Online Appendices for Large Matching Markets: Risk, Unraveling, and Conflation

Aaron L. Bodoh-Creed - Cornell University

A Online Appendix: Strategic Convergence

In section 4 we described the matching mechanism we use and how we determine the set of stable outcomes given the preference declarations of the agents. In this section we address the equilibrium outcomes in games where the agents can make false preference declarations. In section A.1 we provide a strategic convergence result showing that equilibria of our limit game approximate the equilibria of a game with a sufficiently large, finite set of agents. In section A.2 we use these strategic convergence results to show that since there exist matching markets where every equilibrium of the limit game is non-truthful, for sufficiently large N any equilibrium of the N-agent game is non-truthful.

A.1 Agent Preferences and Equilibrium

Our focus is on matching games with incomplete-information, so we will need to be explicit regarding the risk preferences of the students and colleges. This requires us to use the von Neumann-Morgenstern representations defined in section 6. We model the matching game as a revelation mechanism, so the noncooperative game has an action space for all agents equal to $\mathcal{A} = \mathcal{C} \cup \mathcal{S}$. Since agent types are comprised of traits (that are observable) and preferences (that are declared), we restrict the strategies so that agents must declare traits truthfully. The agents declare their preferences without observing the types of the other participants on either side of the market. Even in applications like school matching where colleges can determine the student traits from academic records and most relevant information regarding colleges is publicly available, it may be very costly to acquire and evaluate this information. In appendix B we provide an extension to the case where agents can observe traits, but not preferences, of the other agents in the match. Propositions B.1 and B.3 extend propositions A.1 and A.2 developed in this section to the alternative assumption that the agents observe the distribution of traits on each side of the market.[1]

We focus on equilibria in symmetric strategies for the students (colleges) denoted $\sigma_S : \mathcal{S} \to \Delta(\mathcal{S})$ ($\sigma_C : \mathcal{C} \to \Delta(\mathcal{C})$) with the strategy space denoted Σ^S (Σ^C). Let $\sigma_S^*(s)(s')$ denote the probability that a student of type $s \in \mathcal{S}$ declares he is of type $s' \in \mathcal{S}$ and $\sigma_C^*(c)(c')$ denote the probability that a college of type $c \in \mathcal{C}$ declares it is of type $c' \in \mathcal{C}$. Let $\Delta_N(X)$ denote an empirical measure of N realizations from a random variable over X. Given a realized distribution of the true types for students in the N-agent game π_N^S, we denote the declared distributions of types given the equilibrium strategies as $\widehat{\pi}_N^S(s; \sigma_S^*) \in \Delta_N(\mathcal{S})$. Analogously, $\widehat{\pi}_N^C(c; \sigma_C^*) \in \Delta_{M_N}(\mathcal{C})$ is an empirical distribution of college types declared in the N agent game. If the agents use pure strategies, the distribution of declared types (conditional on the realized distribution of true types) can be written

$$\begin{aligned}
\text{For any } s &\in \mathcal{S},\ \widehat{\pi}_N^S(s'; \sigma_S^*) = \sum_{s \in \mathcal{S}} \sigma_S^*(s)(s') * \pi_N^S(s) \\
\text{For any } c &\in \mathcal{C},\ \widehat{\pi}_N^C(c'; \sigma_C^*) = \sum_{c \in \mathcal{C}} \sigma_C^*(c)(c') * \pi_N^C(c)
\end{aligned}$$

For notational convenience, we suppress arguments and write $\widehat{\pi}_N^S$ and $\widehat{\pi}_N^C$ for $\widehat{\pi}_N^S(s'; \sigma_S^*)$ and $\widehat{\pi}_N^C(c'; \sigma_C^*)$. If the agents employ mixed strategies, then the distributions of declared student and college types conditional on the realized distribution of types in the finite games are random variables defined analogously.

[1] In this alternative model we need to assume that the equilibrium strategies are continuous in the distribution of traits in the economy.

Interim preferences for a student of type $s \in \mathcal{S}$ declaring a feasible type $s' \in \mathcal{S}$ are

$$\mathbb{E}\left[\sum_{c\in\mathcal{C}} x_S(c, s'; \widehat{\pi}_N^C, \widehat{\pi}_N^S) * v(c, s)\right]$$

where the expectation reflects uncertainty over the realization of the types of the other agents and the resulting distributions of declared types $\widehat{\pi}_N^C$ and $\widehat{\pi}_N^S$. We denote college utilities given a declaration of a feasible type $c' \in \mathcal{C}$ as

$$\mathbb{E}\left[\sum_{S\in\mathcal{S}_C} x_C(c', S; \widehat{\pi}_N^C, \widehat{\pi}_N^S) * W(c, S)\right]$$

where $x_C(c', S; \widehat{\pi}_N^C, \widehat{\pi}_N^S)$ denotes the fraction of colleges of type c matched with a class of students S. We interpret x_C and x_S as probabilities of various match outcomes.

We use the following equilibrium concept for games with a finite number of agents. δ_s (δ_c) refer to a measure that places unit mass on the student (college) type s (c), and the notation supp$[\pi]$ denotes the support of the random variable π.

Definition A.1. *Given an N-agent matching game, the strategies $\sigma_S : \mathcal{S} \to \Delta(\mathcal{S})$ and $\sigma_C : \mathcal{C} \to \Delta(\mathcal{C})$ comprise an **Ex Post (ε, ρ)-Nash Equilibrium** if for all types $s \in \mathcal{S}$ and $c \in \mathcal{C}$ in the support of π^S and π^U, any $\widetilde{s} \in supp[\sigma_S(s)], s' \in \mathcal{S}$ and any $\widetilde{c} \in supp[\sigma_C(c)], c' \in \mathcal{C}$ we have that with probability $1 - \rho$*

$$\sum_{c\in\mathcal{C}} x_S(c, \widetilde{s}; \widehat{\pi}_N^C, \widehat{\pi}_N^S) * v(c, s) + \varepsilon \geq \sum_{c\in\mathcal{C}} x_S\left(c, s'; \widehat{\pi}_N^C, \widehat{\pi}_N^S + \frac{1}{N}[\delta_{s'} - \delta_{\widetilde{s}}]\right) * v(c, s)$$

$$\sum_{S\in\mathcal{S}_C} x_C(\widetilde{c}, S; \widehat{\pi}_N^C, \widehat{\pi}_N^S) * W(c, S) + \varepsilon \geq \sum_{S\in\mathcal{S}_C} x_C\left(c', S; \widehat{\pi}_N^U + \frac{1}{M_N}[\delta_{c'} - \delta_{\widetilde{c}}], \widehat{\pi}_N^S\right) * W(c, S)$$

The ex post aspect of the equilibrium requires that agent actions be approximately optimal once all of the actions are common knowledge. Note that deviations change the agents' utilities indirectly by altering the ex post empirical distribution of actions in the game, which is captured by the terms $\frac{1}{N}[\delta_{s'} - \delta_{\widetilde{s}}]$ and $\frac{1}{M_N}[\delta_{c'} - \delta_{\widetilde{c}}]$.

In the limit game, strategies $(\sigma_C^\infty, \sigma_S^\infty)$ induce declared distributions of types defined by

$$\begin{aligned}\text{For any } s &\in \mathcal{S},\ \widehat{\pi}_\infty^S(s;\sigma_S^\infty) = \sum_{s'\in\mathcal{S}} \sigma_S^\infty(s')(s) * \pi^S(s)\\ \text{For any } c &\in \mathcal{C},\ \widehat{\pi}_\infty^U(u;\sigma_U^\infty) = \sum_{c'\in\mathcal{C}} \sigma_C^\infty(c')(c) * \pi^C(c)\end{aligned}$$

The following equilibrium notion is applied to our limit game.

Definition A.2. *The strategies $\sigma_S^\infty : \mathcal{S} \to \Delta(\mathcal{S})$ and $\sigma_C^\infty : \mathcal{C} \to \Delta(\mathcal{C})$ comprise an ε-Nash equilibrium (ε-NE) of the nonatomic game if for all types $s \in \mathcal{S}$ and $c \in \mathcal{C}$ in the support of π^S and π^C, any $\widetilde{s} \in supp[\sigma_S(s)], s' \in \mathcal{S}$, and any $\widetilde{c} \in supp[\sigma_C(c)], c' \in \mathcal{C}$ we have that*

$$\begin{aligned}\sum_{c\in\mathcal{C}} x_S(c,\widetilde{s};\widehat{\pi}_\infty^C,\widehat{\pi}_\infty^S) * v(c,s) + \varepsilon &\geq \sum_{c\in\mathcal{C}} x_S(c,s';\widehat{\pi}_\infty^C,\widehat{\pi}_\infty^S) * v(c,s)\\ \sum_{S\in\mathcal{S}_C} x_C(\widetilde{c},S;\widehat{\pi}_\infty^C,\widehat{\pi}_\infty^S) * W(c,S) + \varepsilon &\geq \sum_{S\in\mathcal{S}_C} x_C(c',S;\widehat{\pi}_\infty^C,\widehat{\pi}_\infty^S) * W(c,S)\end{aligned}$$

The critical difference between the equilibrium definition employed in the nonatomic and finite games lies in the fact that nonatomic agents cannot individually affect the distribution of declared types in the nonatomic limit game. The nonatomic agents take these distributions as exogenous to their own action and optimize with respect to the decision problem these distributions generate.

A subtle issue regarding our equilibrium concept for the nonatomic matching game arises when an agent deviates to declare a type that is not in the support of the declared types. Formally since the deviator is of 0 measure, none of the feasibility or stability conditions would be violated by arbitrarily matching such a deviator (even to partners with declared types that are not present in the match). However, our assumption that the matching mechanism is continuous dictates the payoffs from these forms of deviations.[2]

Our results on equilibrium convergence are based on theorems 1, 2 and 3 of Bodoh-Creed [2]. For completeness we restate these theorems here without proof. Bodoh-Creed [2] uses the notation u for a utility function, θ for an agent type, and $\mathcal{A}$ for an action space. $\mathcal{E}^{NA}$ refers to the space of equilibria of the nonatomic limit game, while $\mathcal{E}^{\infty}$ refers to the limits of convergent sequences of equilibria of the finite game (i.e., sequences of the form $\{(\sigma_C^N, \sigma_S^N)\}_{N=1}^{\infty}$ where (σ_C^N, σ_S^N) is an equilibrium strategy pair in the N-agent game). Bodoh-Creed [2] incorporates an aggregate market variable ω that is not observed by the agents. For our purposes assume ω takes a single value in our matching model, which effectively eliminates ω from our setting.

Theorem 1 of Bodoh-Creed [8]. *Endow $\Delta(\mathcal{A})$ with the weak-* topology. The correspondence $\mathcal{E}$ is upper hemicontinuous as $N \to \infty$ ($\mathcal{E}^{\infty} \subset \mathcal{E}^{NA}$)* ***if*** *the family $\{E^{\Omega}[u(\theta, \omega, \cdot, \cdot)|\theta]\}_{\theta \in \Theta}$ is uniformly equicontinuous in $\mathcal{A} \times \Delta(\mathcal{A})$ and uniformly bounded.*

Theorem 2 of Bodoh-Creed [8]. *Fix $\varepsilon > 0$ and endow $\Delta(\mathcal{A})$ with the Kolmogorov topology. For each Bayesian-Nash equilibrium $\sigma^{\infty} : \Theta \to \Delta(\mathcal{A})$ of the nonatomic game $u : \Theta \times \Omega \times \mathcal{A} \times \Delta(\mathcal{A}) \to \mathbb{R}$, there exists an N^* such that for all $N > N^*$, $\sigma^{\infty}(\circ)$ is an ε-Bayesian-Nash equilibrium of the N-agent game $u_N : \Theta \times \Omega \times \mathcal{A} \times \Delta_N(\mathcal{A}) \to \mathbb{R}$* ***if*** *the family $\{E^{\Omega}[u(\theta, \omega, a, \cdot)|\theta]\}_{\theta \in \Theta, a \in \mathcal{A}}$ is uniformly bounded and uniformly equicontinuous in a relatively open set of $\Delta(\mathcal{A})$ containing $\cup_{\omega \in \Omega}\{\pi_0^{\mathcal{A}}(\omega)\}$*

Theorem 3 of Bodoh-Creed [8]. *Assume $\Omega = \{\omega\}$ and fix $\varepsilon > 0, \rho \in (0, 1]$. Endow $\Delta(\mathcal{A})$ with the Kolmogorov topology. For each Nash equilibrium $\sigma^{\infty} : \Theta \to \Delta(\mathcal{A})$ of the nonatomic game $u : \Theta \times \Omega \times \mathcal{A} \times \Delta(\mathcal{A}) \to \mathbb{R}$, there exists an N^* such that for all $N > N^*$, $\sigma^{\infty}(\circ)$ is an ex post ε-Nash equilibrium of*

[2] This issue (obviously) does not arise in the finite game since the act of deviating changes the distribution of declared types to admit a small positive measure of the deviator's declared type.

*the N-agent game $u_N : \Theta \times \Omega \times \mathcal{A} \times \Delta_N(\mathcal{A}) \to \mathbb{R}$ with probability $1 - \rho$ **if** the family $\{u(\theta, \omega, a, \cdot)\}_{\theta \in \Theta, a \in \mathcal{A}}$ is uniformly equicontinuous in a relatively open set of $\Delta(\mathcal{A})$ containing $\pi_0^{\mathcal{A}}(\omega)$*

We now state our equilibrium convergence results and use the results of Bodoh-Creed [2] to prove our results.

Proposition A.1. *Choose $\rho \in (0, 1], \varepsilon > 0$. Then there exists N^* such that for all $N > N^*$ we have that any 0-NE of the limit game, $(\sigma_C^\infty, \sigma_S^\infty)$, is an ex post (ε, ρ)-Nash equilibrium of the N-agent game. Furthermore, $(\sigma_C^\infty, \sigma_S^\infty)$ is an $\varepsilon-$Bayesian-Nash equilibrium of the N-agent game.*

Proof. From assumption 1 we know that the set of feasible and stable matches is continuous with respect to changes in the distribution of declared types. Given the finiteness of the type-space and the compactness of the space of matches, we have that the agent utility is bounded and uniformly equicontinuous. Our claim that $(\sigma_C^\infty, \sigma_S^\infty)$ is an ex post (ε, ρ)-Nash equilibrium of the N-agent game for sufficiently large N then follows from Theorem 3 of Bodoh-Creed [2]. From the boundedness we have that $(\sigma_C^\infty, \sigma_S^\infty)$ is an $\varepsilon-$Bayesian-Nash equilibrium of the N-agent game for sufficiently large N from Theorem 2 of Bodoh-Creed [2]. □

Proposition A.1 implies that the equilibria of the nonatomic limit game are approximate equilibria of the matching game with sufficiently many participants. The intuition underlying proposition A.1 is that in a sufficiently large game there is only a low probability that the realized distribution of declared types differs significantly from the distribution realized in the nonatomic limit game equilibrium. Since the model is continuous in the distribution of declared types, there is little to gain by deviating from $(\sigma_C^\infty, \sigma_S^\infty)$ by optimizing to account for either the frequent small deviations or the rare large deviations. This suggests that the ε-equilibria found in proposition A.1 may be appealing behavioral predictions.

Corollary A.1. *If $(\sigma_C^\infty, \sigma_S^\infty)$ is a strict 0-NE of the nonatomic game, then it is an ex post $(0, \rho)$-Nash equilibrium of the N-agent game for N sufficiently* *large. Furthermore, $(\sigma_C^\infty, \sigma_S^\infty)$ is an exact Bayesian-Nash equilibrium of the N-agent game for N sufficiently large.*

Proof. Let $\widehat{\pi}^C, \widehat{\pi}^S$ denote the measure of declared agent types in the limit game when all agents follow $(\sigma_C^\infty, \sigma_S^\infty)$, and let $\widehat{\pi}_N^C, \widehat{\pi}_N^S$ denote the corresponding distributions in the N-agent game. Since $(\sigma_C^\infty, \sigma_S^\infty)$ is a strict Nash equilibrium, there exists $\underline{\varepsilon} > 0$ such that for any $c_d \neq \sigma_C^\infty(u)$, $u' = \sigma_C^\infty(u)$, $s_d \neq \sigma_S^\infty(s)$, and $s' = \sigma_S^\infty(s)$ we have[3]

$$\begin{aligned} \sum_{c \in \mathcal{C}} \left[x_S(c, s'; \widehat{\pi}^C, \widehat{\pi}^S) - x_S(c, s_d; \widehat{\pi}^C, \widehat{\pi}^S)\right] * v(c, s) &> \underline{\varepsilon} \qquad \text{(A.1)} \\ \sum_{S \in \mathcal{S}_C} \left[x_C(c', S; \widehat{\pi}^C, \widehat{\pi}^S) - x_C(c_d, S; \widehat{\pi}^C, \widehat{\pi}^S)\right] * W(c, S) &> \underline{\varepsilon} \end{aligned}$$

From the law of large numbers we have that as $N \to \infty$ that $(\widehat{\pi}_N^C, \widehat{\pi}_N^S) \to (\widehat{\pi}^C, \widehat{\pi}^S)$. Therefore we have from the continuity of x (assumption 1) that for any $\varepsilon, \rho > 0$ there is a large enough N that with probability $1 - \rho$ that for all $s \in \mathcal{S}$, $S \in \mathcal{S}_C$, and $c \in \mathcal{C}$

$$\begin{aligned} \sum_{c \in \mathcal{C}} \left\|x_S(c, s; \widehat{\pi}_N^C, \widehat{\pi}_N^S) - x_S(c, s; \widehat{\pi}^C, \widehat{\pi}^S)\right\| * v(c, s) &< \varepsilon \qquad \text{(A.2)} \\ \sum_{S \in \mathcal{S}_C} \left\|x_C(c, S; \widehat{\pi}_N^C, \widehat{\pi}_N^S) - x_C(c, S; \widehat{\pi}^C, \widehat{\pi}^S)\right\| * W(c, S) &< \varepsilon \end{aligned}$$

Combining equations A.2 and A.1 and using the continuity of x_S we have for large enough N and choices of $\varepsilon < \frac{1}{2}\underline{\varepsilon}$

$$\begin{aligned} \sum_{c \in \mathcal{C}} \left(x_S(c, s'; \widehat{\pi}_N^C, \widehat{\pi}_N^S) - x_S(c, s_d; \widehat{\pi}_N^C, \widehat{\pi}_N^S + \frac{1}{N}\left[\delta_{s'} - \delta_{s_d}\right]) \right) * v(c, s) &> 0 \\ \sum_{S \in \mathcal{S}_C} \left(x_C(c', S; \widehat{\pi}_N^C, \widehat{\pi}_N^S) - x_C(c_d, S; \widehat{\pi}_N^C + \frac{1}{M_N}\left[\delta_{c'} - \delta_{c_d}\right], \widehat{\pi}_N^S) \right) * W(c, S) &> 0 \end{aligned}$$

[3]Since the equilibrium is strict, the agents do not mix between actions, so we focus on equilibrum actions rather than "actions in the support of the equilibrium startegy."

which implies that $(\sigma_C^\infty, \sigma_S^\infty)$ it is an ex post Nash equilibrium of the N-agent game with probability $1-\rho$, which proves our claim that $(\sigma_C^\infty, \sigma_S^\infty)$ is an ex post $(0,\rho)$-Nash equilibrium of the N-agent game for N sufficiently large. Since the utility of the agents is uniformly bounded, we have that the loss in the probability ρ event that either π_N^C or π_N^S is not close to π^C or π^S is bounded by M and by $\rho * M$ in expectation. Since $\rho > 0$ can be chosen to be arbitrarily small as $N \to 0$, this implies that $(\sigma_C^\infty, \sigma_S^\infty)$ is an exact Bayesian-Nash equilibrium of the N-agent game for N sufficiently large. □

Corollary A.1 strengthens Proposition A.1 by noting that if $(\sigma_C^\infty, \sigma_S^\infty)$ is an exact 0-NE of the limit game, then any deviation from $(\sigma_C^\infty, \sigma_S^\infty)$ in the limit game results in a strict loss of utility. Since the model is continuous in the realized distribution of types, the payoffs from following $(\sigma_C^\infty, \sigma_S^\infty)$ and deviating are approximately the same in a finite game with sufficiently many players so long as the deviation of the realized distribution of declared types from the distribution in the limit game is small. Corollary A.1 follows after noting that our law of large numbers result holds both for the actions taken in equilibrium and deviations from these actions, so the strict incentives of the equilibria in the nonatomic limit game are inherited by finite games that are sufficiently large.

What remains is to show that the large finite game does not admit equilibria that differ from *any* equilibrium of the nonatomic limit game. With this result in hand, we can claim that the equilibrium strategy set of our nonatomic limit game contains the set of possible Bayesian-Nash equilibria of the finite game. The following proposition encapsulates this by proving that the equilibrium correspondence is upper hemicontinuous in N.

Proposition A.2. *Consider a sequence of strategies $\{(\sigma_C^N, \sigma_S^N)\}_{N=1}^\infty$ where (σ_C^N, σ_S^N) is an Bayesian-Nash equilibrium strategy pair in the N-agent game and assume that $(\sigma_C^N, \sigma_S^N) \to (\sigma_C^\infty, \sigma_S^\infty)$. Then $(\sigma_C^\infty, \sigma_S^\infty)$ is a $0-$NE of the nonatomic limit game.*

Proof. From assumption 1, we know that the set of stable of matchings is continuous with respect to changes in the distribution of declared types. Given the continuity of the utility function with respect to the matching outcome and the compactness of the space of matches, we have that the agent utility is uniformly continuous. Our claim that $(\sigma_C^\infty, \sigma_S^\infty)$ is an ex post (ε, ρ)-Nash equilibrium of the N-agent game for sufficiently large N then follows from Theorem 1 of Bodoh-Creed [2]. □

A.2 Incentive Compatibility

The finite, complete-information matching literature abounds with examples where non-truthful declarations of types to the matching mechanism are optimal for some of the agents. Any such example can be mapped into our structure by letting π^S, π^C place equal probability on the types in the example and 0 probability on other types. From our convergence results (proposition A.2), if we demonstrate an example where no equilibrium in truthful strategies exists, there cannot exist a truthful equilibrium of the large finite game. The following example illustrates one matching market where incentive compatibility fails.

Example A.1. *Suppose $\mathcal{T} = \{t_1, t_2\}$ and $\mathcal{P} = \{\succ^a, \succ^b\}$ where $t_1 \succ^a t_2$ and $t_2 \succ^b t_1$. Assume that $q_c = 1$ for all colleges. Let $\succ^{a_T}$ denote the truncated preference ordering $t_1 \succ^{a_T} \emptyset \succ^{a_T} t_2$, and let $\succ^{b_T}$ denote the truncated preference ordering $t_2 \succ^{b_T} \emptyset \succ^{b_T} t_1$. The types in the economy are*

Student Type	*Probability*	*College Type*	*Probability*
$(t_1, \succ^a)$	$\frac{1}{3}$	$(t_1, \succ^b, 1)$	$\frac{1}{3}$
$(t_1, \succ^{a_T})$	$\frac{1}{3}$	$(t_2, \succ^a, 1)$	$\frac{1}{3}$
$(t_2, \succ^b)$	$\frac{1}{3}$	$(t_2, \succ^{a_T}, 1)$	$\frac{1}{3}$

If the agents declare their preferences truthfully, then the following match *is realized under the mechanism that maximizes student welfare.*[4]

Student Type	*Match Partner*
$(t_1, \succ^a)$	$(t_2, \succ^{a_T}, 1)$
$(t_1, \succ^{a_T})$	$(t_1, \succ^b, 1)$
$(t_2, \succ^b)$	$(t_2, \succ^a, 1)$

If a college of type $(t_2, \succ^a)$ had instead submitted the type $(t_2, \succ^{a_T})$ he would have matched with a partner of type $(t_1, \succ^a, 1)$. Therefore truthfulness is not incentive compatible.

Since the agents in our limit model have complete-information regarding the types of agents in the mechanism and the agents can declare preference lists of arbitrary length, the set of acceptable partners cannot vanish and the market remains thick. The thickness of the markets leaves open the possibility of non-truthful equilibria.

Our result sharply contrasts with the results of Immorlica and Mahdian [4] and Kojima and Pathak [5] that emphasize that uncertainty combined with preference lists of fixed length can render matching mechanisms approximately incentive compatible as $N \to \infty$. Because of the finite length of the preference lists and the rich type spaces assumed in these works, the measure of the set of partners that the an agent deems acceptable vanishes as $N \to \infty$ and the market is thin for all agents on both sides of the match.[5] The thin markets eliminates any benefit of nontruthful declarations.

B Online Appendix: Equilibrium Convergence with Observable Types

In this appendix we study the model of section A.1 under the alternative assumption that the agents can observe the traits of the participants on each

[4]If the students place the same utility on their favorite matches and second favorite partners (respectively), then this match maximizes an egalitarian social welfare function.

[5]We define the notion of a thick market differently than in Kojima and Pathak [5].

side of the market prior to submitting their preference lists to the mechanism. We denote the distribution over the traits realized among the students (colleges) in the N-agent game as ν_N^S (ν_N^C), which is the marginal distribution of π_N^S (π_N^C) over the space of traits, $\mathcal{T}$.

We assume symmetric strategies for the students (colleges) denoted $\sigma_S : \mathcal{S} \times \Delta_N(\mathcal{T}) \times \Delta_{M_N}(\mathcal{T}) \to \Delta(\mathcal{S})$ ($\sigma_C : \mathcal{C} \times \Delta_N(\mathcal{T}) \times \Delta_{M_N}(\mathcal{T}) \to \Delta(\mathcal{C})$). $\Delta_N(\mathcal{T}) \times \Delta_{M_N}(\mathcal{T})$ denotes the space of empirical distributions of traits of the N students ($\Delta_N(\mathcal{T})$) and the M_N colleges ($\Delta_{M_N}(\mathcal{T})$). A particular realization of an action for a student of type $s \in \mathcal{S}$ is $\sigma(s, \nu_N^S, \nu_N^C)$, which makes explicit that we allow the agents to condition on their own type and the distributions of student and college traits in the economy. Let the respective strategy spaces be denoted Σ^S and Σ^C.

As before we restrict the agents to truthfully declaring their observable traits to the mechanism, but the agents may nontruthfully declare their preferences. In equilibrium it will be common knowledge that the students (colleges) employ strategy $\sigma_S^* \in \Sigma^S$ ($\sigma_C^* \in \Sigma^C$). We denote the the empirical distribution of declared student types given the equilibrium strategies as $\widehat{\pi}_N^S(s; \sigma_S^*, \nu_N^S, \nu_N^C) \in \Delta_N(\mathcal{S})$, while the analogous distribution of declared college types is $\widehat{\pi}_N^C(c; \sigma_C^*, \nu_N^S, \nu_N^C) \in \Delta_{M_N}(\mathcal{C})$. For expositional ease, we suppress the arguments of $\widehat{\pi}_N^S$ and $\widehat{\pi}_N^C$.

The following theorem extends proposition A.1 to this setting. Let $\mathcal{E}_\infty(\nu_N^C, \nu_N^S) \subset \Sigma^C \times \Sigma^S$ denote the set of Bayesian-Nash equilibria of the nonatomic matching game given the distributions of traits (ν_N^C, ν_N^S). We now prove that the equilibria of the limit game are approximate equilibria of the large finite game, and we note that it is straightforward to get an analogous extension of Corollary A.1 in this setting.

Proposition B.1. *Choose $\rho \in (0, 1], \varepsilon > 0$. Then there exists N^* such that for all $N > N^*$ we have that any $(\sigma_C^\infty, \sigma_S^\infty) \in \mathcal{E}_\infty(\nu_N^C\ , \nu_N^S)$ is an ex post (ε, ρ)-Nash equilibrium of the N-agent game. Furthermore, $(\sigma_C^\infty, \sigma_S^\infty)$ is an $\varepsilon-$Bayesian-Nash equilibrium of the N-agent game.*

Proof. Consider a sequence of K-agent games, $N < K \to \infty$, with trait distribution (ν_N^C, ν_N^S) and declared type distribution drawn in an i.i.d. fashion from the distributions

$$\begin{aligned}
\widehat{\pi}^C|\nu_N^C(c) &= \sum_{c' \in C, t \in \mathcal{T}} \sigma_\infty^C(c', \nu_N^C, \nu_N^S)(c) * \pi^C(c'|t) * \nu_N^C(t) \\
\widehat{\pi}^S|\nu_N^S(s) &= \sum_{s' \in S, t \in \mathcal{T}} \sigma_\infty^S(s', \nu_N^C, \nu_N^S)(s) * \pi^S(s'|t) * \nu_N^C(t)
\end{aligned}$$

where $\pi^C(\circ|t)$ and $\pi^S(\circ|t)$ represent conditional measures. From assumption 1 we have that x is continuous in the declared type distributions $\widehat{\pi}^C$ and $\widehat{\pi}^S$. Given the continuity of the utility function with respect to the matching outcome and the compactness of the space of matches, we have that the agent utility is uniformly equicontinuous. Our claim that $(\sigma_C^\infty, \sigma_S^\infty)$ is an ex post (ε, ρ)-Nash equilibrium of the N-agent game for sufficiently large N then follows from Theorem 3 of Bodoh-Creed [2]. Noting from the finiteness of the outcome space that our utility functions are uniformly bounded, we have the $(\sigma_C^\infty, \sigma_S^\infty)$ is an $\varepsilon-$Bayesian-Nash equilibrium of the N-agent game for sufficiently large N from Theorem 2 of Bodoh-Creed [2]. □

An interpretational nuance is required with regard to this theorem. Note that since we assume that π^C and π^S are realized exactly in the nonatomic limit game of section A.1, the corresponding equilibria of the nonatomic limit game where the agents can observe the traits of the other players is actually $\mathcal{E}_\infty(\nu_\infty^C, \nu_\infty^S) \subset \Sigma^C \times \Sigma^S$ where ν_∞^S (ν_∞^C) is the marginal distribution of π^S (π^C) on $\mathcal{T}$.[6] Therefore we need to interpret our result carefully - in the case of observable traits, we need to approximate the finite model with a family of nonatomic models that are differentiated by (ν_N^C, ν_N^S).

We can make the additional claim that the equilibrium strategies in $\mathcal{E}_\infty(\nu_\infty^C, \nu_\infty^S) \subset \Sigma^C \times \Sigma^S$ are (ε, ρ)-Nash equilibrium of sufficiently large finite games. A subtle restriction is required - we need to assume that the strategies in the nonatomic limit game do not condition on $(\nu_\infty^C, \nu_\infty^S)$ since the distributions $(\nu_\infty^C, \nu_\infty^S)$ are known exactly. Denote the restricted strategy space of the limit game as

[6] Note that in the nonatomic limit game there is no loss if the agents do not condition their action on the distribution of observable traits since the distribution is known to be $(\nu_\infty^C, \nu_\infty^S)$ exactly.

$\Sigma_R^C \times \Sigma_R^S$

Proposition B.2. *Choose $\rho \in (0,1], \varepsilon > 0$. Then there exists N^* such that for all $N > N^*$ we have that any $(\sigma_C^\infty, \sigma_S^\infty) \in \mathcal{E}_\infty(\nu_\infty^C, \nu_\infty^S) \cap \Sigma_R^C \times \Sigma_R^S$ is an ex post (ε, ρ)-Nash equilibrium of the N-agent game. Furthermore, $(\sigma_C^\infty, \sigma_S^\infty)$ is an $\varepsilon-$Bayesian-Nash equilibrium of the N-agent game.*

Proof. Since $(\sigma_C^\infty, \sigma_S^\infty)$ falls within the space of strategies of section A.1, proposition A.1 implies that $(\sigma_C^\infty, \sigma_S^\infty)$ ex post (ε, ρ)-Nash equilibrium of the N-agent game where agents make declarations without observing the other agents' traits. Therefore $(\sigma_C^\infty, \sigma_S^\infty)$ is also an ex post (ε, ρ)-Nash equilibrium of the N-agent game where agents make declarations after observing the other agents' traits. □

In order to extend proposition A.2 to this setting, we have to assume that the sequence of equilibrium strategies we consider, $\{(\sigma_C^N, \sigma_S^N)\}_{N=1}^\infty$, is equicontinuous in (ν_N^C, ν_N^S).

Proposition B.3. *Consider a sequence of equicontinuous strategies $\{(\sigma_C^N, \sigma_S^N)\}_{N=1}^\infty$ where (σ_C^N, σ_S^N) is a Bayesian-Nash equilibrium of the N-agent game where agents observe the traits of others before taking actions. Assume that $(\sigma_C^N, \sigma_S^N) \to (\sigma_C^\infty, \sigma_S^\infty)$. Then $(\sigma_C^\infty, \sigma_S^\infty)$ is a $0-$NE the nonatomic limit game.*

Proof. From assumption 1, we know that the set of stable of matchings is continuous with respect to changes in the distribution of declared types. Given the continuity of the utility function with respect to the matching outcome and the compactness of the space of matches, we have that the agent utility is uniformly continuous. By assumption we have that $\{(\sigma_C^N, \sigma_S^N)\}_{N=1}^\infty$ is equicontinuous, and from the compactness of the space of declarations we can strengthen this to uniformly equicontinuity. Our claim then follows from Theorem 3 of Bodoh-Creed [3].[7] □

[7] Bodoh-Creed [2] studies dynamic games, but we can use this result by assuming that agents all have discount factors of 0 (or, alternately, the agent types have 0 utility from any match after the first period). The principal aspect of the analysis we leverage is that strategies are allowed to be conditioned on the distribution of other players' types.

C Online Appendix: Information Gathering and Budget Sets

We can define the outcomes that can be obtained by a student of type $s \in \mathcal{S}$ in the limit model from different preference declarations using "budget sets" defined as

$$\mathcal{B}(s|\widehat{\pi}^C_\infty, \widehat{\pi}^S_\infty) = \{x \in \Delta(\mathcal{C}) : x = x(\cdot, (t_s, \succ'); \widehat{\pi}^C_\infty, \widehat{\pi}^S_\infty) \text{ for some } \succ' \in \mathcal{P}\}$$

Note that the budget set depends on the student's own type, the distribution of agent types in the economy, and the equilibrium strategies. Given the budget set, the problem of a student of type $s \in \mathcal{S}$ can be described as

$$\max_{x \in \mathcal{B}(s|\widehat{\pi}^C_\infty, \widehat{\pi}^S_\infty)} \sum_{c \in \mathcal{C}} x(c) * v(c, s) \tag{C.1}$$

The budget sets can be discovered either by solving for the stable set analytically using our limit model or by examining the outcomes realized by students with the same trait in prior iterations of the market. Our convergence results prove that given the problem is described using budgets sets as per equation C.1, students in a sufficiently large matching market implemented through budget sets have only small incentives to gather additional information about the market.

Presenting the problem in terms of budget sets also makes the problem simple for the students. Strategizing requires a belief regarding the types and strategies of the other agents, a process which may be very involved if done from epistemic principles (Aumann and Brandenburger [1]). The effort expended by parents in the Boston mechanism in order to prepare an optimal preference declaration suggests that strategizing is quite effortful for at least some matching mechanisms (Pathak and Sönmez [7]).[8] By requiring each agent to only understand his or her own preferences, the budget set formula-

[8] Pathak and Sönmez [7] describe the West Zone Parents Group (WZPG), which is a group of parents that met regularly and recommended heuristics to its members for submitting preference rankings to the Boston mechanism.

tion of the problem eliminates this complexity and removes a significant source of both real costs (e.g., gathering information on the strategies of the other agents, analyzing past outcomes of the mechanism) and cognitive costs (e.g., mental effort required to convert information into a well-formed belief). Our proposal to implement the matching mechanism through budget sets shares features in common with Segal [8]. While Segal [8] focuses on using budgets sets to verify an equilibrium, our contribution is to show that an implementation through budgets sets is (approximately) incentive compatible as well.

D Online Appendix: Convergence with Short Preference Lists on Both Sides

In this section we compute convergence rates for an alternative version of the matching model of section 5. The advantage of our alternative model is that we can generate tight bounds on the asymptotic convergence rates. We find convergence is $O(\frac{\sqrt{N}}{T_N})$, significantly slower than the lower bound found in section 5. Although we focus on analyzing the risk faced by a particular type of agent, it is straightforward to construct uniform bounds using our results.

In this section we use the model of section 5, but restrict the colleges to having a single acceptable student-type as a partner. Therefore we can index both the students and colleges using $(i, j) \in \{1, .., T_N\}^2$. An agent of type (i, j) has trait $t_i \in \mathcal{T}_N$ and is only willing to match with a partner that has trait t_j, which is formally described as a preference relation of the form $t_j \succ_{(i,j)} \varnothing \succ_{(i,j)} t_k$ for all $t_k \neq t_i$.

Suppose that for some student type (i, j) we have

$$\pi_E^{S,N}(i, j) - \pi_E^{C,N}(j, i) \geq \delta * \pi_E^{S,N}(i, j)$$

This implies that a fraction δ of students of type i cannot find a partner on the other side of the market. Rewriting this relation we find

$$\pi_E^{S,N}(i, j) - \pi_E^{C,N}(j, i) = \big(\pi^{S,N}(i, j) + \eta^{S,N}\big) - \big(\pi^{C,N}(j, i) + \eta^{U,N}\big) \geq \delta * \big(\pi^{S,N}(i, j) + \eta^{S,N}\big)$$

which gives us

$$(1-\delta) * \eta^{S,N} - \eta^{C,N} \geq \delta * \frac{1}{T_N^2}$$

From the central limit theorem, we know that

$$(1-\delta) * \eta^{S,N} - \eta^{C,N} \xrightarrow{d} N\left(0, \frac{(1-\delta)^2+1}{N} * \frac{1}{T_N^2}\left[1-\frac{1}{T_N^2}\right]\right)$$

Since N (and hence T_N) is large, $\left[1-\frac{1}{T_N^2}\right] \cong 1$.

$$\begin{aligned} \Pr\left\{\left[(1-\delta) * \eta^{S,N} - \eta^{C,N}\right] > \frac{\delta}{T_N^2}\right\} &= 1-\Phi\left(\frac{\delta}{T_N} * \sqrt{\frac{N}{(1-\delta)^2+1}}\right) \\ &= 1-\Phi\left(\frac{\delta}{\sqrt{(1-\delta)^2+1}} * \frac{\sqrt{N}}{T_N}\right) \end{aligned}$$

The asymptotic value of this probability will be driven entirely by whether $\frac{N}{T_N^2} \to \infty$ or $\frac{N}{T_N^2} \to c < \infty$. In the first case, we have that

$$\Pr\left\{\frac{\pi_E^{S,N}(i) - \pi_E^{C,N}(i)}{\pi_E^{S,N}(i)} > \delta\right\} \to 0$$

and asymptotically students of type (i, j) are matched with probability 1. In the later case there is a positive probability that a significant fraction of students of type (i, j) remain unmatched as $N \to \infty$. The convergence rate is demonstrated in figure 4.

References

[1] Aumann, R. and A. Brandenburger (1995) "Epistemic Conditions for Nash Equilibrium," *Econometrica*, 63, pp. 1161 - 1180.

[2] Bodoh-Creed, A. (2011) "Approximation of Large Games with Applications to Uniform Price Auctions," *mimeo.*

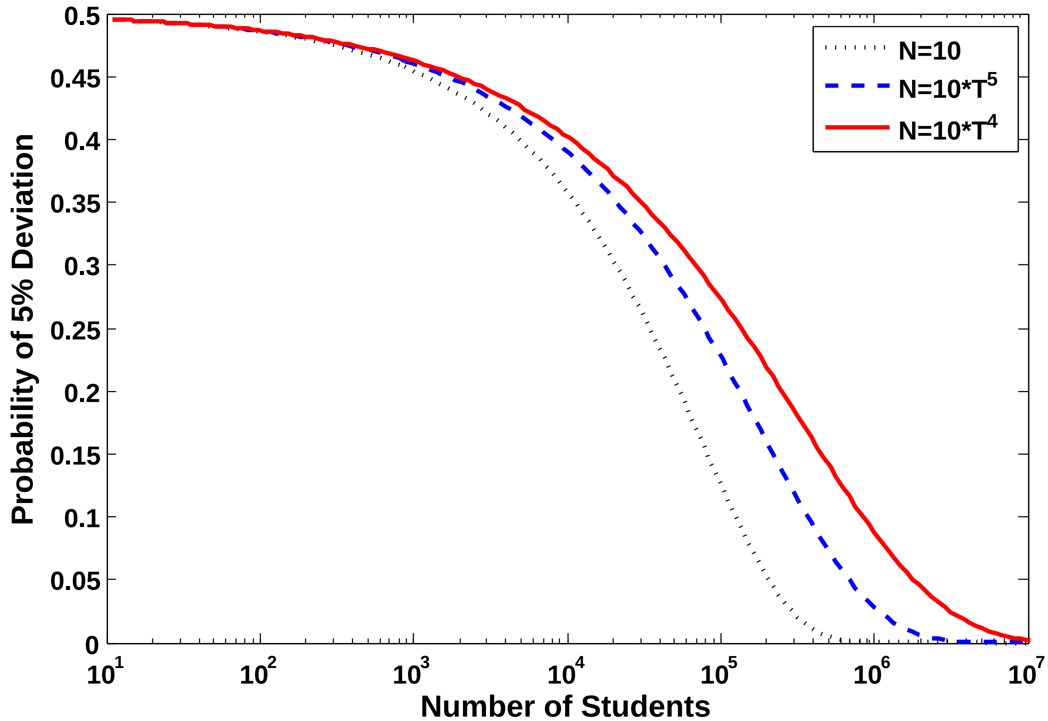

Figure 1: Bound on Large Deviations

[3] Bodoh-Creed, A. (2012) "Approximation of Large Dynamic Games," *mimeo.*

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

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

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

[7] Pathak, P. and T. Sönmez (2008) "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *American Economic Review*, 98, pp. 1636 - 1652.

[8] Segal, I. (2007) "The communication requirements of social choice rules and supporting budget sets," *Journal of Economic Theory,* 136, pp. 341 - 378.