 with a type distribution equal to $\pi^{S,N}$. The set of colleges has measure κ with types distributed as $\pi^{C,N}$ where $\frac{M_N}{N} \to \kappa$ as $N \to \infty$. Since the distribution of types is known exactly, this is a risk-free, complete-information economy. We later show that the limit model predictions represent the expected outcome of the finite model with many participants.

4 Matching in the Finite and the Continuum Case

We now provide a method for describing feasible and stable matches that allows us to use the same structure in matching mechanisms with a finite number of agents or a continuum of agents. We define a *match* as a function $x_S : \mathcal{C} \times \mathcal{S} \times \Delta(\mathcal{C}) \times \Delta(\mathcal{S}) \to [0, 1]$ where $\Delta(X)$ denotes the space of probability measures over X. $x_S(c, s; \pi^C, \pi^S)$ denotes the fraction of students of type s matched with colleges of type c given a distribution

[11] One might object that it is unreasonable to assume colleges seek classes filled with students having the college's most preferred trait for reasons such as a preference for a diverse class, which suggests that the assumption of responsive preferences (rather than our finite type structure) is objectionable. This becomes clear if we assume that every student has a unique trait (as in a conventional matching model), in which case classes are diverse by necessity, but each college would prefer an outcome where its favorite student could fill all of its seats.

of college and student types π^C and π^S respectively. Where confusion will not result, we suppress the arguments π^C and π^S.

Traditionally a matching is defined as a map between individuals on either side of the market (see Roth and Sotomayor [22]). Interpreted within the traditional modeling paradigm x_S represents an equivalence class of matchings, none of which are Pareto ranked. For example, suppose $C = \{c_1, c_2\}$ and we have for some $s \in S$ that $x_S(c_1, s) = x_S(c_2, s) = \frac{1}{3}$. In this case $\frac{1}{3}$ of the students of type s are admitted to c_1, $\frac{1}{3}$ of the students of type s are admitted to c_2, and $\frac{1}{3}$ of the students of type s are unmatched. Our notation does not specify the identity of the particular students assigned to each college.

We require that the following feasibility constraints are satisfied by our match given distributions of types π^C and π^S.

(1) For all $s \in \mathcal{S}$, $\sum_{c \in \mathcal{C}} x_S(c, s) \leq 1$

(2) For all $c \in \mathcal{C}$, $\sum_{s \in \mathcal{S}} x_S(c, s) * \pi^S(s) \leq q_c * \pi^C(c)$

Condition (1) insures that the number of seats occupied by students of type s does not exceed the population of these students. Condition (2) implies that the total number of students assigned to schools of type c, $\sum_{s \in S} x_S(c, s) * \pi^S(s)$, does not exceed the number of seats at these schools, $q_c * \pi^C(c)$.

We could equivalently describe a match in terms of the average number of seats in a college of type c filled by students of type s, which we denote as

$$x_C : \mathcal{C} \times \mathcal{S} \to \left[0, \max_{c \in \mathcal{C}} q_c\right]$$

The two formulations are related by the following flow equation

$$\pi^S(s) * x_S(c, s) = \pi^C(c) * x_C(c, s) \tag{4.1}$$

Where convenient we employ either x_S or x_C and assume equation 4.1 holds.

As noted for x_S, x_C refers to an equivalence class of match outcomes. Although x_S and x_C are tightly related, considering the equivalence classes in terms of class composition throws into sharp light the richness of the outcomes encapsulated by an equivalence class. Example 1 exemplifies this richness.

Example 1. *Suppose for some* $c \in \mathcal{C}$ *we have* $q_c = 2$, $\mathcal{S} = \{s_1, s_2\}$ *and the match outcome* $x_C(c, s_1) = \frac{2}{3}$ *and* $x_C(c, s_2) = \frac{2}{3}$. *There are many assignments supporting this match. For example, it could be that* $\frac{2}{3}$ *of colleges of type* c *are matched with* $\{s_1, s_2\}$ *pairs while* $\frac{1}{3}$ *of the colleges are empty. It is also possible that* $\frac{1}{3}$ *of the colleges are matched with* $\{s_1, s_1\}$ *pairs,* $\frac{1}{3}$ *are matched with* $\{s_2, s_2\}$ *pairs, and* $\frac{1}{3}$ *of the colleges are empty. Individual colleges prefer some of the outcomes over others, but the outcomes within an equivalence class are not Pareto ranked.*

4.1 Stability

A college-student pair $(c, s) \in \Gamma$ is a *blocking pair* if

$$\sum_{\{s' : s' \succeq_c s\}} x_C(c, s') < q_c \text{ and } \sum_{\{c' : c' \succeq_s c\}} x_S(c', s) < 1 \tag{4.2}$$

The first condition implies that some college of type c either has an empty seat or is matched to students to which a student of type s is strictly preferred. The second condition implies that some student of type s is either unmatched or matched to a university to which c is strictly preferred. A match x_S is *stable* if it does not admit blocking pairs.[12] Stability is a desirable property since it implies students and colleges have no mutual desire to recontract after the match. This can be interpreted either as a notion of fairness or as a form of renegotiation-proofness.

We now show that to determine if (c, s) is a blocking pair with respect to x_S, it suffices to consider the least preferred partners assigned to students of type s and colleges of type c. This implies our definition of matches as equivalence classes is without loss of generality since the stability of a match is independent of the element of the equivalence class implemented. Given a match x_S and a student $s \in \mathcal{S}$, denote the worst possible outcome under x_S for that student as $\underline{x}_S(s) \in \mathcal{C} \cup \{\varnothing\}$, where $\varnothing$ denotes that the agent is unmatched. If $\sum_{c \in \mathcal{C}} x_S(s, c) = 1$ we let

$$\underline{x}_S(s) = \max\{c : c \preceq_s c' \text{ for all } c' \text{ such that } x_s(c', s) > 0\} \tag{4.3}$$

where the maximum is taken with respect to $\succeq_s$. For $\sum_{c \in \mathcal{C}} x_S(c, s) < 1$, we define

[12] In matching models with indifferences, stronger notions of stability such as *strong stability* and *super stability* have been proposed. A stable match in our setting (in general) does not exist under these stronger notions of stability.

$\underline{x}_S(s) = \varnothing$.

To describe the worst partner of colleges of type c given a match x_S we use the function $\underline{x}_C(c) \in \mathcal{S} \cup \{\varnothing\}$. If $\sum_{s\in\mathcal{S}} x_C(c,s) = q_c$ we define $\underline{x}_C$ as

$$\underline{x}_C(c) = \max\{s : s \preceq_c s' \text{ for all } s' \text{ such that } x_C(c,s') > 0\} \tag{4.4}$$

where the maximum is taken with respect to $\succeq_c$. If $\sum_{s\in\mathcal{S}} x_C(c,s) < q_c$, we let $\underline{x}_C(c) = \varnothing$. We abuse our notation by letting (for example) $c \succeq_s \underline{x}_S(s)$ denote that c is preferred by s to all members of $\underline{x}_S(s)$.

Lemma 1. *$(c,s) \in \Gamma$ is a blocking pair (i.e., equation 4.2 holds) if and only if $c \succ_s \underline{x}_S(s)$ and $s \succ_c \underline{x}_C(c)$ hold.*

Suppose that the "if" condition holds, so there is some student of type s matched to a school to which c is strictly preferred (i.e., $c \succ_s \underline{x}_S(s)$) and some college of type c who admitted a student to which s is strictly preferred (i.e., $s \succ_c \underline{x}_C(c)$). In this case some student of type s and college of type c would wish to recontract ex post (i.e., equation 4.2 holds). The intuition for the reverse direction is similar.

We can describe the set of stable matches as the set of $x_S \in [0,1]^{|\mathcal{C}|*|\mathcal{S}|}$ that satisfy the following constraints given distributions of types π^C and π^S.

$$\text{For all } s \in \mathcal{S}, \ \sum_{c\in\mathcal{C}} x_S(c,s) \le 1 \tag{4.5}$$

$$\text{For all } c \in \mathcal{C}, \ \sum_{s\in\mathcal{S}} x_S(c,s) * \pi^S(s) \le q_c * \pi^C(c) \tag{4.6}$$

$$\text{For all } (c,s) \in \Gamma, \ \left(1 - \sum_{\{c'\in\mathcal{C}:c'\succeq_s c\}} x_S(c',s)\right) * \left(q_c - \sum_{\{s'\in\mathcal{S}:s'\succeq_c s\}} x_C(c,s')\right) \le 0 \tag{4.7}$$

$$\text{For all } (c,s) \notin \Gamma \text{ we have } x_S(c,s) = 0 \tag{4.8}$$

The first two conditions are feasibility constraints. Equation 4.5 insures that each student can be matched with at most one college. Equation 4.6 implies that each college has at most q_c enrolled students. Equation 4.7 encapsulates all of the restrictions imposed by stability as per equation 4.2. Equation 4.8 represents the individual rationality constraints.

Proposition 1. *Equations 4.5, 4.6, 4.7 and 4.8 are necessary and sufficient for a match x_S to be feasible and stable.*

The following lemma argues that there is at least one match satisfying our feasibility and stability criteria. The existence result is known for the case of markets with a finite number of agents. To prove the existence of a feasible and stable match in the limit game we develop a modified version of the deferred acceptance algorithm modified for the continuum setting.

Proposition 2. *There is at least one feasible and stable match.*

In the continuum case, the match outcomes described by equations 4.5 through 4.8 can be implemented exactly and deterministically. In the finite game, deterministic outcomes require integer constraints be satisfied as well. However, the match can also be implemented in the finite case as a randomization over deterministic, stable outcomes without adding any integer constraints.

To formalize a stochastic implementation in the finite case, we must delineate the individual students and colleges of each type. An *assignment* is a $N \times (M_N + 1)$ matrix defining the probability that each of the N students will be assigned to each of the M_N colleges (the first M_N columns) or remain unmatched (the final column). Consider a stable match x_S. For each (c, s) pair all cells of the matrix corresponding to students of type s and colleges of type c are assigned a value equal to $x_S(c, s)$ divided by the number of colleges of type c.[13] The resulting (stochastic) assignment, which we denote as $\mathbf{X}$, formalizes the interpretation of $x_S(c, s)$ as the probability that a student of type s is assigned to some college of type c.

Since each student must be, in the end, deterministically assigned to a college, the final match will be a *pure assignment*. In other words, the final assignment must contain only 0 or 1 values - a student is either matched to a particular college (1) or not matched with that college (0). We say that our stochastic match can be *implemented* if the stochastic match is equivalent to randomizing over pure matches. Our match can be implemented if we can find a set of pure assignments $\{\mathbf{X_i}\}_{i=1}^{A}$ and positive

[13] The choice to make the assignment of a student of type s to colleges of type c uniform was arbitrary. Many different choices could have been made, each of which would have altered the distribution of classes in each college of type c. The issue of class composition is discussed more in section 7.2.

numbers $\{\lambda_i\}_{i=1}^{A}$ such that $\sum_{i=1}^{A} \lambda_i = 1$ where

$$\mathbf{X} = \sum_{i=1}^{A} \lambda_i \mathbf{X_i}$$

Our proof uses theorem 1 of Budish et al. [4], which proves that the feasible set of a particular class of linear programs can be implemented using a mixture over pure assignments. Since our stability constraints are not linear (much less of the particular form studied in Budish et al. [4]), we need to argue that they can be re-written in a way that fits within the framework of Budish et al. [4].

Proposition 3. *Any (stochastic) match satisfying equations 4.5 through 4.8 can be implemented.*

Our main result, Theorem 1, proves that the set of stable outcomes is continuous with respect to changes in the distributions of declared types, π^C and π^S. Matching mechanisms that treat each agent as a unique good (e.g., Gale and Shapley [9]) are by necessity not continuous - when the distribution of type declarations changes, outcomes for an agent are either unchanged or changed radically as the agent is assigned a different partner. By allowing multiple agents to share the same type and by analyzing the market outcomes in the aggregate, we can define an appropriate notion of continuity.

The fact that our model admits a continuous selection is not obvious. Since the set of stable matches is defined by a set of weak inequalities, upper hemicontinuity of the stable set is straightforward. In order to prove lower hemicontinuity, we must show that given any distribution of types (π^C, π^S), any stable match in $x_S(\pi^C, \pi^S)$, and any perturbation of (π^C, π^S), we can find a small perturbation of the match that is stable under the perturbed distribution of types. We prove lower hemicontinuity constructively through the application of our modified deferred acceptance algorithm. Moreover, our constructive techniques allow us to show that a selection from the correspondence of stable matches is locally Lipschitz continuous.[14]

Theorem 1. *Let $G(\pi^C, \pi^S)$ denote the correspondence of feasible and stable matches given (π^C, π^S). $G(\pi^C, \pi^S)$ is continuous on the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$. Furthermore*

[14]The term *locally Lipschitz continuous* means the selection is continuous and within any open subset $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$ the selection is Lipschitz continuous. The Lipschitz constant may be different in each such subset, however.

there is a continuous selection $x(c, s; \pi^C, \pi^S)$ from G that is locally Lipschitz continuous on the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$.

Theorem 1 implies that we can define our matching function $x_S(c, s; \pi^C, \pi^S)$ as a selection that is continuous (with respect to the parameters π^C and π^S) from the correspondence of feasible and stable matches. While it is in general possible to have discontinuous selections, such selections could exhibit abrupt changes in outcome in the N-agent case if the mechanism has a discontinuity at (or near) π^C and π^S. The following assumption presumes that market designers avoid such mechanisms in practice since (among other reasons) they may impose unnecessary risk on the participants.

Assumption 1. *$x(c, s; \pi^C, \pi^S)$ is locally Lipschitz continuous in π^C and π^S.*

Assumption 1 insures that the limit model with a continuum of agents approximates the finite game with many participants. To see this, note that the distribution of types in a large market is with high probability close to the exact distribution from which the types are drawn. Assumption 1 implies that when the empirical and exact distributions are close, the outcome of the finite model is nearly the same as the outcome predicted by the limit model with a continuum of agents.

5 When the set of traits can grow

Since each type of agent asymptotically is realized in infinite numbers in the economy, the economy becomes thick in the sense that there are many copies of each type available as potential partners. When the economy becomes thick the match approaches the complete-information outcome predicted by our limit model, which implies that agents face no risk in the $N \to \infty$ limit.

However it is often more natural to assume that as $N \to \infty$ the number of possible partner traits increases. For example, as the number of students in a school district increases, the district will have to build additional facilities in new locations to accommodate the added students. As the set of schools expands, the new schools will be imperfect substitutes for the older schools if for no other reason than their different locations. We can think of new traits generated in this fashion as arising for *logistical* reasons.

The creation of new traits as $N \to \infty$ may also result from the desire to cater to participants with specialized preferences. For example a large school system could introduce schools to cater to narrow segments of the population (e.g., The Bronx High School of Science in NYC, the Newcomers Academy in Boston). These new school programs generate traits that arise for *innovative* reasons.

We argue in section 5.1 that modest rates of growth in the number of traits as N increases can stall the convergence of the outcome of the finite match to the complete-information outcome of the limit model, which implies that the participants may be exposed to significant uncertainty even when markets are very large. Further, in applications such as school matching, the market designer might desire a bound on the risk that holds uniformly across all students to insure that no student suffers an unusually adverse outcome. Such a uniform bound can be harder to satisfy as N increases over realistic ranges for even slow rates of increase in the richness of the type-space.

In sections 5.2 and 5.3 we develop our ideas through an example that illustrates how the level of risk changes as the set of traits grows. In section 5.4 we apply our analysis techniques to compute exact distributions of match outcomes for general economies with many participants.

5.1 Trait Growth and the Rate of Convergence

In order to illustrate the possibility of slow convergence, consider a matching market with N agents on each side of the market. Traits are drawn from the set $\mathcal{T}_N$ where $|\mathcal{T}_N| = T_N$. We assume that as $N \to \infty$ we have $T_N \to \infty$ and $\frac{N}{T_N} \to \infty$. In this case the law of large numbers holds, so for any $\delta, \rho > 0$ we can choose $N^* < \infty$ such that with probability $1 - \rho$ we have $\left\| \pi^{C,N} - \pi_E^{C,N} \right\| < \delta$ and $\left\| \pi^{S,N} - \pi_E^{S,N} \right\| < \delta$. The continuity of the match (assumption 1) implies that our limit model provides a close approximation of the likely outcomes of a finite market with many participants. However, the convergence to the limit outcome may be slow.

The slow convergence of equilibria is driven by two factors:

1. If $\frac{N}{T_N}$ grows slowly, N will have to be very large before the empirical probability distribution converges to the true distribution. Therefore the probability of fluctuations in $\pi_E^{C,N}$ and $\pi_E^{S,N}$ from $\pi^{C,N}$ and $\pi^{S,N}$ is significant for large N.

2. As T_N grows, the true probability of realizing any given type falls. Therefore, the

effect of mild fluctuations in the empirical distribution of types has an increasing *percentage* effect on the match as T_N increases.

Effect (1) is not surprising and follows from standard results on the law of large numbers.[15] Effect (2) is more subtle. Fix $\mathcal{T}_N$ and suppose that all traits are realized with equal probability in the population, so each trait t_i has a probability $\frac{1}{T_N}$ of being realized. A deviation of δ in the empirical frequency of t_i induces a percentage deviation of the empirical frequency equal to

$$\frac{(1/T_N)+\delta}{(1/T_N)} - 1 = \delta * T_N$$

It is these percentage changes that matter to agents seeking a partner with trait t_i since these capture the demand and supply imbalances for trait t_i. If the deviation of the frequency of t_i is $-\delta$, a fraction $\delta * T_N$ of agents seeking to match with a partner with trait t_i will be unable to do so. If T_N grows quickly, then would-be partners for agents with trait t_i face a significant amount of risk in the market even though δ is small with high probability.

5.2 Lower Bounds on Probability of Remaining Unmatched

To analyze how risk changes as the size of the market (and the number of traits) grows, we consider a sequence of markets indexed by N, the number of students. In the N student market agents on both sides of the markets are characterized by T_N traits indexed $\{1, .., T_N\}$. There are T_N^2 types of students indexed by $(i,j) \in \{1, .., T_N\}^2$, where a student of type (i,j) has trait $t_i \in \mathcal{T}_N$ and a preference relation of the form $t_j \succ_{(i,j)} \varnothing \succ_{(i,j)} t_k$ for all $t_k \neq t_j$. In other words, a student of type (i,j) has trait t_i and is only willing to match with partners with trait t_j. In the N agent game student types are drawn uniformly over the set $\{1, .., T_N\}^2$.

We assume that there is an equal numbers of students and colleges in each economy. The preferences of the colleges are drawn from the set $\mathcal{P}_N$ of strict orderings over the traits $\mathcal{T}_N$. We allow elements of $\mathcal{P}_N$ to denote some (but not all) partners as unacceptable. We consider a marriage market (so $q_c = 1$). We denote a generic college type as $(i, p_j) \in \{1, .., T_N\} \times \mathcal{P}_N$. We assume that the college trait-preference pairs are distributed uniformly over $\mathcal{T}_N \times \mathcal{P}_N$.

[15] It is straightforward to derive bounds on the probability that $\left\| \pi^{C,N} - \pi_E^{C,N} \right\| < \delta$.

Two alternative interpretations of our model can be offered. First, we could be dealing with an economy wherein students actually view a very small fraction of possible colleges as acceptable. Second, our market can be viewed as a model of a matching mechanism wherein the students have longer preference orderings over acceptable colleges, but are restricted by the mechanism to declaring preference lists of a fixed, finite length as $N \to \infty$.

Our strategy in this section is to analyze each economy in our sequence with the respective limit model approximation. For each such economy, our approximation of that economy implies that each student is matched with his most preferred college in the limit. For the N-agent economy, we use the term *risk* to refer to the probability of a large deviation in the outcome of the N-agent match from the approximation provided by the respective limit model. Any such large deviation leaves a significant number of students unmatched. By analyzing the risk in each economy in the sequence, we can calculate how the risk changes with N and T_N.

Our goal is to establish a lower bound on the probability that a student desiring a match with a college with trait j fails to match. The percentage of excess demand for colleges with trait t_j is at least δ if the following holds[16]

$$\sum_{i=1}^{T_N} \pi_E^{S,N}(i,j) - \sum_{p \in \mathcal{P}_N} \pi_E^{C,N}(j,p) \geq \delta \sum_{i=1}^{T_N} \pi_E^{S,N}(i,j) \tag{5.1}$$

If equation 5.1 holds, then at least a fraction δ of at least one type of student desiring a match with a college with trait t_j fails to match. Rewriting the left hand side of equation 5.1 we find

$$\begin{aligned} \sum_{i=1}^{T_N} \pi_E^{S,N}(i,j) - \sum_{p \in \mathcal{P}_N} \pi_E^{C,N}(j,p) &= \sum_{i=1}^{T_N} \left(\pi^{S,N}(i,j) + \eta_{i,j}^{S,N}\right) - \sum_{p \in \mathcal{P}_N} \left(\pi^{C,N}(j,p) + \eta_{i,p}^{C,N}\right) \\ &= \frac{1}{T_N} + \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \frac{1}{T_N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} \end{aligned}$$

where $\eta_{i,j}^{k,N} = \pi_E^{k,N}(i,j) - \pi^{k,N}(i,j)$ for $k \in \{C, S\}$. The right hand side of equation

[16]Equation 5.1 is a sufficient, but not necessary, condition for a large deviation, which is why we obtain a lower bound on the probability of such deviations.

5.1 can be written

$$\delta \sum_{i=1}^{T_N} \pi_E^{S,N}(i,j) = \delta * \left[\frac{1}{T_N} + \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} \right]$$

Combining these formula we have that equation 5.1 is equivalent to

$$(1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{i,p}^{C,N} \geq \frac{\delta}{T_N} \tag{5.2}$$

From the central limit theorem we know that[17]

$$\begin{aligned} \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} &\xrightarrow{d} N\left(0, \frac{1}{N}\frac{1}{T_N}\left[1 - \frac{1}{T_N^2}\right]\right) \\ \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} &\xrightarrow{d} N\left(0, \frac{1}{N}\frac{1}{T_N}\left[1 - \frac{1}{T_N * T_N!}\right]\right) \end{aligned} \tag{5.3}$$

If T_N is sufficiently large that $\left(1 - \frac{1}{T_N^2}\right) \simeq 1$, equations 5.2 and 5.3 yield

$$(1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} \xrightarrow{d} N\left(0, \frac{(1-\delta)^2 + 1}{N} * \frac{1}{T_N}\right)$$

Using this asymptotic approximation, we find that for large N

$$\Pr\left\{(1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} > \frac{\delta}{T_N}\right\} \simeq 1 - \Phi\left(\frac{\delta}{\sqrt{(1-\delta)^2+1}} * \sqrt{\frac{N}{T_N}}\right) \tag{5.4}$$

which is a lower bound on the probability that a fraction δ of colleges with trait t_j are overdemanded.

Equation 5.4 is plotted in figure 3 for different growth rates of T_N where $\delta = 5\%$. The growth rates reflected in each line are

T_N	Probability of $\delta = 5\%$ unmatched
$\Theta(1)$	$\Omega(N^{-1/2})$
$\Theta(N^{1/3})$	$\Omega(N^{-1/3})$
$\Theta(N^{1/2})$	$\Omega(N^{-1/4})$

[17]Note that the numbers of students and colleges of each type are binomial variables with parameters T_N^{-2} and $(T_N * T_N!)^{-1}$ respectively.

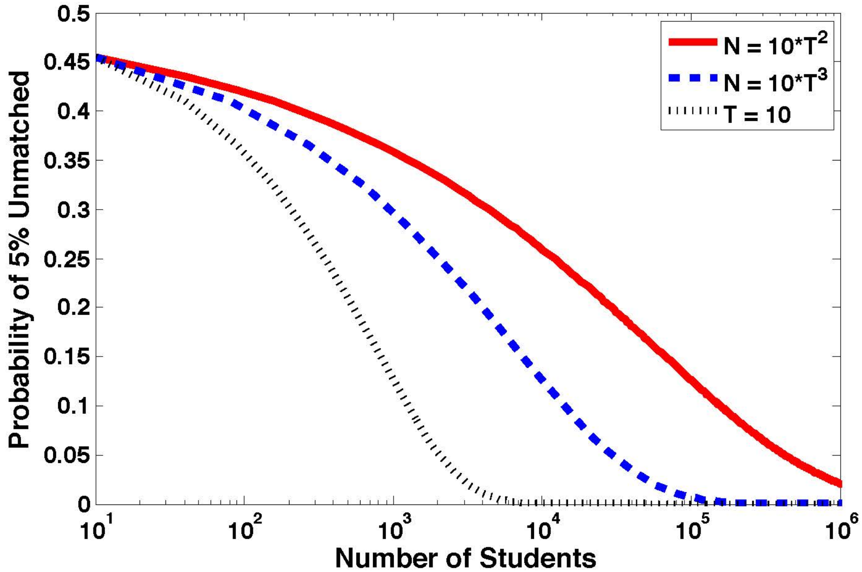

Figure 3: Bound on Large Deviations

where Θ denotes asymptotic equality and Ω denotes an asymptotic lower bound. We emphasize again that since equation 5.1 is a lower bound on the probability of a large deviation, our results provide an upper bound on the convergence rate.

Abdulkadiroğlu et al.[1], Table 1, provides simulations based on preferences submitted to the NYC high school match that suggest roughly 7.5% of grade 8 applicants (roughly 5,500 out of over 73,000 applicants in total) are unmatched. Our model of the $N = 10 * T_N^2$ case, which is the closest analog to the NYC school match, implies a lower bound of at least 14% of traits that have more than 5% of the students seeking that trait unmatched. Therefore our model predictions are (as expected) more conservative than found in the data.

5.3 Uniform Lower Bounds

In this section we seek uniform lower bounds on the probability that *some* type of student fails to match in the limit as $N \to \infty$. If the mechanism designer is a public agency (e.g., a school district), then a uniform bound provides an assurance that no group of participants is exposed to significant risk by the mechanism. These concerns can be acute when the agency has a mandate to serve disadvantaged groups (e.g., minority students, new immigrants) or may be subject to lawsuits or political pressure by any group that obtains an adverse match outcome.

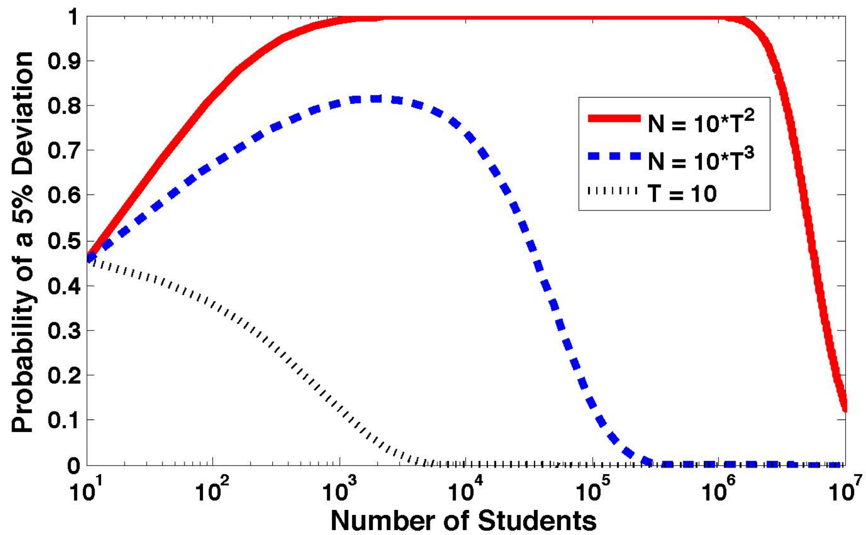

Figure 4: Uniform Bound on Large Deviations

We study the following uniform lower bound on the probability of deviations

$$\Pr\left\{\text{for all } j \in \{1,..,T_N\},\ (1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} \leq \frac{\delta}{T_N}\ \right\}$$

For large T_N, deviations in the measure of each type are only weakly correlated. By treating the demand and supply imbalance for colleges with each of the T_N traits as independent events we find

$$\Pr\left\{\text{for all } j \in \{1,..,T_N\},\ (1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} \leq \frac{\delta}{T_N}\ \right\} = \left[\Pr\left\{\ (1-\delta) * \sum_{i=1}^{T_N} \eta_{i,j}^{S,N} - \sum_{p \in \mathcal{P}_N} \eta_{j,p}^{C,N} < \frac{\delta}{T_N}\right\}\right]^{T_N} \tag{5.5}$$

Figure 4 plots the values of equation 5.5 using the asymptotic approximation described in equation 5.4. The non-monotone lower bound reflects a tension between the decreasing probability with which a single trait violates our bound as N increases (discussed in section 5.2) and the increasing difficulty of satisfying the uniformity condition as T_N increases with N. Note that in large markets with 100,000 students, when $T_N = O(\sqrt{N})$ the uniform lower bound is likely violated.

To obtain asymptotic convergence rates for equation 5.5, we use extreme value theory to provide a tail approximation of equation 5.5 as T_N grows large (see Embrechts et al. [6] propositions 3.2.3 and 3.3.2 as well as example 3.3.29 for details). The variance of the supply-demand imbalance of a single trait in the N student market, which we denote σ_N, is $O(\frac{T_N}{N})$. Given a set of traits $\mathcal{T}_N$, let the set of corresponding supply-demand imbalances be $\{\delta_i\}_{i=1}^{T_N}$ and let $\delta_N^{(1)} = \max\{\delta_i\}_{i=1}^{T_N}$. Standard results from extreme value theory imply

$$\sqrt{2 * \ln T_N} * \left[\frac{1}{\sigma_N}\delta_N^{(1)} - \sqrt{2 * \ln T_N} + \frac{\ln \ln T_N + \ln 4\pi}{2 * \sqrt{2 * \ln T_N}}\right] \xrightarrow{d} \Lambda \text{ as } N \to \infty$$

where Λ is the Gumbel distribution. Using the formula for the expectation of the Gumbel distribution, we find that the expectation of the largest deviation is $O\left(\sqrt{\frac{T_N}{N}} \frac{2\ln T_N - \ln\ln T_N}{\sqrt{2\ln T_N}}\right)$. Since $\delta^{(1)}$ is bounded, $\delta^{(1)} \to 0$ in probability at this rate as well. However, as Figure 2 suggests, for modest rates of increase in T_N, N must be on the order of $100,000$ for the Gumbel approximation to be useful.

5.4 General Asymptotic Bounds

The essential result from our example in sections 5.2 and 5.3 is that if the set of traits grows with the market, then the risk facing participants fades slowly as the market grows. In this section we provide *exact* asymptotic distributions of match outcomes for economies with arbitrary type distributions and preference list lengths. We use these exact distributions to show that the main qualitative result of sections 5.2 and 5.3, high risk even in large markets, continues to hold It is straightforward to use our results to find uniform lower bounds as in section 5.3.

To quantify the rate at which the asymptotic bounds are approached, continue to consider a marriage market for simplicity (so $q_c = 1$), which is without loss of generality since we could divide a college into q_c identical schools with a one student capacity.[18] Students are allowed to have arbitrary preferences over college traits, and the types of students and colleges are distributed according to measures $\pi^{S,N}$ and $\pi^{C,N}$ over $\{1, .., T_N\} \times \mathcal{P}_N$. We adopt the convention that in the N agent game the types are indexed $\mathcal{S} = \{s_1, .., s_K\}$ and $\mathcal{C} = \{c_1, .., c_L\}$.

For notational convenience, we let $p_S(s, t)$ and $p_C(c, t)$ represent the percentage

[18] This is not completely correct in a strategic setting since treating the seats of a college as separate agents eliminates the possibility that colleges may withhold capacity for strategic reasons.

change in the measure of a particular student or college type that match with a partner with trait t relative to the prediction of the limit model.

$$
\begin{aligned}
p_S(s,t) &= \frac{\sum_{\{c:t_c=t\}}\left[x_S(c,s;\pi^{C,N},\pi^{S,N}) - x_S(c,s;\pi_E^{C,N},\pi_E^{C,N})\right]}{\sum_{\{c:t_c=t\}} x_S(c,s;\pi^{C,N},\pi^{S,N})} \\
p_C(c,t) &= \frac{\sum_{\{s:t_s=t\}}\left[x_C(c,s;\pi^{C,N},\pi^{S,N}) - x_C(c,s;\pi_E^{C,N},\pi_E^{S,N})\right]}{\sum_{\{s:t_s=t\}} x_C(c,s;\pi^{C,N},\pi^{S,N})}
\end{aligned}
$$

for $\sum_{\{c:t_c=t\}} x_S(c,s;\pi^{C,N},\pi^{S,N}), \sum_{\{s:t_s=t\}} x_C(c,s;\pi^{C,N},\pi^{S,N}) > 0$. If $p_S(s,t) > 0$ this implies that a smaller percentage of students of type s match with colleges with trait t than predicted by the limit model. We can use our prior results on the continuity of the match, the central limit theorem, and the delta method to provide the following approximation of the probability of a large deviation of $x_S(c,s;\pi_E^{C,N},\pi_E^{C,N})$ from $x_S(c,s;\pi^{C,N},\pi^{S,N})$. In any such event, a significant fraction of students of type s are matched with colleges with an unexpected trait.

Proposition 4. *As $N \to \infty$ we we have*

$$
\begin{aligned}
\Pr\{p_S(s,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{c:t_c=t\}} x_S(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{c:t_c=t} \upsilon_S^2(c,s)\right]^{-0.5}\right) \\
\Pr\{p_C(c,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{s:t_s=t\}} x_C(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{\{s:t_s=t\}} \upsilon_C^2(c,s)\right]^{-0.5}\right)
\end{aligned}
$$

where we define $\sigma_S^2(s) = \pi^{S,N}(s) * (1 - \pi^{S,N}(s))$ and $\sigma_C^2(c) = \pi^{C,N}(c) * (1 - \pi^{C,N}(c))$, the diagonal matrices

$$
\Sigma_S = \begin{pmatrix} \sigma_S^2(s_1) & ... & 0 \\ ... & & ... \\ 0 & ... & \sigma_S^2(s_K) \end{pmatrix}, \Sigma_C = \begin{pmatrix} \sigma_C^2(c_1) & ... & 0 \\ ... & & ... \\ 0 & ... & \sigma_C^2(c_L) \end{pmatrix}
$$

and the variances

$$\begin{aligned}
\upsilon_S^2(c,s) &= \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right] + \\
&\qquad \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right] \\
\upsilon_C^2(c,s) &= \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right] + \\
&\qquad \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right]
\end{aligned}$$

We obtain qualitatively the same result as in section 5.2 - the risk vanishes slowly if $\sum_{\{c:t_c=t\}} \upsilon_S^2(c,s)$ or $\sum_{\{s:t_s=t\}} \upsilon_C^2(c,s)$ grows with T_N. While the terms are notationally cumbersome, they are easy to compute numerically from our continuum model. The distribution defined in Proposition 4 is exact (i.e., not just a lower bound on the probability), but the results of section 5.2 only required us to know the value of x_S at $(\pi^{C,N},\pi^{S,N})$.

Even without using the delta method, we can provide some insight into which agents are affected if a large deviation is observed. First note that when a trait is overdemanded, the deviation affects the least preferred partners of agents with that trait. Second, in each matched pair (c,s), lemma 1 implies that either $\underline{x}_C(c) \succeq_c s$ or $\underline{x}_S(s) \succeq_s c$. In other words, either c is s's least favorite partner or vice versa. Therefore, at least one agent in each matched pair is vulnerable to large deviations. This generalizes the insight of the example of section 5.2, in which every student is vulnerable to overdemand since each student is the single admitted student type (and hence the least preferred admitted student type) of their respective college.

For welfare analyses we care about fluctuations in the utility from the match rather than the match outcomes per se. However when students are forced to declare truncated preference lists, large deviations may leave students unmatched when there are acceptable, but unranked, partners available. Students that remain unmatched when acceptable partners are available clearly suffer a welfare loss. Furthermore, faced with the need to truncate her preference list, it is plausible that a student would include schools of varying degrees of desirability at the top of her preference list and end her list with one or more schools to which she is confident of being admitted. Being matched with such a "safety school" also entails a large loss of utility.[19]

[19] The use of "safety schools" makes it difficult to make welfare judgments even when most of the students are assigned to school programs at the close of a school match mechanism.

6 Conflation

The design of a matching market requires the selection of the traits that are used to describe the agents on each side of the market and define the preferences that can be expressed to the mechanism. When partners can (or must) be described in terms that do not distinguish between the underlying primitive traits, we refer to the traits as *conflated*. For example, students in public high schools in the United States are classified by a grade point average scheme that combines a variety of measures of performance within a course into a letter grade. Student performance within a course could be given by a complete ranking, which would be more precise and have the benefit of removing ambiguity as to the meaning of the grade (e.g., grade inflation). As noted earlier, kidney exchanges are based on compatibility between donor and recipient, but the threshold for a compatible match can be defined with varying degrees of precision (Roth, Sönmez, and Unver [21]).

Assume that agents have exogenous preferences over a set $\mathcal{T}$ of *primitive traits* that allows for the finest possible description of the market participants' preferences. The set of *designed traits* that the matching mechanism uses to describe the market participants, $\mathcal{T}^*$, is a partition of $\mathcal{T}$. The complete set of traits that can be ranked in a preference relation submitted to the matching mechanism is $\mathcal{T} \cup \mathcal{T}^*$. An agent's preferences over $\mathcal{T} \cup \mathcal{T}^*$ incorporate the agent's beliefs about the distributions and actions of the types comprising each designed trait.

Example 2. *In the NYC school match $\mathcal{T}$ includes one trait for each high-school program. $\mathcal{T}^*$ could include a designed trait t_M that includes all high school programs between 14^{th} and 60^{th} street with a graduation rate above 80%. By allowing preference rankings to include any elements from $\mathcal{T} \cup \mathcal{T}^*$, parents would have the ability to rank individual high school programs as well as a category corresponding to high quality schools in Midtown Manhattan.*

Designed traits can be put into practice in a variety of ways. The most straightforward, albeit restrictive, use of designed traits prevents students from ranking the primitive traits comprising a designed trait. This has the benefit of transparency in that the mechanism runs exactly like the deferred acceptance algorithm in use today.

The most flexible use of designed traits allows students to rank any combination of primitive and designed traits. The market could even allow the students to create custom designed traits in addition to any designed traits that are defined by the market designer. Allowing these flexible preference declarations requires some rule for how to resolve situations where a student has ranked a designed trait as well as some of the designed trait's constituent primitive traits. In such a situation we propose treating this as if the student had declared herself indifferent between the primitive traits comprising the designed trait unless she had otherwise ranked these primitive traits. The following example illustrates the procedure:

Example 3. *Let the primitive traits be* $\mathcal{T} = \{t_1, t_2, ..., t_{50}\}$ *and assume a single designed trait is available,* $t_A = \{t_1, t_2, t_3, t_4\}$*. Suppose a student declares the ranking* $t_1 \succ_s t_A \succ_s t_4$*. The market treats this as if the student had declared* $t_1 \succ_s t_2 \sim_s t_3 \succ_s t_4$

Allowing this flexible use of designed traits requires the mechanism to break these indifferences. Delving into the optimal technique for breaking these indifferences is beyond the scope of this work, but the interested reader is referred to Erdil and Ergin [7] and Abdulkadiroğlu, Pathak, and Roth [1].

We illustrate three first-order effects of conflation:

1. Allows for the ranking of more partners within a truncated preference list.

2. Can reduce the risk of being unmatched or matched with an inferior partner.

3. Welfare effects are, in general, ambiguous for both sides of the market.

If we assume that agents can rank as many conflated traits as they can primitive traits, then point (1) is immediate since each conflated trait corresponds to at least one primitive trait. In some cases it may in fact be easier to rank conflated traits than primitive traits. For example, suppose schools are conflated based on location and high school graduation rates. Given the conflated traits, it is easy for parents to express their quality-location preferences by ranking these conflated traits. If the traits were not conflated, then the parents would have to discover the quality and location information for each high school program and rank them accordingly.

Point (2) follow from the discussion in section 5. If we assume that a preference list declared by a student to the mechanism must be incomplete, then there is a possibility that the student will receive an outcome when a superior, but unlisted college, would have accepted the student. Section 5 shows that the probability of such an outcome shrinks very slowly if the number of traits grows at a moderate rate. From this result we can infer that since conflation allows agents to rank more partners, moderate use of conflation can greatly reduce the amount of uncertainty participants face due to risky mechanism outcomes.

Point (3) is subtle. Holding fixed the behavior of the other market participants, a student capable of learning his or her entire preference ranking and declaring the ranking to a student-optimal stable mechanism finds it incentive compatible to truthfully reveal this ranking. However, conflating traits can cause the participants to behave differently than they would in a mechanism without conflated traits, and these differences in behavior can cause the expected outcome of a match with conflation to differ significantly from the expected outcome of a match that uses only primitive traits.

A market designer would be reasonable to worry that conflation could provide a moderate reduction of risk, as suggested by point (2), at the cost of much worse expected outcomes for students (or colleges). Point (3) assesses the welfare differences between the expected outcomes of the conflated and unconflated economies as described by the respective complete-information limit economies. We find that the welfare effects are ambiguous and that, in fact, the welfare of up to half of the participants (e.g., all of the students) may be improved by conflation. [20]

6.1 Welfare

In this subsection we study the welfare effects of the behavioral changes caused by including designed traits in the match. We conduct our analysis by comparing the predictions of the complete-information limit models of the match with and without conflation. By examining the limit model we remove any welfare gains from conflation due to the reduction of risk. The welfare effects we identify are solely due to the change in expected match outcome that conflation may cause.

The welfare effects of conflating traits is complex. However, we can provide an upper bound on the possible welfare gains. The following proposition implies that if we consider each colleges' seat as a separate agent (or consider a marriage market), then at most $\frac{1}{2}$ of agents can have their welfare improved through conflation. Example

[20]The ambiguity of the welfare effects of conflation, especially when agents can only rank designed traits, may be surprising given results that show removing participants from one side of the market unambiguously lowers welfare for the other side of the match (Theorems 2.25 and 2.26 of Roth and Sotomayor [22]). These intuitions do not hold under conflation since by conflating two primitive traits we remove both of the primitive traits from the market and replace these traits with a designed trait that is a lottery over the underlying traits.

5 shows that this bound is tight. Proposition 5 holds for any set of designed traits and regardless of whether the use of designed traits is optional or required.

Proposition 5. *Suppose under conflation a student-college pair* (c, s) *are matched. Let* $\underline{x}_C(c)$ *and* $\underline{x}_S(s)$ *refer to the unconflated match. Then* $c \succ_s \underline{x}_S(s)$ *implies* $\underline{x}_C(c) \succeq_c s$ *and* $s \succ_c \underline{x}_C(c)$ *implies* $\underline{x}_S(s) \succeq_s c$.

Proof. Note that if $c \succ_s \underline{x}_S(s)$ implies $\underline{x}_C(c) \succeq_c s$ were violated, then (c, s) would be a blocking pair under the unconflated match, which contradicts the stability of the unconflated match. A similar argument applies for the second statement. □

We provide three examples of marriage markets to demonstrate the ambiguous welfare effects of conflation. The goal of our examples is to illustrate that conflation can improve the welfare of the vast majority of the participants on one side of the market and that there is no general relationship between the side of the market whose welfare is improved (or depressed) and the side of the market where traits are conflated. Where possible, we assume agents declare their preferences truthfully.

In each of our examples we assume that the agents are required to use the same set of designed traits. Requiring the use of designed traits could be a mandate from the market designer. However, we prefer the broader interpretation that the use of designed traits by the agents is a reflection of the limited capacity of the agents in our examples to describe their true preferences to the mechanism.

Example 4 demonstrates the intuitive idea that more finely described traits may yield higher welfare from match outcomes since the agents can more accurately describe their desired partners and obtain higher match values.

Example 4. *Let* $\mathcal{T} = \{t_1, t_2\}$ *and* $\mathcal{T}^* = \{t_A\}$*, so the matches are random given* $\mathcal{T}^*$*. Suppose that agents on each side find all partners acceptable, all traits are uniformly distributed in the population, and both* $t_1 \succ t_2$ *and* $t_2 \succ t_1$ *are equally likely and distributed independently of traits. In the nonatomic limit game under* $\mathcal{T}^*$*,* $\frac{1}{2}$ *of the agents are matched with a partner with their least-preferred trait. If the mechanism had instead used* $\mathcal{T}$*, then all agents would be matched with their ideal partner and a Pareto improvement is realized.*

Our next example demonstrates that the welfare of almost all students can be improved when two college traits are conflated. This contrasts with the usual result that adding colleges to the match is welfare improving for all students. For expositional ease, we denote the sets of student and college traits as $\mathcal{T}_S$ and $\mathcal{T}_C$.

Example 5. *Consider a market with primitive traits* $\mathcal{T}_S = \mathcal{T}_C = \{t_1, t_2, ..., t_K\}$. *The traits of the students are not conflated, so* $\mathcal{T}_S^* = \mathcal{T}$. *The traits of the colleges are conflated, so* $\mathcal{T}_C^* = \{t_A, t_3, ..., t_K\}$ *where* t_A *conflates traits* t_1 *and* t_2.

We assume that the following types of agents are in the economy

Student Type	*Probability*	*College Type*	*Probability*
$(t_1, t_1 \succ t_K \succ t_2 \succ ... \succ t_{K-1})$	*1/K*	$(t_1, t_1 \succ t_2 \succ ... \succ t_K, 1)$	*1/K*
$(t_2, t_1 \succ t_2 \succ ... \succ t_K)$	*1/K*	$(t_2, t_1 \succ t_2 \succ ... \succ t_K, 1)$	*1/K*
...	...	...	...
$(t_K, t_1 \succ t_2 \succ ... \succ t_K)$	*1/K*	$(t_K, t_1 \succ t_2 \succ ... \succ t_K, 1)$	*1/K*

Under the conflated match, assume the agents make the following declarations

Student Type	*Declaration*
$(t_1, t_1 \succ t_K \succ t_2 \succ ... \succ t_{K-1})$	$(t_1, t_K \succ t_A \succ t_3 \succ ... \succ t_{K-1})$
$(t_2, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_2, t_A \succ t_3 \succ ... \succ t_K)$
...	...
$(t_K, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_K, t_A \succ t_3 \succ ... \succ t_K)$

The student-optimal match under conflation is

Student Type	*Match Partner*
$(t_1, t_1 \succ t_K \succ t_2 \succ ... \succ t_{K-1})$	$(t_K, t_1 \succ t_2 \succ ... \succ t_K, 1)$
$(t_2, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_1, t_1 \succ t_2 \succ ... \succ t_K, 1)$
...	...
$(t_K, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_{K-1}, t_1 \succ t_2 \succ ... \succ t_K, 1)$

If the trait space had not been conflated, the student-optimal match would have been

Student Type	*Match Partner*
$(t_1, t_1 \succ t_K \succ t_2 \succ ... \succ t_{K-1})$	$(t_1, t_1 \succ t_2 \succ ... \succ t_K, 1)$
$(t_2, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_2, t_1 \succ t_2 \succ ... \succ t_K, 1)$
...	...
$(t_K, t_1 \succ t_2 \succ ... \succ t_K)$	$(t_K, t_1 \succ t_2 \succ ... \succ t_K, 1)$

All of the students but those of type $(t_1, t_1 \succ t_K \succ t_2 \succ ... \succ t_{K-1})$ *prefer the match* *under conflation to the match without conflation. For K large, this implies that essentially all of the students prefer that colleges with traits t_1 and t_2 be conflated.*

In our final example, we show that there is no simple relationship between the side of the match with conflated traits and the welfare effect of the conflation.

Example 6. *Consider a market where students have traits $\mathcal{T}_S = \{t_1, t_2\}$ and colleges have traits $\mathcal{T}_C = \{t_3, t_4, t_5\}$. The traits of the students are not conflated, so $\mathcal{T}_S^* = \{t_1, t_2\}$. The traits of the colleges are conflated, so $\mathcal{T}_C^* = \{t_A, t_5\}$ where t_A conflates traits t_3 and t_4. The types of agents are*

Student Type	*Probability*	*College Type*	*Probability*
$(t_1, t_3 \succ t_5 \succ t_4)$	*25%*	$(t_3, t_1 \succ t_2, 1)$	*25%*
$(t_1, t_4 \succ t_5 \succ t_3)$	*25%*	$(t_4, t_2 \succ t_1, 1)$	*25%*
$(t_2, t_3 \succ t_5 \succ t_4)$	*50%*	$(t_5, t_1 \succ t_2, 1)$	*50%*

Under the conflated match, assume the agents make the following declarations

Student Type	*Declaration*	*College Type*	*Declaration*
$(t_1, t_3 \succ t_5 \succ t_4)$	$(t_1, t_5 \succ t_A)$	$(t_3, t_1 \succ t_2, 1)$	$(t_A, t_1 \succ t_2, 1)$
$(t_1, t_4 \succ t_5 \succ t_3)$	$(t_1, t_5 \succ t_A)$	$(t_4, t_2 \succ t_1, 1)$	$(t_A, t_2 \succ t_1, 1)$
$(t_2, t_3 \succ t_5 \succ t_4)$	$(t_2, t_A \succ t_5)$	$(t_5, t_1 \succ t_2, 1)$	$(t_5, t_1 \succ t_2, 1)$

The student-optimal match given the type declarations under conflation is

Student Type	*Match Partner*
$(t_1, t_3 \succ t_4 \succ t_5)$	$(t_5, t_1 \succ t_2, 1)$
$(t_1, t_5 \succ t_4 \succ t_3)$	$(t_5, t_1 \succ t_2, 1)$
$(t_2, t_3 \succ t_4 \succ t_5)$	50% $(t_3, t_1 \succ t_2, 1)$, 50% $(t_4, t_2 \succ t_1, 1)$

If the trait space is not conflated and agents declare their preferences truthfully, the student-optimal match is instead

Student Type	*Match Partner*
$(t_1, t_3 \succ t_5 \succ t_4)$	$(t_3, t_1 \succ t_2, 1)$
$(t_1, t_4 \succ t_5 \succ t_3)$	$(t_4, t_2 \succ t_1, 1)$
$(t_2, t_3 \succ t_5 \succ t_4)$	$(t_5, t_1 \succ t_2, 1)$

In this case we have that colleges of type $(t_4, t_2 \succ t_1)$ *and* $(t_5, t_1 \succ t_2)$ *strictly prefer the match under conflation, while students* $(t_1, t_3 \succ t_4 \succ t_5)$ *and* $(t_1, t_5 \succ t_4 \succ t_3)$ *and colleges of type* $(t_3, t_1 \succ t_2)$ *prefer the match without conflation. The welfare of student* $(t_2, t_3 \succ t_5 \succ t_4)$ *depends on the specification of the utility function.*

7 Other Applications and Extensions

In this section we highlight a few fashions in which our framework can be extended and applied. First, our model and the associated results can be extended to many-to-many matching if agents on both sides of the match have responsive preferences. Second, because our match only describes the expected number of seats in each type of college occupied by each type of student, the market designer has some room to accommodate goals related to the composition of the classes (e.g., diversity) . Third, we can use our model to describe the marginal value of relaxing the stability constraints or adding capacity to the schools. Finally, we can extend our model to the case of matching with contracts.

7.1 Many-to-Many Matches

Our analysis has focused on many-to-one matching markets as this is usually the case of interest for applied matching market design (e.g., school choice). However, our framework extends gracefully to cover the case of many-to-many matching markets if we assume that student and college preferences are responsive and we use pairwise stability as our equilibrium concept. This allows us to handle applications such as the matching of workers within a firm to one or more tasks in the firm.

In the many-to-many case, we assume that students of type s can be matched with at most q_s colleges. In the many-to-many matching framework, we adjust the capacity constraint for the students by replacing equation 4.5 with

$$\text{For all } s \in \mathcal{S}, \ \sum_{c \in \mathcal{C}} x_S(c, s) \leq q_s$$

and replace our stability condition (equation 4.7) with

$$\text{For all } (c,s) \in \Gamma, \quad \left(q_s - \sum_{\{c' \in \mathcal{C} : c' \succeq_s c\}} x_S(c', s)\right) * \left(q_c - \sum_{\{s' \in \mathcal{S} : s' \succeq_c s\}} x_C(c, s')\right) \leq 0$$

Conveniently, our results on existence of a pairwise stable match (proposition 2), the implementability of the match as a mixture of pure assignments (proposition 3), and the continuity of the correspondence of stable matches (theorem 1) all continue to hold. We leave applying this framework to future work.

7.2 Class Composition

We now consider whether the flexibility of our framework provides leverage to achieve goals that depend on the structure of classes. As noted before, there are a variety of class compositions that implement the same stable match. This means that we have some flexibility to satisfy objectives that depend on the allocation of students into classes without interfering with the implementability or stability of the match.

We focus on two class composition desiderata: balanced class sizes and diversity. For under-demanded school types (i.e., those with empty seats), it is entirely possible for some schools to be filled to capacity while other schools of the same type are empty as demonstrated in example 1. To address this the market designer may wish to move students between schools to balance the class sizes. Note that we can move students between colleges of the same type without interfering with the stability or the feasibility of the match. If balanced class sizes are desirable, then we can move students in this fashion until there is less than a one student difference between the largest and smallest classes assigned to colleges of a particular type.

We now provide a technique for accommodating a designer's preferences for diverse classes composed of a variety of student types. To formalize a designer's preference for diverse classes, suppose the market designer wants each school's class to be as close to a canonical distribution of types as possible. Denote the market designer's ideal class composition for colleges of type c as $x_S^{Ideal}(c, \circ)$. We do not require $x_S^{Ideal}(c, s)$ to be either feasible or stable given the distributions of declared types (π^C, π^S). Finally, we assume that the market designer has a social welfare function over the matches (given the ideal match) that we denote as $W(x_S(c,s), x_S^{Ideal}(c,s))$. One obvious candidate for the social welfare function is

$$W(x_S(c,s), x_S^{Ideal}(c,s)) = \sum_{(c,s)\in\mathcal{C}\times\mathcal{S}} \lambda_{(c,s)} \left\| x_S(c,s) - x_S^{Ideal}(c,s) \right\|$$

where $\lambda_{(c,s)} \geq 0$ is a weighting factor.

The market designer solves the class composition optimization problem in two steps. First, the designer maximizes the social welfare function subject to the feasibility and stability constraints

$$\begin{aligned}
x_S^* &= \underset{x_S\in[0,1]^{\mathcal{C}\times\mathcal{S}}}{\arg\max} \; W(x_S(c,s), x_S^{Ideal}(c,s)) \text{ such that} \\
\text{For all } s &\in \mathcal{S}, \; \sum_{c\in\mathcal{C}} x_S(c,s) \leq 1 \\
\text{For all } c &\in \mathcal{C}, \; \sum_{s\in\mathcal{S}} x_S(c,s) * \pi^S(s) \leq q_c * \pi^C(c) \\
\text{For all } (c,s) &\in \Gamma, \; \left(1 - \sum_{\{c'\in\mathcal{C}:c'\succeq_s c\}} x_S(c',s)\right) * \\
&\qquad \left(q_c - \sum_{\{s'\in\mathcal{S}:s'\succeq_c s\}} x_C(c,s')\right) \leq 0 \\
\text{For all } (c,s) &\notin \Gamma \text{ we have } x_S(c,s) = 0
\end{aligned}$$

Proposition 3 implies $x_S^*(c,s)$ can be implemented as a mixture over pure assignments. Once a particular pure assignment has been realized, the second step in solving the class composition problem is to move students between colleges of the same type until the composition of classes at each college of type c is as close to $x_S^*(c,\circ)$ as possible. By moving students in this fashion, we can insure that realized classes are within one seat of the description provided by $x_S^*(c,s)$. For example if $x_S^*(c,s) = 4.25$, then one can insure that each college of type c admits either 4 or 5 students of type s.

7.3 The Marginal Cost of Stability and Capacity Constraints

Matching market designers strive to create economies that maximize participant welfare while maintaining incentive compatibility (i.e., straightforwardness) and stability (i.e., fairness). Obviously imposing either incentive compatibility or stability as desiderata will limit the scope for welfare enhancing design choices. The welfare cost of stability has been previously studied by considering a move from a stable mechanism to an unstable, but welfare improving, alternative mechanism (Erdil and Ergin [7], Abdulkadiroğlu, Pathak, and Roth [1], Featherstone [8]). In our framework we can compute the welfare benefit of marginal increases in school capacity or marginal relaxations of the stability constraint.

To make the notion of welfare concrete, assume that each agent assigns utility $-n$ to a college that possesses that student's n^{th} ranked trait and denote the associated utility function for students of type s as u_s. We can then write the student-optimal stable match as the solution to the following program, which maximizes the average student utility while satisfying the capacity, stability, and individual rationality constraints defined in section 4.1.

$$\begin{aligned}
x_S^{SO} &= \underset{x_S \in [0,1]^{\mathcal{C}\times\mathcal{S}}}{\arg\max} \sum_{s\in\mathcal{S}} \pi^S(s) \sum_{c\in\mathcal{C}} x_S(c,s) u_s(c) \text{ such that} \\
\text{For all } s &\in \mathcal{S},\ \sum_{c\in\mathcal{C}} x_S(c,s) \leq 1 \\
\text{For all } c &\in \mathcal{C},\ \sum_{s\in\mathcal{S}} x_S(c,s) * \pi^S(s) \leq q_c * \pi^C(c) \\
\text{For all } (c,s) &\in \Gamma,\ \left(1 - \sum_{\{c'\in\mathcal{C}: c' \succeq_s c\}} x_S(c',s)\right) * \\
&\qquad \left(q_c - \sum_{\{s'\in\mathcal{S}: s' \succeq_c s\}} x_C(c,s')\right) \leq 0 \\
\text{For all } (c,s) &\notin \Gamma \text{ we have } x_S(c,s) = 0
\end{aligned}$$

First consider the case of capacity constraints. Weakening a capacity constraint for college type c can be accomplished by either adding seats to colleges of a particular type (i.e., increasing q_c) or by building more schools (i.e., increasing $\pi^C(c)$). Let the set of student types who are admitted to colleges of type c be

$$\mathcal{A}(c) = \{s : x_S^{SO}(c, s) > 0\}$$

Let the set of students who would be *eager* to attend a college of type c if capacity were available be defined as

$$\mathcal{E}(c) = \{s : c \succ_s \underline{x}_S(s)\}$$

The increased capacity will be allocated to students types in $\mathcal{A}(c) \cap \mathcal{E}(c)$. Note that an increase in capacity can only benefit the students if the college is overdemanded. If the college is overdemanded, for generic values of (π^C, π^S) the set $\mathcal{A}(c) \cap \mathcal{E}(c)$ will be nonempty. The total welfare gain for a student $s \in \mathcal{A}(c) \cap \mathcal{E}(c)$ of obtaining one of the new seats in college c is

$$u_s(c) - u_s(\underline{x}_S(s)) \tag{7.1}$$

The welfare improvement for the student $s \in \mathcal{A}(c) \cap \mathcal{E}(c)$ that maximizes equation 7.1 describes the welfare gain due to marginal increases in school capacity.

We now consider the effect of slackening a stability constraint of a college-student pair while leaving the college capacities fixed. In real-life matching problems, it is common for students (or their parents) to lobby schools to open a seat for their child. This lobbying, if successful, results in a (c, s) assignment for which there is a blocking pair. We refer to such a (c, s) assignment as an unstable pair. The question then is what are the welfare effects of allowing these unstable pairs to form. Relatedly, if our model suggests that a student would benefit greatly from being allowed to form an unstable pair, we should expect that the lobbying efforts in favor of this unstable pair on the part of the parents to be intense.

The marginal welfare gain to a student of type s who is allowed to form an unstable pair with c is described by equation 7.1. Note that s need not be in either $\mathcal{A}(c)$ or $\mathcal{E}(c)$. Therefore the maximum possible marginal welfare gain for allowing a college of type c to form an unstable pair is

$$R = \max_{s \in \mathcal{S}} u_s(c) - u_s(\underline{x}_S(s))$$

Allowing a blocking pair (c, s) to form means that some student assigned to college c would now be unable to claim that seat. The displaced student is forced to match with his or her least preferred partner, $\underline{x}_S(s)$. To minimize these social welfare costs we must deny admission to the (previously admitted) student $s \in \mathcal{A}(c)$ with the minimal welfare loss. To compute this minimal loss we need to solve

$$C = \min_{s \in \mathcal{A}(c)} u_s(c) - u_s(\underline{x}_S(s))$$

The total welfare effect is then $R - C \geq 0$.

7.4 Matching with Contracts

Our framework extends easily to settings where agents on each side of the market can match with varying terms of trade that are encapsulated in contracts (Kelso and Crawford [13], Hatfield and Milgrom [11]). We refer to one side of the market as workers and the other side as firms. A firm and a worker match through a contract that dictates terms of trade such as wage and job tasks. A set of worker-firm matches is stable when there are no blocking pairs - in other words, when no unmatched worker-firm pair have a contract that is mutually preferred to a current match.[21]

To handle this extension, we treat each worker-contract combination as a different good on the worker-side of the market. Firms (without an associated contract) comprise the other side of the match. If we assume that the set of possible contracts is finite, then equations 4.5 through 4.8 continue to characterize the set of stable matches with the caveat that equation 4.5, which captures the capacity constraints for a worker of type s, must now sum over all worker-contract pairs that involve a worker of type s. Once this modification is made, it is straightforward to show that our results on existence of a pairwise stable match (proposition 2), the implementability of the match as a mixture of pure assignments (proposition 3), and the continuity of the correspondence of matches (theorem 1) all continue to hold. We could then use our model to (for example) describe the risk workers face with regard to their wage and the uncertainty of the composition of each firm's work force.

[21] Throughout we assume workers can match with a single firm, but firms may employ multiple workers through different contracts.

8 Conclusion

In this paper we introduce a novel model of large matching markets that we use to study the trade-off between rich spaces of traits used to describe agents and the risk faced by the participants. We provide a characterization of the stable set of our model and prove that the match can be chosen to be continuous in the distribution of preferences declared to the mechanism. The continuity allows us to use central limit theorem results to characterize the distribution of market outcomes when the number of participants is large.

We study the rate at which market outcomes become predictable (i.e., how rapidly the finite model approaches the complete-information limit) when the number of traits grows with N. We are able to use our structure to provide exact asymptotic distributions of match outcomes as a function of the number of agents, N, and the number of traits that are used to describe the agents. We argue that asymptotic convergence can be slow even for mild rates of growth in the number of traits and that the welfare effects of the risk are particularly acute when preference rankings declared by the agents are truncated. This suggests that participants may be exposed to a high degree of risk in markets of practical size. In addition, we argue that if the market designer requires a uniform bound on the level of risk faced by participants, then the required bound can be harder to satisfy as N increases for realistic market sizes and moderate rates of increase in the richness of the type-space.

Although most matching markets employed at present allow agents to declare preferences from a very rich class of truncated preferences (or stated differently, use a rich type space), our results suggest that this design could have significant costs. We propose that matching market designers ought to consider allowing (or even requiring) participants to rank categories of potential partners through a process we call conflation. By allowing for the coarse ranking of a greater number of partners, conflation can result in a significant reduction in the uncertainty faced by the participants and mitigate the resulting welfare losses.

One can view the choice of the space of preferences that can be expressed to the mechanism as an exercise in providing a useful way for the agents to describe an approximate preference relation. Using approximations may be optimal given the constraints on the agents' ability to describe their true preferences. Standard implementations of matching markets focus on allowing agents to precisely rank a small number of partners, while conflation allows agents to coarsely rank a much larger number of partners. If agent preferences over partners can be represented by a utility function over a relatively small number of hedonic attributes, then it may be better still to let the agents describe linear approximations of the utility functions over the hedonic attributes that can be used to approximately rank all of the potential partners.

As a final note, if a student ranks a designed trait in his preference list and is matched to such a partner, instabilities may result since the student is not actually indifferent between the conflated primitive traits. However, it is obvious that when conducting a match in an incomplete information environment, truncated preference lists alone can result in unstable outcomes as in the example in the introduction. The important question, which we leave for future work, is whether the instabilities caused by conflation are more numerous or important than those caused by truncated preferences. The answer will clearly depend both on the form of the conflation analyzed and what one considers an "important" instability.

References

[1] Abdulkadiroğlu, A.; P. Pathak; and A. Roth (2009) "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99, pp. 1954 - 1978.

[2] Abdulkadiroğlu, A. and T. Sönmez (2009) "School Choice: A mechanism design approach, "*American Economic Review*, 93 (3), pp. 1954 - 1978.

[3] Azevedo, E. and J. Leshno (2011) "The College Admissions Problem with a Continuum of Students," *mimeo.*-

[4] Budish, E., Y-K Che, F. Kojima, and P. Milgrom (2013) "Designing Random Allocation Mechanisms: Theory and Applications," *American Economic Review*, 103, pp. 585-623.

[5] Echenique, F.; S. Lee; M. Shum; and M.B. Yenmez (2013) "The Revealed Preference Theory of Stable And Extremal Stable Matchings," *Econometrica*, 83, pp. 153 - 171.

[6] Embrechts, P.; C. Klüppelberg; and T. Mikosch (1997) *Modelling Extremal Events for Insurance and Finance*, Springer: New York.

[7] Erdil, A. and H. Ergin (2006) "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98, pp. 669-689.

[8] Featherstone, C. (2006) "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice," *American Economic Review*, 98, pp. 669-689.

[9] Gale, D. and L.S. Shapley (1962) "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, pp. 9-15

[10] Haeringer, G. and F. Klijn (2009) "Constrained School Choice," *Journal of Economics Theory*, 144, pp. 1921-1947.

[11] Hatfield, J. and Milgrom, P. (2005) "Matching with Contracts," *The American Economic Review*, 95 (4), pp. 913-935.

[12] Immorlica, N. and M. Mahdian (2005) "Marriage, Honesty, and Stability," *SODA 2005*, pp. 53-62.

[13] Kelso, A. and V. Crawford (1982) "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50 (6), pp. 1483-1504.

[14] Kojima, F. and P. Pathak (2007) "Incentives and Stability in Large Two Sided Matching Markets," *mimeo.*

[15] Lee, S. (2012) "Incentive Compatibility of Large Centralized Matching Markets," *mimeo.*

[16] Levin, J. and P. Milgrom (2010) "Online Advertising: Heterogeneity and Conflation in Market Design," *American Economic Review: Papers and Proceedings*, pp. 603 - 607.

[17] Niederle, M.; A. Roth; U. Unver (2009) "Unraveling Results From Comparable Demand and Supply: An Experimental Investigation," *mimeo.*

[18] P. Pathak and T. Sönmez (2013) "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation," *American Economic Review*, 103 (1), pp. 80-106.

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

[20] Roth, A. and E. Peranson (1999) "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Market Design," *The American Economic Review,* 89 (4), pp. 748 - 780.

[21] Roth, A.; T. Sönmez; and M.U. Unver (2007) "Efficient Kidney Exchange: Coincidence of Wants in Markets with Compatability-Based Preferences," *The American Economic Review,* 97 (3), pp. 828-851.

[22] Roth, A. and M. Sotomayor (1990) *Two Sided Matching: a Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press: Cambridge, England.

[23] Roth, A. and J. Vande Vate (1991) "Incentives in Two-Sided Matching with Random Stable Matches," *Economic Theory,* 1, pp. 31 - 44.

[24] Sönmez, T. and T. Switzer (2013) "Matching with (Branch-of-Choice) Contracts at the United States Military Academy" *Econometrica,* 81, pp. 451-488.

[25] van der Vaart, A. and J. Wellner (1996) *Weak Convergence and Empirical Processes*, Springer-Verlag: New York.

A Proofs

Lemma 1. *Equation 4.2 holds for a pair (c, s) if and only if $c \succ_s \underline{x}_S(s)$ and $s \succ_c \underline{x}_C(c)$*

Proof. Note that if $c \succ_s \underline{x}_S(s)$, then there must be a positive mass of students either unmatched or matched to colleges to which c is strictly preferred by s. Therefore

$$\sum_{\{c':c' \succeq_s c\}} x_S(c', s) < 1.$$

Similarly, if $s \succ_c \underline{x}_C(c)$ then then it must be that a positive mass of college seats are either unoccupied or occupied by students to which s is strictly preferred by c. Therefore $\sum_{\{s':s' \succeq_c s\}} x_C(c, s') < q_c$.

The logic of the reverse direction is similar (and omitted). □

Proposition 1. *Equations 4.5, 4.6, 4.7, and 4.8 are necessary and sufficient for a match x_S to be feasible, stable, and individually rational.*

Proof. Equations 4.5 and 4.6 are required feasibility constraints and the if-and-only-if argument for feasibility is immediate. Similarly individual rationality is (by definition) satisfied if-and-only-if equation 4.8 holds. Equation 4.7 is a compressed formulation of our definition of stability. □

Proposition 2. *There is at least one feasible and stable match.*

Proof. We can find a stable match using a modification of the student-proposing deferred acceptance algorithm to handle the continuum of agents, but otherwise the algorithm functions as in the case with a finite number of agents.

At each step of the algorithm, a type of student is chosen where 1) the type of student has a positive measure of unmatched agents at the current step of the algorithm and 2) students of this type have not already proposed to all acceptable colleges. Consider such a type of student $s \in \mathcal{S}$. All unmatched students of type s propose to the same college type $c \in \mathcal{C}$, which is one of the most preferred colleges[22] to whom students of type s have not previously proposed. Denote the measure of such proposing students as $\phi > 0$. If c finds students of type s unacceptable, then all of the students' proposals are rejected. Otherwise, suppose that a measure ψ of seats at college c are either unmatched or are filled by students the college finds strictly worse than students of type s.

If $\phi > \psi$ then a fraction $\frac{\psi}{\phi}$ of the proposers are accepted and the remainder have their match refused. All less preferred students currently accepted at colleges of type c are rejected and the seats are filled with a measure ψ of students of type s.

If $\phi < \psi$ then all of the proposers are accepted. The empty seats in the colleges are filled first. Students currently matched with c and strictly inferior to s then have their matches broken until the full measure ϕ of type s students are matched. We assume that current matches are broken from the least preferred type(s) of students to the most with indifferences broken arbitrarily.

In the next step of the algorithm, another type of student is selected and the proposal procedure above is repeated. Since preference lists over traits are finite, the

[22]Since the students are indifferent between colleges that have the same trait, there may be multiple such colleges. For expositional ease, we demand that all student propose to the same college type.

algorithm halts in finite time. By construction, the final result is a feasible and stable match. □

Theorem 1. *Let $G(\pi^C, \pi^S)$ denote the correspondence of feasible and stable matches given (π^C, π^S). $G(\pi^C, \pi^S)$ is continuous on the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$. Furthermore there is a continuous selection $x(c, s; \pi^C, \pi^S)$ from G that is locally Lipschitz continuous on the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$.*

Proof. As a preliminary, note that the correspondence of feasible and stable matches is upper hemicontinuous over $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$ since the weak inequalities defining the set are continuous. What remains is to show that the correspondence is also lower hemicontinuous over the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$.

Consider (π^C, π^S) in the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$, so (π^C, π^S) have full support. Choose an arbitrary $x_S \in G(\pi^C, \pi^S)$ and two distributions of declared types (π^C, π^S) and $(\widetilde{\pi}^C, \widetilde{\pi}^S)$ where $(\widetilde{\pi}^C, \widetilde{\pi}^S)$ has full support and $\left\|\widetilde{\pi}^C - \pi^C\right\| + \left\|\widetilde{\pi}^S - \pi^S\right\| < \varepsilon$ for some $\varepsilon > 0$. We use the modified deferred-acceptance algorithm described in lemma 2 to construct an element $\widetilde{x}_S \in G(\widetilde{\pi}^C, \widetilde{\pi}^S)$ and argue that given our construction we have $\|\widetilde{x}_S - x_S\| \to 0$ as $\varepsilon \to 0$.

First note that we can link (π^C, π^S) to $(\widetilde{\pi}^C, \widetilde{\pi}^S)$ using a finite chain of perturbed measures $\{(\pi_i^C, \pi_i^S)\}_{i=1}^n$ where $(\pi^C, \pi^S) = (\pi_1^C, \pi_1^S)$, $(\widetilde{\pi}^C, \widetilde{\pi}^S) = (\pi_n^C, \pi_n^S)$, and for each i we have one of the following conditions

$$\begin{aligned}
\pi_i^C &= \pi_{i+1}^C,\ \pi_i^S(s) \neq \pi_{i+1}^S(s) \text{ for some } s, \text{ and } \pi_i^S(s') = \pi_{i+1}^S(s') \text{ for all } s' \neq s \\
\pi_i^S &= \pi_{i+1}^S,\ \pi_i^C(c) \neq \pi_{i+1}^C(c) \text{ for some } c, \text{ and } \pi_i^C(c') = \pi_{i+1}^C(c') \text{ for all } c' \neq c
\end{aligned}$$

Note that $\left\|\pi_i^C - \pi_{i+1}^C\right\| \leq \left\|\widetilde{\pi}^C - \pi^C\right\| \leq \varepsilon$ and $\left\|\pi_i^S - \pi_{i+1}^S\right\| \leq \left\|\widetilde{\pi}^S - \pi^S\right\| \leq \varepsilon$. Also note that n can be chosen to be smaller than $|\mathcal{S}| + |\mathcal{C}|$. We now recursively construct a sequence $\{x_{S,i}\}_{i=1}^n$ where $x_{S,1} = x_S$, $x_{S,i}$ is feasible and stable under (π_i^C, π_i^S), and $\|x_{S,i} - x_{S,i+1}\| < \kappa\varepsilon$ for some $\kappa > 0$. Therefore

$$\begin{aligned}
\|\widetilde{x}_S - x_S\| &\leq \sum_{i=1}^n \|x_{S,i} - x_{S,i+1}\| \\
&< n\kappa\varepsilon
\end{aligned}$$

which implies $\|\widetilde{x}_S - x_S\| \to 0$ as $\varepsilon \to 0$, generating our desired continuity result.

Consider some arbitrary $x_{S,i}$ that is feasible and stable with respect to (π_i^C, π_i^S).

There are four cases to consider. First suppose that $\pi^S_{i+1}(s) - \varepsilon_i = \pi^S_i(s)$ for some s and $0 < \varepsilon_i \leq \varepsilon$. We begin our construction of $x_{S,i+1}$ by letting $x_{S,i+1} = x_{S,i}$ and letting a measure ε of students of type s be unmatched. We construct $x_{S,i+1}$ by running our modified version of the modified Gale-Shapley procedure described in the proof of proposition 2 with the first step being the unmatched students of type s proposing to a favorite college c. Since $x_{S,i}$ is stable, we know that the first college to accept an offer from the students of type s will be a college that is no more preferred than the worst match partner under $x_{S,i}$. If the initially unmatched students of type s displace some previously matched students, let the second step of the Gale-Shapley procedure allow proposals by any remaining students of type s and these displaced students. Repeat this procedure until either none of the students are unmatched or the unmatched students have proposed to all acceptable partners. Note that at each stage at most ε students have their match altered, and the Gale-Shapley procedure must halt in at most $|\mathcal{S}| * |\mathcal{C}|$ steps, implying that $\|x_{S,i} - x_{S,i+1}\| \leq \kappa\varepsilon$ where we define $m = \min\{\pi^S(s)\}$ and

$$\kappa > \frac{|\mathcal{S}| * |\mathcal{C}|}{m}$$

Our other three cases are handled similarly. If $\pi^C_{i+1}(c) - \varepsilon_i = \pi^C_i(c)$ for some c and $0 < \varepsilon_i \leq \varepsilon$, proceed as in the first case but use the college-proposing version of our modified deferred acceptance algorithm. In the third case, suppose $\pi^C_{i+1}(c) + \varepsilon_i = \pi^C_i(c)$ for some c and $0 < \varepsilon_i \leq \varepsilon$ where $\pi^C_{i+1}(c) > 0$. We break a measure ε_i of matches of colleges of type c with students with types in the set

$$\max\{s : s \preceq_c s' \text{ for all } s' \text{ such that } x_C(c, s'), x_S(c, s) > 0\}$$

where the maximum is with respect to $\preceq_c$. For sufficiently small ε_i, this creates a match $x'_{S,i}$ and a measure ε_i of unmatched students, and $x'_{S,i}$ is stable under π^C_{i+1} (except for blocking pairs composed of an agent from the broken matches). Use the student-proposing modified deferred acceptance algorithm to find a stable and feasible $x_{S,i+1}$. The fourth case, $\pi^S_{i+1}(s) + \varepsilon_i = \pi^S_i(s)$, proceeds similarly by breaking matches with colleges and reassigning the newly unmatched colleges using the college-proposing modified deferred acceptance algorithm.

Note that in all cases above, the stable matches are within $n\kappa\varepsilon$ of each other for some value of κ that holds over an open subset of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$ that contains

(π^C, π^S). This implies our claim that the selection can be chosen to be locally Lipschitz continuous on the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$.

The only remaining issue is $G(\pi^C, \pi^S)$ where (π^C, π^S) do not have full support. We define $\overline{G}$ as the closure of G defined over the interior of $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$, then we have that $\overline{G}$ is a subset of G and that we may draw a continuous selection from $\overline{G}$ that has the domain $\Delta(\mathcal{C}) \times \Delta(\mathcal{S})$. The issue is that for (π^C, π^S) without full support, we have the freedom to arbitrarily match agents with measure 0 under (π^C, π^S). Some of these matches will violate continuity, and these are the only matches in G not contained in $\overline{G}$. □

Proposition 3. *Any stochastic match satisfying equations 4.5 through 4.8 can be implemented.*

Proof. We rely on theorem 1 of Budish et al. [4]. From this result, it suffices to argue that the constraints defining our match are a bihierarchy. The capacity constraints on students and colleges and the individual rationality constraints obviously form a bihierarchy. What remains are the stability constraints, which are difficult to interpret in the context of Budish et al. [4] since they do not fit into the linear constraint structure studied therein. Given a particular stable match, a binding stability constraint for student-college pair (c, s) requires one or both of the following to hold

$$\sum_{\{c' \in \mathcal{C} : c' \succeq_s c\}} x_S(c', s) = 1 \tag{A.1}$$

$$\sum_{\{s' \in \mathcal{S} : s' \succeq_c s\}} x_C(c, s') = q_c \tag{A.2}$$

When written in this way, it is clear that each equation involves a sum over a subset of the (c, s) pairs that appear in the capacity constraint for the student and college. So equations 4.5 and A.1 form a hierarchy, while equations 4.6 and A.2 form a second hierarchy. The individual rationality constraints mandate that certain cells of $\mathbf{X}$ be 0, which implies that the corresponding cells of each of $\{\mathbf{X_i}\}_{i=1}^{A}$ also be zero. Since each $\{\mathbf{X_i}\}_{i=1}^{A}$ is weakly positive, these requirements do not affect our ability to implement $\{\mathbf{X}\}$. With these difficulties resolved, theorem 1 of Budish et al. [4] implies $\mathbf{X}$ can be implemented. □

Proposition 4. *As $N \to \infty$ we we have for $i \in \{C, S\}$*

$$\begin{aligned}
\Pr\{p_S(s,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{c:t_c=t\}} x_S(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{c:t_c=t} \upsilon_S^2(c,s)\right]^{-0.5}\right) \\
\Pr\{p_C(c,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{s:t_s=t\}} x_C(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{\{s:t_s=t\}} \upsilon_C^2(c,s)\right]^{-0.5}\right)
\end{aligned}$$

where we define the diagonal matrices

$$\Sigma_S = \begin{pmatrix} \sigma_S^2(s_1) & \dots & 0 \\ \dots & & \dots \\ 0 & \dots & \sigma_S^2(s_K) \end{pmatrix}, \Sigma_C = \begin{pmatrix} \sigma_C^2(c_1) & \dots & 0 \\ \dots & & \dots \\ 0 & \dots & \sigma_C^2(c_L) \end{pmatrix}$$

and the variances

$$\begin{aligned}
\upsilon_S^2(c,s) &= \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right] + \\
&\qquad \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right] \\
\upsilon_C^2(c,s) &= \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right] + \\
&\qquad \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right]
\end{aligned}$$

Proof. We now provide lower bounds on the probability of large deviations of p_S and p_C using the convergence results for $\pi_E^{S,N}$ and $\pi_E^{C,N}$. Central limit theorems imply

$$\begin{aligned}
\sqrt{N}\left(\pi_E^{S,N}(s) - \pi^{S,N}(s)\right) &\xrightarrow{d} N(0, \sigma_S^2(s)) \\
\sqrt{N}\left(\pi_E^{C,N}(c) - \pi^{C,N}(c)\right) &\xrightarrow{d} N(0, \sigma_C^2(c))
\end{aligned}$$

where $\sigma_S^2(s) = \pi^{S,N}(s) * (1 - \pi^{S,N}(s))$ and $\sigma_C^2(c) = \pi^{C,N}(c) * (1 - \pi^{C,N}(c))$. We use the delta method[23] to compute a central limit approximation for the terms

$$x_i(c,s;\pi_E^{C,N},\pi_E^{S,N}) - x_i(c,s;\pi^{C,N},\pi^{S,N}), i \in \{C,S\}$$

[23]Since the selection from the correspondence of feasible and stable matches can be chosen to be locally Lipschitz continuous, we know that our matching function, x_S, is differentiable for generic choices of $\pi^{S,N}$ and $\pi^{C,N}$.

We use the following notation where $i \in \{C, S\}$

$$\nabla_S x_i(c,s;\pi^{C,N},\pi^{S,N}) = \begin{pmatrix} \frac{\partial x_i}{\partial \pi^{S,N}(s_1)} \\ ... \\ \frac{\partial x_i}{\partial \pi^{S,N}(s_K)} \end{pmatrix}, \nabla_C x_i(c,s;\pi^{C,N},\pi^{S,N}) = \begin{pmatrix} \frac{\partial x_i}{\partial \pi^{C,N}(c_1)} \\ ... \\ \frac{\partial x_i}{\partial \pi^{C,N}(c_L)} \end{pmatrix}$$

and we define the diagonal matrices

$$\Sigma_S = \begin{pmatrix} \sigma_S^2(s_1) & ... & 0 \\ ... & & ... \\ 0 & ... & \sigma_S^2(s_K) \end{pmatrix}, \Sigma_C = \begin{pmatrix} \sigma_C^2(c_1) & ... & 0 \\ ... & & ... \\ 0 & ... & \sigma_C^2(c_L) \end{pmatrix}$$

and the variances

$$\begin{aligned} \upsilon_S^2(c,s) &= \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_S(c,s;\pi^{C,N},\pi^{S,N})\right] + \\ &\quad \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_S(c,s;\pi^{C,N},\pi^{S,N})\right] \\ \upsilon_C^2(c,s) &= \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_S \left[\nabla_S x_C(c,s;\pi^{C,N},\pi^{S,N})\right] + \\ &\quad \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right]^T \Sigma_C \left[\nabla_C x_C(c,s;\pi^{C,N},\pi^{S,N})\right] \end{aligned}$$

While these terms are notationally cumbersome, they are easy to compute numerically from our continuum model. Using the delta method we then have

$$\begin{aligned} \sqrt{N}\left(x_S(c,s;\pi^{C,N},\pi^{S,N}) - x_S(c,s;\pi_E^{C,N},\pi_E^{S,N})\right) &\xrightarrow{d} N(0,\upsilon_S^2(u,s)) \\ \sqrt{N}\left(x_C(C,s;\pi^{C,N},\pi^{S,N}) - x_C(c,s;\pi_E^{C,N},\pi_E^{S,N})\right) &\xrightarrow{d} N(0,\upsilon_C^2(c,s)) \end{aligned}$$

This implies that for large N we can use the following approximation for $i \in \{C, S\}$

$$\begin{aligned} \Pr\{p_S(s,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{c:t_c=t\}} x_S(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{c:t_c=t} \upsilon_S^2(c,s)\right]^{-0.5}\right) \\ \Pr\{p_C(c,t) &\geq \delta\} \approx 1 - \Phi\left(\frac{\delta\sqrt{N}}{\sum_{\{s:t_s=t\}} x_C(c,s;\pi^{C,N},\pi^{S,N})}\left[\sum_{\{s:t_s=t\}} \upsilon_C^2(c,s)\right]^{-0.5}\right) \end{aligned}$$

□