Online Appendix for Demand Analysis using Strategic Reports: An application to a school choice mechanism

Nikhil Agarwal* Paulo Somaini†

October 12, 2017

The appendix follows the organization of the paper. Appendix B describes the data sources and the cleaning process, Appendix C presents results related to the first step estimator, preliminaries for the convergence results and details on RSP+C mechanisms. Appendix D presents technical details relevant for Section 5. Appendix E details the Gibbs' sampler used in Section 6 and the bootstrap.

B Data Appendix

The primary data for the study come from Cambridge Public Schools. Under a non-disclosure agreement, we use data from student registration records, assignment files, and data on student characteristics.

The student registration records contain the school/program the student is registered at, student's grade, language spoken at home and the paid-lunch status at registration.

The assignment files include the rank-order list of the student, sibling or proximity priority at the ranked school, the randomly generated tie-breaker used in the assignment, and the paid-lunch/free-lunch status of the student. Cambridge pre-assigns about 40% of the students to public elementary schools via arrangements with pre-kindergarten schools. The assignment files provide detail on whether the student is pre-assigned and if the student participated in the school choice process (the Cambridge mechanism) studied in this paper.

We also obtained reports from the school district containing the overall capacity of each school/program in each year and the numbers assigned through each process. We use these reports as the primary source for computing the number of seats available at various schools and programs in the mechanism. In rare cases, the rank order lists, the random tie-breaker

*Agarwal: Department of Economics, MIT and NBER, email: agarwaln@mit.edu
†Somaini: Stanford Graduate School of Business and NBER, email: soma@stanford.edu

and the priority codes indicated an inconsistency in the capacity data. We used the knowledge of the mechanism to adjust these capacities and were able to compute the correct assignment for almost all students with these modified capacities.

The student characteristics file duplicates several of the variables in the registration and school choice ranking and assignment file. Importantly, it also includes the home address of the student. The Network Analyst Toolbox in ArcGIS and information in ESRI's Datamaps 10.1 on the US road network was used to compute the distance by road between the student's home and the school address based on brochures from the relevant years. This computation ignores one-way restrictions because Cambridge uses walking distance to compute proximity priority.

These files were merged using a unique student identifier.[1] Schools and programs are also uniquely identified in the dataset.

C Theory: Mechanisms, Convergence and Equilibrium

This section presents results related to the large sample properties of our estimator for $L_{R,t}$ in the class of Report-Specific Priority and Cutoff (RSP+C) Mechanisms. Section C.1 presents examples of RSP+C mechanisms. Section C.2 provides preliminaries for Theorem A.3, which shows consistency and asymptotic normality of our estimator. The result requires that the limit economy has a unique market clearing cutoff. Section C.3 derives conditions under which a market-clearing cutoff exists in an economy and shows that the limit economy (generically) has a unique market-clearing cutoff. Section C.4 shows that equilibrium strategies in a large market approximate equilibria in the limit game.

C.1 Report-Specific Priority and Cutoff Mechanisms

This section formally shows that several school choice mechanisms belong to the class of Report-Specific Priorities + Cutoff (RSP+C) mechanisms. For simplicity, we assume that each school has only one program and that there are no priorities. These examples can be easily modified to accomodate these details.

In the interest of completeness, we start by formally defining the two most commonly used mechanisms, the Student Proposing Deferred Acceptance mechanism, and the Immediate Acceptance mechanism (also known as the Boston mechanism).

The **Student Proposing Deferred Acceptance mechanism**: For reports $R_1, \ldots, R_N$ and priorities $t_1, \ldots, t_N$,

[1] We are grateful to Parag Pathak for sharing the dataset for this project.

Step* 1*: Students apply to their first ranked choice and their applications are *tentatively* held in order of priority and a tie-breaker until the capacity has been reached. Schools reject the remaining students.

Step* k*: Students that are rejected in the previous round apply to their highest ranked choice that has not rejected them. Schools pool new applications with those held from previous steps and *tentatively hold* applications in order of priority and a tie-breaker until the school's capacity has been reached. The remaining students are rejected. The algorithm continues if any rejected student has not been considered at all of her listed schools. Otherwise, each student is assigned to the school that currently holds her application.

This mechanism is strategy-proof for the students if the students can rank all J schools (Dubins and Freedman, 1981; Roth, 1982), but provides strategic incentives for students if students are constrained to list $K < J$ schools (see Abdulkadiroglu et al., 2009; Haeringer and Klijn, 2009, for details).

The **Immediate Acceptance mechanism**: For reports $R_1, \ldots, R_N$ and priorities $t_1, \ldots, t_N$,

Step* 1*: Assign students to their first choice in order of priority and a random tie-breaker until the capacity has been reached. Reject the remaining students.

Step* k*: Assign students that are rejected in the previous round to their k-th choice in order of priority and a random tie-breaker until the capacity has been reached. Schools reject the remaining students. Continue if any rejected student has not been considered at all their listed schools.

This mechanism is a canonical example for one that provides strategic incentives to students (Abdulkadiroglu et al., 2006). Our next result shows that all mechanisms in table I except the TTC is report-specific priority + cutoffs mechanisms.

Proposition C.1. *The Deferred Acceptance mechanism, the Immediate Acceptance mechanism, Serial Dictatorship, First Preferences First, Chinese Parallel Mechanism, the Pan London Admissions scheme and the New Haven Mechanism with tie-breakers are RSP+C mechanisms.*

Proof. We assume that there are no priority types for simplicity, though the proof can be easily rewritten to incorporate finitely many priority types as done for the Cambridge Controlled Choice Plan.

Deferred Acceptance:

We show that Deferred Acceptance is equivalent to a report-specific priority + cutoff mechanisms with

$$e_j = f_j(R_i, \nu_i) = \nu_{ij}.$$

Let $\underline{\nu}_j$ be supremum of the priority scores of the rejected students in school j. We claim that $p^n = \underline{\nu}$ are the cutoffs with the desired properties (if a school does not reject any students, set $p_j^n = 0$).

Let $\underline{\nu}_j^r$ be the supremum the priority scores of students that were rejected in round r. Set $\underline{\nu}_j^r = 0$ if no students are rejected. Observe that for each school, $\underline{\nu}_j^r \leq \underline{\nu}_j^{r+1}$. If the algorithm terminates in round k, then $\underline{\nu}_j^k = \underline{\nu}_j$. Note that the algorithm terminates in finitely many rounds for every n because there are finitely many students and schools and no student applies to the same school twice.

Assume that student i is assigned to school j' and consider any school j with jR_jj'. Let r be round in which student i was rejected by j. By definition, it must be that $\nu_{ij} < \underline{\nu}_j^r$. Therefore, $\nu_{ij} < \underline{\nu}_j$ and we have that each student is assigned to $D^{(R_i,\nu_i)}(p^n)$.

Finally, the aggregate demand cannot exceed q_j by construction of p^n.

Immediate Acceptance mechanism:

We show that the Immediate Acceptance mechanism is report-specific priority + cutoff mechanisms for

$$e_{ij} = f_j(R_i, \nu_i) = \frac{\nu_{ij} + J - 1 - \#\{k : kR_ij\}}{J}$$

by constructing market cutoffs p^n for each profile $((R_1, \nu_1), \ldots, (R_N, \nu_N))$ such that (i) the assignment of each agent is given by $D^{(R_i,\nu_i)}(p^n)$ and (ii) p^n clears the market.

Note that if a school rejects a student in round k, then it rejects students in all further rounds since it is full at the end of round k. Let k_j denote the pivotal round for school j, and let $\underline{\nu}_j$ be supremum of the random priorities of the rejected students in round k_j. We claim that $p_j^n = 1 - \dfrac{k_j - \underline{\nu}_j}{J}$ are the cutoffs with the desired properties (if a school does not reject any students, set $k_j = J$ and $p_j = 0$).

We first show that the assignment of each student in the Immediate Acceptance mechanism is given by $D^{(R_i,\nu_i)}(p^n)$. Assume that student i is assigned to school j' and consider any school j with jR_ij'. Since jR_ij', it must be that the student was rejected at j, and could not have applied to j before round k_j. If student applied to j after round k_j, then $\nu_{ij} - \#\{k : kR_ij\} < \underline{\nu}_j - k_j$ since $|\nu_{ij} - \underline{\nu}_j| \leq 1$. If $\#\{k : kR_ij\} = k_j$, then $\nu_{ij} < \underline{\nu}_j$. In either case, $f_j(R_i, \nu_i) < p_j$. Therefore, the student is assigned to $D^{(R_i,\nu_i)}(p^n)$.

Next, we show that p^n clears the market for economy $((R_1, \nu_1), \ldots, (R_N, \nu_N))$. As noted earlier, each agent is assigned to $D^{(R_i,\nu_i)}(p^n)$. By construction of p^n, the aggregate demand must be less than q_j, and $p_j^n = 0$ if aggregate demand is strictly less than q_j.

Serial Dictatorship:

The Serial Dictatorship mechanism orders the students according to a single priority and then assigns the top student to her top ranked choice. The k-th student is then assigned to her top ranked choice that has remaining seats. It is straightforward to show that this mechanism is equivalent to a Deferred Acceptance mechanism in which all students have identical tie-breakers at all schools. Hence, it is a report-specific priority + cutoff mechanism.

First Preferences First:

The First Preferences First mechanism assigns students to their top ranked choice if seats are available, with tie-breaking according to priorities and a random number. Rejected students are then processed for the remaining seats according to the Deferred Acceptance mechanism. Arguments identical to the ones above show that the First Priority First mechanism is a report-specific priority + cutoff mechanism for

$$e_{ij} = f_j(R_i, \nu_i) = \frac{\nu_{ij} + 1\{jR_ij' \quad \forall j' \neq j\}}{2}.$$

Chinese Parallel (Chen and Kesten, 2013):

The chinese parallel mechanism operates in t rounds, each with t_c-subchoices. In each round, rejected students applies to the next t_c highest choices that have not yet rejected her. Within each round, the algorithm implements a deferred acceptance procedure in which applications are held tentatively until no new proposals are made. Assignments are finalized after all t_c choices have been considered. It is straightforward to show that the Chinese Parallel mechanism is a report-specific priority + cutoff mechanism for

$$f_j(R_i, \nu_i) = \frac{\nu_{ij} - \left\lfloor \dfrac{\#\{k : kR_ij\}}{t_c} \right\rfloor}{\left\lfloor \dfrac{J}{t_c} \right\rfloor} + \frac{\left\lfloor \dfrac{J-1}{t_c} \right\rfloor}{\left\lfloor \dfrac{J}{t_c} \right\rfloor}.$$

Pan London Admissions (Pennell et al., 2006):

The Pan London Admissions system uses the Student Proposing Deferred Acceptance mechanism, except that a subset of schools upgrade the priority of students that rank the school highly. Suppose school j upgrades students that rank it first. For such schools, we set

$$f_j(R_i, \nu_i) = \frac{\nu_{ij} + 1\{jR_ij' \quad \forall j' \neq j\}}{2},$$

and $f_j(R_i, \nu_i) = \nu$ otherwise. With this modification, the Pan London Admissions scheme is a report-specific priority + cutoff mechanism.

We use $e_{ij} = f_j(R_i, \nu_i) = \nu_{ij}$ for schools that do not modify the priority and $e_{ij} = f_j(R_i, \nu_i) = \dfrac{\nu_{ij} - \#\{k : kR_ij\}}{J} + \dfrac{J-1}{J}$ for school that use the Immediate Acceptance rule.

New Haven Mechanism:

See Kapor et al. (2017) for description and proof.

□

C.2 Preliminaries for the proof of Theorem A.3

Lemma C.1. *Suppose that the tie-breaker ν is non-degenerate. Then, (i) for each $j \in \mathcal{J}$, $\sup_p |D_j(p|\eta) - D_j(p|\eta^n)|$ and $\sup_p |D_j(p|\eta) - D_j(p|\eta_b^{n-1})|$ converge in probability to 0.*

(ii) for any p^*, we have that

$$\sqrt{n}\left(\frac{1}{B}\sum_b D(p^*|\eta_b^{n-1}) - D(p^*|\eta)\right) \overset{d}{\to} \mathcal{N}(0, \Omega)$$

where

$$\Omega = \left(1 + \frac{1}{B}\right) V\left(\int D^{(R,t,\nu)}(p^*) \mathrm{d}\gamma_\nu\right) + \frac{E\left[V\left(D^{(R,t,\nu)}(p^*)\middle| R, t\right)\right]}{B}.$$

(iii) For any p^* and any sequence of δ_n decreasing to 0,

$$\sup_{\|p - p^*\| \leq \delta_n} \sqrt{n}\|D(p|\eta^n) - D(p|\eta) + D(p^*|\eta) - D(p^*|\eta^n)\| = o_p(1).$$

Likewise,

$$\sup_{\|p - p^*\| \leq \delta_n} \sqrt{n}\|D(p|\eta_b^{n-1}) - D(p|\eta^n) + D(p^*|\eta^n) - D(p^*|\eta_b^{n-1})\| = o_p(1).$$

Proof. Part (i): Let v_{pj} be the set of tuples of priority types, random tie-breakers and rank order lists, (R_i, t_i, ν_i), that are assigned to programs j under cutoffs p. This set can be written as:

$$v_{pj} = \{(R_i, t_i, \nu_i) : f_j(R_i, t_i, \nu_{ij}) \geq p_j,\ jR_i0; \forall j' R_i j, f_{j'}(R_i, t_i, \nu_{ij'}) < p_{j'}\}.$$

Let $\mathcal{V} = \{v_{pj} : p, j\}$ be the class of sets v_{pj} indexed by p and j.

Since f in increasing in the last argument, for each j, R_i, t_i, the class of sets $\{\{\nu_i : f_j(R_i, t_i, \nu_{ij}) \geq p_j\} : p_j\}$ is a Vapnik-Chervonenkis (VC) class. Hence, the class $\mathcal{B} = \{\{\nu_i : f_j(R_i, t_i, \nu_{ij}) \geq p_j\} : p_j, j, R, t\}$ is a VC class because (j, R, t) belong to a finite set. Hence, $\mathcal{V}$ is a VC-class since it is a subset of finite unions and intersections of sets in $\mathcal{B}$ and their complements. Therefore, $\mathcal{V}$ is a uniform Glivenko-Cantelli class. Part (i) follows from the Glivenko-Cantelli Theorem.

Part (ii): We first re-write

$$
\begin{aligned}
& \frac{1}{B}\sum_b D(p^*|\eta_b^{n-1}) - D(p^*|\eta) \\
= \; & \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{i_b} D^{(R_{i_b}, t_{i_b}, \nu_{i_b})}(p^*) - D(p^*|\eta) \\
= \; & \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{i_b} D^{(R_{i_b}, t_{i_b}, \nu_{i_b})}(p^*) - \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i, t_i, \nu)}(p^*)\mathrm{d}\gamma_\nu \\
& + \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i, t_i, \nu)}(p^*)\mathrm{d}\gamma_\nu - D(p^*|\eta).
\end{aligned}
$$

We now derive the distribution of

$$
\mathbb{G}_{n,b} = \sqrt{n}\left(\frac{1}{n}\sum_{i_b} D^{(R_{i_b}, t_{i_b}, \nu_{i_b})}(p^*) - \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i, t_i, \nu)}(p^*)\mathrm{d}\gamma_\nu\right)
$$

conditional on the sample $(R_1, t_1), \ldots, (R_n, t_n)$, and fixed b. To do this, we adapt the proof for the bootstrap distribution of the sample mean (Theorem 23.4, van der Vaart, 2000).

Note that

$$
\begin{aligned}
E\left[D^{(R_{i_b}, t_{i_b}, \nu_{i_b})}(p^*)\middle| (R_1, t_1), \ldots, (R_n, t_n)\right] &= E\left[E\left[D^{(R_{i_b}, t_{i_b}, \nu)}(p^*)\middle| R_{i_b}, t_{i_b}\right]\middle| (R_1, t_1), \ldots, (R_n, t_n)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} E[D^{(R_i, t_i, \nu)}(p^*)|R_i, t_i] \\
&= \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i, t_i, \nu)}(p^*)\mathrm{d}\gamma_\nu.
\end{aligned}
$$

By the law of total variance, the conditional variance of $D^{(R_{i_b}, t_{i_b}, \nu_{i_b})}(p^*)$ given $(R_1, t_1), \ldots, (R_n, t_n)$

is

$$
\begin{aligned}
& E\left[V\left(D^{(R_{i_b},t_{i_b},\nu_{i_b})}(p^*)\middle|\, R_{i_b},t_{i_b}\right)\middle|\,(R_1,t_1),\ldots,(R_n,t_n)\right] \\
& +V\left[E\left(D^{(R_{i_b},t_{i_b},\nu_{i_b})}(p^*)\middle|\, R_{i_b},t_{i_b}\right)\middle|\,(R_1,t_1),\ldots,(R_n,t_n)\right] \\
= \; & \frac{1}{n}\sum_{i=1}^{n} V\left(D^{(R_i,t_i,\nu_i)}(p^*)\middle|\, R_i,t_i\right) + V\left(\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu\middle|\,(R_1,t_1),\ldots,(R_n,t_n)\right),
\end{aligned}
$$

where $V\left(\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu\middle|\,(R_1,t_1),\ldots,(R_n,t_n)\right)$ is the sample variance of $\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu$. Since D is uniformly bounded, the variance above is bounded. By the strong law of large numbers, the conditional variance of $D^{(R_{i_b},t_{i_b},\nu_{i_b})}(p^*)$ converges to

$$
\tilde{\Omega} = E\left[V\left(D^{(R_i,t_i,\nu_i)}(p^*)\right)\right] + V\left(\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu\right)
$$

almost surely for sequences $(R_1,t_1),(R_2,t_2),\ldots$.

Note that since $D^{(R_{i_b},t_{i_b},\nu_{i_b})}$ is uniformly bounded, we have that for every $\varepsilon > 0$,

$$
E\left[\|D^{(R_{i_b},t_{i_b},\nu_{i_b})}\|^2 1\{\|D^{(R_{i_b},t_{i_b},\nu_{i_b})}\| > \varepsilon\sqrt{n}\}\right] \to 0.
$$

Therefore, by the Lindeberg-Feller central limit theorem (Theorem 2.27, van der Vaart, 2000), conditionally on $(R_1,t_1),\ldots,(R_n,t_n)$, for almost every sequence $(R_1,t_1),(R_2,t_2),\ldots$, $\mathbb{G}_{n,b} \xrightarrow{d} \mathcal{N}(0,\tilde{\Omega})$. An identical argument shows that $\frac{1}{B}\sum_b \mathbb{G}_{n,b} \xrightarrow{d} N\left(0,\frac{1}{B}\tilde{\Omega}\right)$ conditionally on $(R_1,t_1),\ldots,(R_n,t_n)$, for almost every sequence $(R_1,t_1),(R_2,t_2),\ldots$, since i_b is independent of $i_{b'}$ conditional on $(R_1,t_1),\ldots,(R_n,t_n)$ for all $b \neq b'$. Therefore, we have that conditionally on $(R_1,t_1),\ldots,(R_n,t_n)$, for almost every sequence $(R_1,t_1),(R_2,t_2),\ldots$,

$$
\sqrt{n}\left(\frac{1}{B}\sum_{b=1}^{B} D(p^*|\eta_b^{n-1}) - \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu\right) \xrightarrow{d} N\left(0,\frac{1}{B}\tilde{\Omega}\right).
$$

Now consider the stacked random vector

$$
\sqrt{n}\begin{pmatrix} \frac{1}{B}\sum_{b=1}^{B} D(p^*|\eta_b^{n-1}) - \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu \\ \frac{1}{n}\sum_{i=1}^{n}\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu - D(p^*|\eta) \end{pmatrix}. \tag{C.1}
$$

Conditional on $(R_1,t_1),\ldots,(R_n,t_n)$, the second element is deterministic and the first element converges in distribution to $Z_1 \sim N\left(0,\frac{1}{B}\tilde{\Omega}\right)$ for almost every sequence $(R_1,t_1),(R_2,t_2),\ldots$.

By the central limit theorem, the second element converges in distribution to

$$Z_2 \sim N\left(0, V\left(\int D^{(R_i,t_i,\nu)}(p^*)\mathrm{d}\gamma_\nu\right)\right).$$

Since Z_1 is (almost surely) independent of $(R_1, t_1), \ldots, (R_n, t_n)$, we have that the stacked random vector in expression (C.1) converges in distribution to (Z_1, Z_2) where Z_1 and Z_2 are independent. Hence,

$$\sqrt{n}\left(\frac{1}{B}\sum_{b=1}^{B} D(p^*|\eta_b^{n-1}) - D(p^*|\eta)\right) \xrightarrow{d} \mathcal{N}(0, \Omega).$$

Part (iii): Note that

$$\begin{aligned} & \sqrt{n}\|D(p|\eta^n) - D(p|\eta) + D(p^*|\eta) - D(p^*|\eta^n)\| \\ \leq \quad & J|\sqrt{n}\left(\eta^n(\upsilon_{p\wedge p^*,p\vee p^*}) - \eta(\upsilon_{p\wedge p^*,p\vee p^*})\right)|, \end{aligned}$$

where $\upsilon_{p,p'} = \{\nu : p \leq f(R,T,\nu) \leq p'\}$. We now bound the variance of the right-hand side. For any p, p' with $p \leq p'$,

$$\begin{aligned} V(\eta^n(\upsilon_{p,p'}) - \eta(\upsilon_{p,p'})) &= V\left(\frac{1}{n}\sum_i 1\{f(R_i,T_i,\nu_i) \in \upsilon_{p,p'}\} - \eta(\upsilon_{p,p'})\right) \\ &= \frac{1}{n}\eta(\upsilon_{p,p'})(1-\eta(\upsilon_{p,p'})). \end{aligned}$$

Therefore, $V(J|\sqrt{n}\left(\eta^n(\upsilon_{p\wedge p^*,p\vee p^*}) - \eta(\upsilon_{p\wedge p^*,p\vee p^*})\right)|)$ is at most $J\eta(\upsilon_{p\wedge p^*,p\vee p^*})$. By Chebychev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\left(J|\sqrt{n}\left(\eta^n(\upsilon_{p\wedge p^*,p\vee p^*}) - \eta(\upsilon_{p\wedge p^*,p\vee p^*})\right)| > \varepsilon\right) \leq \frac{J^2\eta(\upsilon_{p\wedge p^*,p\vee p^*})^2}{\varepsilon^2}.$$

Since $\eta(\upsilon_{p\wedge p^*,p\vee p^*}) \leq \kappa\|p\wedge p^* - p\vee p^*\|_\infty$, we therefore have that for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{\|p-p^*\|\leq\delta_n}\sqrt{n}\|D(p|\eta^n) - D(p|\eta) + D(p^*|\eta) - D(p^*|\eta^n)\| > \varepsilon\right) \leq \frac{\kappa^2\delta_n^2 J^2}{\varepsilon^2}.$$

Hence, for any sequence of δ_n decreasing to zero, we have that

$$\sup_{\|p-p^*\|\leq\delta_n}\sqrt{n}\|D(p|\eta^n) - D(p|\eta) + D(p^*|\eta) - D(p^*|\eta)\| = o_p(1).$$

By a similar argument, we have that

$$\mathbb{P}\left(\sup_{\|p-p*\|<\delta_n}\sqrt{n}\|D(p|\eta_b^{n-1})-D(p|\eta^n)+D(p^*|\eta^n)-D(p^*|\eta_b^{n-1})\|>\varepsilon\right)<\frac{J^2V(\eta_b^{n-1}(v_{p,p'})-\eta^n(v_{p,p'}))}{\varepsilon^2}.$$

Since $E[\eta_b^{n-1}(v_{p,p'})|\eta^n]=\eta^n(v_{p,p'})$, by the law of total variance,

$$\begin{aligned}V(\eta_b^{n-1}(v_{p,p'})-\eta^n(v_{p,p'})) &= E\left[V(\eta_b^{n-1}(v_{p,p'})-\eta^n(v_{p,p'})|\eta^n)\right]\\ &= E\left[\eta^n(v_{p,p'})(1-\eta^n(v_{p,p'}))\right]\\ &\le E\left[\eta^n(v_{p,p'})\right]=\eta(v_{p,p'}).\end{aligned}$$

Hence, we have that

$$\mathbb{P}\left(\sup_{\|p-p*\|<\delta_n}\sqrt{n}\|D(p|\eta_b^{n-1})-D(p|\eta^n)+D(p^*|\eta^n)-D(p^*|\eta_b^{n-1})\|>\varepsilon\right)<\frac{k^2J^2\delta_n^2}{\varepsilon^2}.$$

□

Lemma C.2. *Suppose there is a unique p^* such that for all $k\in\mathcal{J}\cup\mathcal{S}$, $D_k(p^*|\eta)-q_k\le 0$ with equality if $p_k^*>0$. Also assume that there exists p^n such that $D_k(p^n|\eta^n)-q_k^n\le 0$ with equality if $p_k^n>0$. and likewise assume that there exists p_b^{n-1} such that $D_k(p_b^{n-1}|\eta_b^{n-1})-q_k^n\le 0$ with equality if $p_{b,k}^{n-1}>0$.*

1. *If (i) $|D(p|\eta_b^{n-1})-D(p|\eta)|\xrightarrow{p}0$ and $|D(p|\eta^n)-D(p|\eta)|\xrightarrow{p}0$ uniformly in p, (ii) $q^n\to q$, (iii) $D(p|\eta)$ is continuous in p, then $\sup_{j\in\mathcal{J}}|p_{b,j}^{n-1}-p_j^*|\xrightarrow{p}0$ and $\sup_{j\in\mathcal{J}}|p_j^n-p_j^*|\xrightarrow{p}0$.*
2. *Further, if the hypotheses of part 1 hold, (iv) $E[D(p^*|\eta^n)]=D(p^*|\eta)$, (v) for any p^**

$$\sqrt{n}\left(\frac{1}{B}\sum_b D(p^*|\eta_b^{n-1})-D(p^*|\eta)\right)\ \xrightarrow{d}\ Z$$

   (vi) For any p^* and any sequence of δ_n decreasing to 0,

$$\sup_{\|p-p^*\|\le\delta_n}\sqrt{n}\|D(p|\eta_b^{n-1})-D(p|\eta)+D(p^*|\eta)-D(p^*|\eta_b^{n-1})\|=o_p(1).$$

   (vii) $\nabla_{p_+^*}D_+(p^*|\eta)$ exists and is invertible at p^*, and (viii) $q^n-q=o_p(n^{-1/2})$, then

$$\sqrt{n}\left(\frac{1}{B}\sum_b p_b^{n-1}-E[p^n]\right)\xrightarrow{d}\nabla DZ$$

$$\textit{where } \nabla D = \begin{bmatrix} (\nabla_{p_+^*} D_+(p^*|\eta))^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof. **Part 1:** The result is similar in spirit to Azevedo and Leshno (2016), theorem 2, though the techniques are different and generalized to mechanisms.

We only show the result for p^n since the argument for p_b^{n-1} is identical. Let

$$Q_n(p) = \left\| \begin{bmatrix} \max\left\{ z\left(p \,|\eta^n, q^n\right), 0 \right\} \\ p * z\left(p \,|\eta^n, q^n\right) \end{bmatrix} \right\|,$$

where $*$ represents the Hadamard product and $z(p|\eta, q) = D(p|\eta, q) - q$. Note that p^n solves $Q_n(p) = 0$. Let Q_0 be the limiting objective function,

$$Q_0(p) = \left\| \begin{bmatrix} \max\left\{ z(p|\eta, q), 0 \right\} \\ p * z(p|\eta, q) \end{bmatrix} \right\|.$$

By the continuous mapping theorem, $\sup_p |Q_n(p) - Q_0(p)| \xrightarrow{p} 0$. Also, $Q_0(p)$ is continuous since $D(p|\eta)$ is continuous. Further, $Q_0(p)$ is uniquely minimized at p^*. For $\varepsilon > 0$, let $\delta_\varepsilon = \inf_{p:\|p-p^*\|>\varepsilon} Q_0(p)$. Since Q_0 is continuous, p is an element of a compact space and $Q_0(p) = 0$ only at p^*, $\delta_\varepsilon > 0$. Pick N such that for all $n > N$, $\mathbb{P}(\sup_p |Q_0(p) - Q_n(p)| > \delta_\varepsilon) < \varepsilon$. For p^n, we have that $Q_n(p^n) = 0$. Note that

$$\begin{aligned} & |Q_0(p^n) - Q_0(p^*)| \\ \leq \; & |Q_0(p^n) - Q_n(p^n)| + |Q_n(p^n) - Q_0(p^*)| \\ \leq \; & \sup_p |Q_0(p) - Q_n(p)| + 0. \end{aligned} \tag{C.2}$$

Hence, we have that for all $n > N$,

$$\begin{aligned} \mathbb{P}\left(\sup_{k \in \mathcal{J} \cup \mathcal{S}} |p_k^n - p_k^*| > \varepsilon \right) & \leq \mathbb{P}\left(|Q_0(p^n) - Q_0(p^*)| > \delta_\varepsilon \right) \\ & \leq \mathbb{P}\left(\sup_p |Q_0(p) - Q_n(p)| > \delta_\varepsilon \right) < \varepsilon \end{aligned}$$

where the first inequality follows from set inclusion, the second from equation (C.2), and the third by our choice of N.

Part 2: We can re-write

$$\sqrt{n}\left(\frac{1}{B} \sum_b p_b^{n-1} - E[p^n] \right) = \sqrt{n}\left(\frac{1}{B} \sum_b p_b^{n-1} - p^* \right) + \sqrt{n}(p^* - E[p^n]).$$

We first derive the limit distribution of $\sqrt{n}\left(\frac{1}{B}\sum_b p_b^{n-1} - p^*\right)$.

Let K^0 be the set of k such that $p_k^* = 0$, i.e. $D_k(p^*|\eta) < q_k$, and let $\delta = \min_{k\in K^0}\{q_k - D_k(p^*|\eta)\}$. Since $D_k(p|\eta)$ is continuous, there exists $\kappa > 0$ such that for all $\|p - p^*\| < \kappa$ and all $k \in K^0$, we have that $D_k(p|\eta) - q_k < -\frac{\delta}{3}$. For any $\varepsilon > 0$, pick N such that for all $n > N$, $\mathbb{P}(\|p_b^{n-1} - p^*\| < \kappa) < \varepsilon$ and $\|q_k^n - q_k\| < \frac{\delta}{3}$. Such an N exists since $p_b^{n-1} \xrightarrow{p} p^*$ and $q_k^n \to q_k$. For all $n > N$, we have that

$$\begin{aligned}
D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k^n \quad & < D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k + \frac{\delta}{3} \\
& < |D_k(p_b^{n-1}|\eta_b^{n-1}) - D_k(p^*|\eta)| + D_k(p^*|\eta) - q_k + \frac{\delta}{3} \\
& < |D_k(p_b^{n-1}|\eta_b^{n-1}) - D_k(p^*|\eta)| - \frac{2\delta}{3} \\
\implies \quad \mathbb{P}\left(D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k^n > -\frac{\delta}{3}\right) \quad & < \mathbb{P}\left(|D_k(p_b^{n-1}|\eta_b^{n-1}) - D_k(p^*|\eta)| > \frac{\delta}{3}\right) \\
& < \mathbb{P}\left(\|p_b^{n-1} - p^*\| > \kappa\right) < \varepsilon
\end{aligned}$$

where the second last inequality follows from set inclusion and the choice of κ. Since $p_b^{n-1} = 0$ if $D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k^n < 0$, we have that for all $n > N$, $\mathbb{P}(p_{b,k}^{n-1} > 0) < \varepsilon$. Therefore, $\sqrt{n}|p_{b,k}^{n-1} - p_k^*| \xrightarrow{p} 0$ for all $k \in K^0$.

The limit distribution of $\sqrt{n}(p_{b,+}^{n-1} - p_+^*)$ is a consequence of the Delta Method. For simplicity of notation, we omit the subscript $+$ and treat $p_k^n = 0$ if $p_k^* = 0$ since $p_k^n = o_p(n^{-1/2})$.

Note that for all $k \notin K^0$, we have that $D_k(p^*|\eta) - q_k = 0$. Let $\delta = \min_{k\notin K^0} p_k^*$. Since $\|p_b^{n-1} - p^*\| \xrightarrow{p} 0$, we have that for any $\varepsilon > 0$, there exists N such that for all $n > N$, $\mathbb{P}(p_{b,k}^{n-1} = 0 \text{ for any } k \notin K^0) < \varepsilon$. Since $p_{b,k}^{n-1} > 0$ implies that $D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k^n = 0$, for all $n > N$, p_b^{n-1} solves $0 = D_k(p|\eta_b^{n-1}) - q_k^n$ with probability at least $1 - \varepsilon$. Therefore, $D_k(p_b^{n-1}|\eta_b^{n-1}) - q_k^n = o_p(n^{-1/2})$ for all $k \notin K^0$.

Since $\|p_b^{n-1} - p^*\| \xrightarrow{p} 0$, condition (v) implies that there exists a sequence of δ_n decreasing to 0, such that

$$D(p_b^{n-1}|\eta_b^{n-1}) - D(p_b^{n-1}|\eta) + D(p^*|\eta) - D(p^*|\eta_b^{n-1}) = o_p(n^{-1/2}).$$

Together with $D(p_b^{n-1}|\eta_b^{n-1}) - q^n = o_p(n^{-1/2})$, condition (v) implies that

$$q - q^n + D(p^*|\eta_b^{n-1}) - q + D(p_b^{n-1}|\eta) - D(p^*|\eta) \quad = o_p(n^{-1/2}).$$

Since $\|q - q^n\| = o_p(n^{-1/2})$, and $D(p^*|\eta) = q$, we have that

$$\begin{aligned} & D(p^*|\eta_b^{n-1}) - D(p^*|\eta) + D(p_b^{n-1}|\eta) - D(p^*|\eta) && = o_p(n^{-1/2}) \\ \implies \quad & \sqrt{n}(D(p^*|\eta_b^{n-1}) - D(p^*|\eta)) + \nabla_{p^*} D(p^*|\eta)\sqrt{n}(p_b^{n-1} - p^*) + o_p(\|p_b^{n-1} - p^*\|) && = o_p(1), \end{aligned}$$

where the implication results form the Delta Method. Since, $o_p(\|p_b^{n-1} - p^*\|) = o_p(1)$, and $\nabla_{p^*} D(p^*|\eta)$ is invertible, we have that

$$\sqrt{n}(p_b^{n-1} - p^*) = \sqrt{n}(\nabla_{p^*} D(p^*|\eta))^{-1}(D(p^*|\eta_b^{n-1}) - D(p^*|\eta)) + o_p(1).$$

Since $E[D(p^*|\eta^n)] = D(p^*|\eta)$, by a similar argument,

$$\sqrt{n}(E[p^n] - p^*) = \sqrt{n}(\nabla_{p^*} D(p^*|\eta))^{-1}(E[D(p^*|\eta^n)] - D(p^*|\eta)) + o_p(1) = o_p(1).$$

Therefore,

$$\begin{aligned} & \sqrt{n}\left(\frac{1}{B}\sum_b p_b^{n-1} - E[p^n]\right) \\ = \quad & \sqrt{n}\left(\frac{1}{B}\sum_b p_b^{n-1} - p^*\right) + o_p(1) \\ = \quad & \sqrt{n}(\nabla_{p^*} D(p^*|\eta))^{-1}\left(\frac{1}{B}\sum_b D(p^*|\eta_b^{n-1}) - D(p^*|\eta)\right) + o_p(1) \end{aligned}$$

By condition (vi) and Slutsky's theorem, we have that

$$\sqrt{n}\left(\frac{1}{B}\sum\nolimits_b p_b^{n-1} - E[p^n]\right) \xrightarrow{d} \nabla DZ.$$

□

C.3 Existence and (Generic) Uniqueness of Cutoffs

This section shows that the cutoffs for RSP+C mechanisms have (generically) unique cutoffs. The main results are Propositions C.2 and C.4. The former provides a general high level condition for (generic) uniqueness in RSP+C mechanisms and the latter provides a weaker condition for the Cambridge mechanism. To do so, we first need to introduce some notation and definitions.

Definition C.1. *The function $D : [0,1]^J \to [0,1]^J$ satisfies **weak-substitutes** if $D_j(p)$ is* *non-increasing in p_j and non-decreasing in $p_{j'}$, where $p \in [0,1]^J$.*

The next definition is a stricter notion of substitutes in a neighborhood around a given cutoff. This borrows from the notion of connected substitutes introduced in Berry et al. (2013) and Berry and Haile (2010) to show conditions when demand is invertible.

Definition C.2. *The function $D : [0,1]^J \to [0,1]^J$ satisfies **local connected substitutes** at p^* if there exists an $\varepsilon > 0$, such that for all $p \in [0,1]^J$ with $\|p - p^*\| < \varepsilon$, we have that*

1. *for all $j \in \{0, 1, \ldots, J\}$ and $k \in \{1, \ldots, J\} \backslash \{j\}$, $D_j(p)$ is nondecreasing in p_k*
2. *for all non-empty subsets $K \subset \{1, \ldots, J\}$, there exists $k \in K$ and $l \notin K$ such that $D_l(p)$ is strictly increasing in p_k*

Local connected substitutes is implied by strict gross substitutes, and the condition that $D(p|\eta)$ as defined in equation (11) satisfies local connected substitutes for all $p \in [0,1]$ is testable.

Definition C.3 (Azevedo and Leshno (2016))**.** *The function $D : [0,1]^J \to [0,1]^J$ is **regular** if the image $D(\bar{P})$, where*

$$\bar{P} = \{p \in [0,1]^J : D(p) \text{ is not continuously differentiable at } p\}$$

has Lebesgue measure 0.

For a fixed $q \in [0,1]^J$, let $p^* \in [0,1]^J$ be a solution to the problem

$$D(p) - q \leq 0 \text{ and } p * (D(p) - q) = 0, \tag{C.3}$$

where $*$ is the Hadamard product. We now observe that (generically for $q \in [0,1]^J$) there exists a unique solution to equation (C.3) if D satisfies local connected substitutes at any market clearing cutoff (is regular).

Proposition C.2. *Let $D(\cdot|\eta)$ be defined as in equation (11). If $D(\cdot|\eta)$ satisfies weak substitutes, then there exists a solution to equation (C.3) for all q.*

Further, for a fixed $D(\cdot|\eta)$, let $Q \subset [0,1]^J$ be the set of capacities, q, such that there are multiple solutions to equation (C.3).

1. *$Q \cap \{q : \sum_{j=1}^J q_j < \sum_j D(0|\eta)\}$ has Lebesgue measure zero if $D_j(\cdot|\eta)$ is regular*
2. *Q is empty if $D(\cdot|\eta)$ satisfies local connected substitutes at any solution p^* to equation (C.3). In particular, Q is empty if $D(\cdot|\eta)$ satisfies local connected substitutes at every cutoff p.*

Proof. Existence of cutoffs that solve equation (C.3) follows from corollary A1 and lemma 1 of Azevedo and Leshno (2016). Statement 1 is a consequence of Azevedo and Leshno (2016), theorem 1(2) and lemma 1. Statement 2 is a strengthening of Azevedo and Leshno (2016), theorem 1(1). By the Lattice Theorem (Azevedo and Leshno, 2016), there exist minimum and maximum cutoffs $p^- \leq p^+$ that solve equation (C.3). By the Rural Hospitals Theorem (Azevedo and Leshno, 2016), for all $C \subseteq S$,

$$\sum_{j \in C} D_j(p^+|\eta) = \sum_{j \in C} D_j(p^-|\eta). \tag{C.4}$$

Let p^* be a solution to equation (C.3) such that $D(\cdot|\eta)$ satisfies local connected substitutes at p^*. Let $C^+ = \{j \in S : p_j^* < p_j^+\}$ and $C^- = \{j \in S : p_j^* > p_j^-\}$. We will show that $C^+ = \emptyset$ i.e. $p^+ = p^*$. The proof to show that $C^- = \emptyset$ is symmetric and together, these claims imply that $p^+ = p^- = p^*$.

Towards a contradiction, assume that $C^+ \neq \emptyset$. Since $D(p|\eta)$ satisfies local connected substitutes at p^* (Definition C.2), there exist $\varepsilon \in (0, 1)$, $k \in C^+$, and $l \notin C^+$ such that

$$D_l(p^*|\eta) < D_l(p^\varepsilon|\eta),$$

where $p_k^\varepsilon = \varepsilon p_k^+ + (1-\varepsilon)p_k^*$ and $p_j^\varepsilon = p_j^*$ for $j \neq k$. Hence, we have that

$$\sum_{j \in S \setminus C^+} D_j(p^*|\eta) < \sum_{j \in S \setminus C^+} D_j(p^\varepsilon|\eta) \leq \sum_{j \in S \setminus C^+} D_j(p^+|\eta),$$

where the implication on the summation and the second inequality are implied by weak substitutes, which follows from the definition of $D(p|\eta)$. Since this inequality contradicts equation (C.4), it must be that $C^+ = \emptyset$. □

As shown in Proposition A.2, p^* is a market clearing cutoff for $D(p|\eta)$ and q if and only if $\tilde{p}^*$ solves equation (C.3), where $p^* = \tilde{A}\tilde{p}^*$. Below, we state uniqueness of a market clearing cutoff in terms of the uniqueness of $\tilde{p}^*$.

Proposition C.3. *Let $\tilde{D}(\tilde{p}|\eta)$ be defined as in equation (12), and for each $\tilde{p}_{\mathcal{S}}$, define $\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}})$ such that $D_j(\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}}|\eta) - q_j \leq 0$ with equality if $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}_{\mathcal{S}}) > 0$.*

If $D(p|\eta)$ is continuous in p and satisfies weak substitutes, then for each $q \in [0,1]^{J+S}$, there exists a $\tilde{p}$ that solves the problem in equation (C.3) for $\tilde{D}(\tilde{p}|\eta)$ and q.

Further, if $D^*(\tilde{p}_{\mathcal{S}}|\eta) = A'D(\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}}|\eta)$ and $D(p|\eta)$ satisfy local connected substitutes at $\tilde{p}^*_{\mathcal{S},s} = \min\{p_j^* : s_j = s\}$ and p^* respectively for some market clearing cutoff, then p^* is unique.

Proof. We first show existence. Since $D(\cdot|\eta)$ satisfies weak substitutes, for each $\tilde{p}_{\mathcal{S}}$, $\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}})$ exists. Lemma C.3 below shows that $D^*(\tilde{p}_{\mathcal{S}}|\eta)$ satisfies weak substitutes. Therefore, by Proposition C.2, there exists $\tilde{p}^*_{\mathcal{S}}$ such that $D^*_s(\tilde{p}^*_{\mathcal{S}}|\eta) - q_s \le 0$ with strict equality if $\tilde{p}^*_{\mathcal{S},s} > 0$. Hence, for $\tilde{p}^* = (\tilde{p}^{*\prime}_{\mathcal{J}}, \tilde{p}^{*\prime}_{\mathcal{S}})'$ and $q \in [0,1]^{J+S}$, and for all $k \in \mathcal{J} \cup \mathcal{S}$, $\tilde{D}_k(\tilde{p}^*|\eta) - q_k \le 0$ with strict equality if $\tilde{p}^*_k > 0$.

To show uniqueness, note that $D(\tilde{p}_{\mathcal{J}} + A\tilde{p}^*_{\mathcal{S}}|\eta)$ satisfies local connected substitutes at $\tilde{p}^*_{\mathcal{J}}$. By Proposition C.2, we have that $D(\tilde{p}_{\mathcal{J}} + A\tilde{p}_{\mathcal{S}}|\eta)$ admits a unique solution $\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}})$ in a neighborhood of $\tilde{p}^*_{\mathcal{S}}$. Further, since $D^*(\tilde{p}_{\mathcal{S}}|\eta)$ satisfies local connected substitutes at $\tilde{p}^*_{\mathcal{S}}$, Proposition C.2 implies that $\tilde{p}^*_{\mathcal{S}}$ is unique. □

We now verify that if

$$f_j(R_i, t_i, \nu_i) = \frac{3 - R_i(j) + \dfrac{t_{ij} + \nu_i}{4}}{3} \tag{C.5}$$

for $\nu_i \in [0,1]$ as in the Cambridge mechanism, then the market clearing cutoff p^* is unique if

$$D_j(p) = E\left[1\{f_j(R_i, t_i, \nu_i) > p_j, jR_i0\} \prod_{j' \neq j} 1\{jR_ij' \text{ or } f_{j'}(R_i, t_i, \nu_i) \le p_{j'}\}\right] \tag{C.6}$$

is strictly decreasing in p_j in a neighborhood around any market-clearing cutoff p^*.

Proposition C.4. *Let f and $D(p)$ be defined as in equations (C.5) and (C.6). If for every program $j \in 1, ..., J$, $D_j(p)$ is strictly decreasing in p_j in a neighborhood of p^*, then the market clearing cutoff p^* is unique. Moreover, if for every program $j \in 1, ..., J$, $D_j(p)$ is differentiable at p^*, then $\nabla_{p_+} D_+(p^*)$ is nonsingular.*

Proof. Fix any market clearing cutoff p^*. For each j, let $r^*_j \in \{1, 2, 3, 4\}$ be the pivotal rank for program j, i.e. $f_j(R_i, t_i, \nu_i) > p^*_j$ if $R_i(j) < r^*_j$ and $f_j(R_i, t_i, \nu_i) < p^*_j$ if $R_i(j) > r^*_j$. We use the convention that $r^*_j = 4$ if the program cutoff is 0, and $r^*_0 = 5$ for the outside option.

For $\varepsilon > 0$, define $p^\varepsilon_k = p^*_k$ if $k \neq j$ and $p^\varepsilon_j = p^*_j + \varepsilon$. By the hypothesis of the theorem, for $0 < \varepsilon < \varepsilon_1 \in (0,1)$, $D_j(p^\varepsilon) < D_j(p^*)$. The definitions of f and D imply that for $\varepsilon < \varepsilon_2 \in (0,1)$, $D_k(p^\varepsilon) = D_k(p^*)$ if $r^*_j \ge r^*_k$. Since $\sum_{j=0}^J D_j(p)$ is constant, it must be that for $\varepsilon < \min\{\varepsilon_1, \varepsilon_2\}$, we have that $D_k(p^\varepsilon) > D_k(p^*)$ for some k such that $r^*_k > r^*_j$.

For any non-empty subset $K \subset \{1, \ldots, J\}$, let $k = \arg\max_{k' \in K} r^*_{k'}$. By the argument above, there exists $l \in \{0, \ldots, J\}$ such that $r^*_l > r^*_k$ such that $D_l(p)$ is strictly increasing in p^k at p^*. Therefore, $D(p)$ satisfies local-connected substitutes at p^*.

We now show that $D^*(\tilde{p}_{\mathcal{S}}) = A'D(\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}})$ satisfies local connected substitutes at $\tilde{p}_{\mathcal{S}}$, where $\tilde{p}_{\mathcal{S},s} = \min\{p_j : s_j = s\}$, and $\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}})$ such that $D_j(\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}}) - q_j \le 0$ with equality if $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}_{\mathcal{S}}) > 0$.

Lemma C.3 implies that $D^*(\tilde{p}_{\mathcal{S}})$ satisfies weak substitutes. For small enough $\varepsilon > 0$, define $\tilde{p}^{\varepsilon}_{\mathcal{S},s'} = \tilde{p}^*_{\mathcal{S},s'}$ for $s' \neq s$, and $\tilde{p}^{\varepsilon}_{\mathcal{S},s} = \tilde{p}^*_{\mathcal{S},s} + \varepsilon$. Observe that this implies that $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}^{\varepsilon}_{\mathcal{S}}) + \tilde{p}^{\varepsilon}_{\mathcal{S},s}) > \tilde{p}^*_{\mathcal{J},j}(\tilde{p}^*_{\mathcal{S}}) + \tilde{p}^*_{\mathcal{S},s})$ for some j with $s_j = s$. Define $r^*_s = \max\{r^*_j : s_j = s\}$. For all programs j with $r^*_j \leq r^*_s$, $D_j(p^*) = D_j(\tilde{p}^*_{\mathcal{J}}(\tilde{p}^{\varepsilon}_{\mathcal{S}}) + A\tilde{p}^{\varepsilon}_{\mathcal{S}})$. Therefore, $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}_{\mathcal{S}}) + \tilde{p}_{\mathcal{S},s_j} = \tilde{p}^*_{\mathcal{J},j}(\tilde{p}^{\varepsilon}_{\mathcal{S}})\tilde{p}^{\varepsilon}_{\mathcal{S},s_j}$ if $r^*_j \leq r^*_s$. Since the $\sum_{s=0}^{S} D^*(\tilde{p}_{\mathcal{S}})$ is constant, an identical argument to the one above implies that for some s' such that $r^*_{s'} > r^*_s$, $D^*_{s'}(\tilde{p}^{\varepsilon}_{\mathcal{S}}) > D^*_{s'}(\tilde{p}^*_{\mathcal{S}})$ for small enough $\varepsilon > 0$. As above, $D^*(\tilde{p}_{\mathcal{S}})$ satisfies local connected substitutes at $\tilde{p}^*_{\mathcal{S}}$.

By Proposition C.3, the market clearing cutoff p^* is unique. Further, part (i) of Theorem 2 in (Berry et al., 2013) ensures that $\nabla_{p_+} D_+(p^*)$ is nonsingular.

□

Preliminaries for Propositions C.3 and C.4

Lemma C.3. *If $D(\cdot|\eta)$ is continuous in its arguments and satisfies weak substitutes, then $D^*(\tilde{p}_{\mathcal{S}}|\eta) = A'D(\tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}}|\eta)$ satisfies weak substitutes.*

Proof. Fix $\tilde{p}_{\mathcal{S}}$, $\tilde{p}_{\mathcal{J}} = \tilde{p}^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}})$ and $s \in \mathcal{S}$. Let J_s be the set of programs in school s, J^+_s be the set of programs in school s with $\tilde{p}_{\mathcal{J},j} > 0$ and J^0_s be the set of programs in school s with $\tilde{p}_{\mathcal{J},j} = 0$. Consider $\tilde{p}'_{\mathcal{S}}$ such that $\tilde{p}'_{\mathcal{S},s} = \tilde{p}_{\mathcal{S},s} + \varepsilon$ for $\varepsilon > 0$ such that $\varepsilon < \min\{\tilde{p}^*_j(\tilde{p}_{\mathcal{S}}) : j \in J^+_s\}$, and $\tilde{p}'_{\mathcal{S},t} = \tilde{p}_{\mathcal{S},t}$ if $t \in \mathcal{S}\backslash\{s\}$.

There are two cases to consider:

Case 1 $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}_{\mathcal{S}}) > 0$ ***for all*** $j \in J_s$***:*** Consider $\tilde{p}'_{\mathcal{J}}$ such that $\tilde{p}'_{\mathcal{J},j} = \tilde{p}_{\mathcal{J},j}$ for $j \notin J_s$ and $\tilde{p}'_{\mathcal{J},j} = p_{\mathcal{J},j} - \varepsilon$. By construction, $\tilde{p}'_{\mathcal{J}} + A\tilde{p}'_{\mathcal{S}} = \tilde{p}_{\mathcal{J}} + A\tilde{p}_{\mathcal{S}}$. Hence, $\tilde{p}'_{\mathcal{J}} = \tilde{p}^*_{\mathcal{J}}(\tilde{p}'_{\mathcal{S}})$. Therefore, $D^*(\tilde{p}_{\mathcal{S}}|\eta) = D^*(\tilde{p}'_{\mathcal{S}}|\eta)$, satisfying Assumption C.1.

Case 2 $\tilde{p}^*_{\mathcal{J},j}(\tilde{p}_{\mathcal{S}}) = 0$ ***for some*** $j \in J_s$***:*** We will construct a convergent sequence of cutoffs $\tilde{p}^k_{\mathcal{J}}$, such that $\lim_{k\to\infty} \tilde{p}^k_0 = p^*_0(\tilde{p}'_{\mathcal{S}})$, and show that $D^*_s(\tilde{p}_{\mathcal{S}}|\eta)$ is non-increasing in $\tilde{p}_{\mathcal{S},s}$ and $D^*_k(\tilde{p}_{\mathcal{S}}|\eta)$ is non-decreasing in $\tilde{p}_{\mathcal{S},s}$ for $k \neq s$.

Set $\tilde{p}^0_{\mathcal{J},j} = \tilde{p}_{\mathcal{J},j}$ for $j \in \mathcal{J}\backslash J^+_s$ and $\tilde{p}^0_{\mathcal{J},j} = \tilde{p}_{\mathcal{J},j} - \varepsilon$ otherwise. Note that for all $j \in \mathcal{J}\backslash J^0_s$, $\tilde{p}^0_j + \tilde{p}'_{\mathcal{S},s_j} = \tilde{p}_{\mathcal{J},j} + \tilde{p}_{\mathcal{S},s_j}$ and for $j \in J^0_s$, $\tilde{p}^0_j + \tilde{p}'_{\mathcal{S},s} = \tilde{p}_{\mathcal{S},s} + \varepsilon$. For each $j \in \mathcal{J}$ and $k \in \mathbb{N}$, construct the sequence $\tilde{p}^k_{\mathcal{J},j}$ such that $D_j((\tilde{p}^k_{\mathcal{J},j}, \tilde{p}^{k-1}_{\mathcal{J},-j}) + A\tilde{p}'_{\mathcal{S}}|\eta) - q_j \leq 0$ with equality if $\tilde{p}^k_{\mathcal{J},j} > 0$. Since $D_j((\tilde{p}^k_{\mathcal{J},j}, \tilde{p}^{k-1}_{\mathcal{J},-j}) + A\tilde{p}'_{\mathcal{S}}|\eta)$ satisfies weak substitutes, if $\tilde{p}^k_{\mathcal{J},-j} \geq \tilde{p}^{k-1}_{\mathcal{J},-j}$, then $\tilde{p}^{k+1}_{\mathcal{J},j} \geq \tilde{p}^k_{\mathcal{J},j}$. Therefore, $\tilde{p}^k_{\mathcal{J}}$ is a monotonically increasing sequence. Since $\tilde{p}^k_{\mathcal{J}}$ is bounded above, it must be that $\lim_{k\to\infty} \tilde{p}^k_{\mathcal{J}} = \tilde{p}^{\infty}_{\mathcal{J}}$ exists. Further, since $D_j(\tilde{p}_{\mathcal{J}} + A\tilde{p}'_{\mathcal{S}}|\eta)$ is continuous in $\tilde{p}_{\mathcal{J}}$, we have that $D_j(\tilde{p}^{\infty}_{\mathcal{J}} + A\tilde{p}'_{\mathcal{S}}|\eta) \leq 0$ with equality if $\tilde{p}^{\infty}_{\mathcal{J},j} > 0$. Hence, $\tilde{p}^{\infty}_{\mathcal{J}} = p^*_{\mathcal{J}}(\tilde{p}'_{\mathcal{S}}) \geq \tilde{p}^0_{\mathcal{J}}$, and we have that $\tilde{p}^*_{\mathcal{J}}(\tilde{p}'_{\mathcal{S}}) + A\tilde{p}'_{\mathcal{S}} \geq p^*_{\mathcal{J}}(\tilde{p}_{\mathcal{S}}) + A\tilde{p}_{\mathcal{S}}$.

We now show that $D_j(\tilde{p}_0^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) \geq D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$ $j \notin J_s$. Fix $j \in \mathcal{J}\backslash J_s$. If $\tilde{p}_{\mathcal{J},j}^*(\tilde{p}'_{\mathcal{S}}) > 0$, then it must be that $D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) = q_j \geq D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$. If $\tilde{p}_{\mathcal{J},j}^*(\tilde{p}'_{\mathcal{S}}) = 0$, then $D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) \geq D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$ from weak substitutes, since $\tilde{p}_{\mathcal{J},j}^*(\tilde{p}_{\mathcal{S}})+\tilde{p}_{\mathcal{S},s_j} = \tilde{p}_{\mathcal{J},j}^*(\tilde{p}_{\mathcal{S}})+\tilde{p}'_{\mathcal{S},s_j}$ and $\tilde{p}_{\mathcal{J},k}^*(\tilde{p}_{\mathcal{S}})+\tilde{p}_{\mathcal{S},s_k} \geq \tilde{p}_{\mathcal{J},k}^*(\tilde{p}_{\mathcal{S}})+\tilde{p}'_{\mathcal{S},s_k}$ for all $k \neq j$.

Finally, we show that $\sum_{j\in\mathcal{J}_s} D_j(\tilde{p}_0^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) \leq \sum_{j\in\mathcal{J}_s} D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$. Note that $D_0(\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) \geq D_0(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$ since $\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}} \geq \tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}$. The proof is complete by noting that $\sum_{j\in\mathcal{J}\cup\{0\}} D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) = \sum_{j\in\mathcal{J}\cup\{0\}} D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$ must be constant since each student can be assigned to only one program and $D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}'_{\mathcal{S}})+A\tilde{p}'_{\mathcal{S}}|\eta) \geq D_j(\tilde{p}_{\mathcal{J}}^*(\tilde{p}_{\mathcal{S}})+A\tilde{p}_{\mathcal{S}}|\eta)$ for all $j \in \{0\}\cup(\mathcal{J}\backslash J_s)$.

□

C.4 Convergence of Equilibrium Probabilities

In this section, we consider a sequence of n-player Bayesian games defined by a sequence of RSP+C mechanisms Φ^n. Let $\sigma(v,t) = (\sigma_{R_1}(v,t),\ldots,\sigma_{R_{|\mathcal{R}|}}(v,t))$ be a (type-symmetric) strategy for a player with utility vector v and priority type t. We allow $\sigma(v,t)$ to be a mixed strategy profile, although players generically have a pure strategy best-reponse. For each n, the lotteries are given by

$$\begin{aligned} L_{R_i,t_i}^{n,\sigma} &= \mathbb{E}_\sigma\left[\Phi^n((R_i,t_i),(R_{-i},T_{-i})|R_i,T_i\right] \\ &= \sum_{R_{-i},t_{-i}} \Phi^n((R_i,t_i),(R_{-i},T_{-i})\prod_{k\neq i} m^\sigma(R_k,t_k), \end{aligned}$$

where $m^\sigma(R_k,t_k) = f_T(t_k)\int \sigma_{R_k}(v;t)dF_{V|t_k}$. The strategy $\sigma^{*,n}$ a Bayesian Nash Equilibrium if for all R such that $\sigma_R^{*,n}(v;t) > 0$, we have that $v\cdot L_{R,t}^{\sigma^{*,n}} \geq v\cdot L_{R',t}^{n,\sigma^{*,n}}$ for all $R' \in \mathcal{R}$.

Define the **Large-Market Limit Mechanism** in the spirit of Azevedo and Budish (2017) as follows:

$$L_{R_i,t_i}^{\infty,\sigma} = \lim_{n\to\infty} \sum_{R_{-i},t_{-i}} \Phi^n((R_i,t_i),(R_{-i},T_{-i})\prod_{k\neq i} m^\sigma(R_k,t_k), \tag{C.7}$$

if it exists. Further, σ^* is a **Limit Equilibrium** if $\sigma_R^*(v,t) > 0$ implies that $v\cdot L_{R,t}^{\infty,\sigma^*} \geq v\cdot L_{R',t}^{\infty,\sigma^*}$ for all $R' \in \mathcal{R}$.

We now show that Bayesian Nash Equilibria of the mechanism in a large economy approximate equilibria of the large-market limit mechanism.

Proposition C.5. *Suppose Φ^n is an RSP+C mechanism. Fix a strategy σ^* such that the limit in equation (C.7) exists, the tie-breakers ν are non-degenerate and $D(p|\eta)$ and q admit a unique market clearing cutoff, where $\eta = m^{\sigma^*} \times \gamma_\nu$.*

1. *If $\sigma^{*,n}$ is a sequence BNE such that $\|\sigma^{*,n} - \sigma^*\|_F \to 0$, then $\|L^{n,\sigma^{*,n}}_{R_i,t_i} - L^{\infty,\sigma^*}_{R_i,t_i}\| \to 0$, where $\|\sigma^{*,n} - \sigma^*\|_F = \sup_R \int |\sigma^{*,n}_R(v,t) - \sigma^*_R(v,t)| \mathrm{d}F_{V,T}$.*

2. *If $\sigma^{*,n}$ is a sequence BNE such that $\|\sigma^{*,n} - \sigma^*\|_F \to 0$, the strategy σ^* is a limit equilibrium.*

3. *If σ^* is a limit equilibrium, then for each $\varepsilon > 0$, and large enough n, $\sigma^*_R(v,t) > 0$ implies that for all $R' \in \mathcal{R}$,*

$$v \cdot L^{n,\sigma^*}_{R,t} \geq v \cdot L^{n,\sigma^*}_{R',t} - \varepsilon\|v\|.$$

The result shows that a convergent sequence of Bayesian Nash Equilibria converge to a limit equilibrium, and that all limit equilibria are approximate BNE for large enough n. The result is similar in spirit to Kalai (2004), which shows that equilibria in limit games are approximate BNE in large games. From an empirical perspective, it also shows that equilibrium behavior in the game does not depend dramatically on the exact number of players once there are sufficiently many players.

Proof. **Part 1:** By the triangle inequality,

$$\|L^{n,\sigma^{*,n}}_{R_i,t_i} - L^{\infty,\sigma^*}_{R_i,t_i}\| \leq \|L^{n,\sigma^{*,n}}_{R_i,t_i} - L^{n,\sigma^*}_{R_i,t_i}\| + \|L^{n,\sigma^*}_{R_i,t_i} - L^{\infty,\sigma^*}_{R_i,t_i}\|.$$

By the assumptions of the proposition, the second term converges to 0. Now consider the first term:

$$L^{n,\sigma^{*,n}}_{R_i,t_i} - L^{n,\sigma^*}_{R_i,t_i} = \mathbb{E}_{\sigma^{*,n}}\left[\Phi^n((R_i,t_i),(R_{-i},t_{-i}))|R_i,t_i\right] - \mathbb{E}_{\sigma^*}\left[\Phi^n((R_i,t_i),(R_{-i},t_{-i}))|R_i,t_i\right],$$

where $\mathbb{E}_\sigma$ denotes the expectation taken with respect to draws of (R_k,t_k) taken from m^σ. Since Φ^n is an RSP+C mechanism, we have that

$$L^{n,\sigma^{*,n}}_{R_i,t_i} - L^{n,\sigma^*}_{R_i,t_i} = \mathbb{E}_{\sigma^{*,n}}\left[\int D^{(R_i,t_i,\nu)}(p^n)\mathrm{d}\gamma_\nu \middle| R_i,t_i\right] - \mathbb{E}_{\sigma^*}\left[\int D^{(R_i,t_i,\nu)}(p^n)\mathrm{d}\gamma_\nu \middle| R_i,t_i\right]. \quad \text{(C.8)}$$

Therefore, to complete the proof, we need to show that the right-hand side of this expression converges to zero.

Let $\eta^{*,n} = m^{\sigma^{*,n}} \times \gamma_\nu$ and $\eta^* = m^{\sigma^*} \times \gamma_\nu$, and observe that

$$\begin{aligned}
\|D(p|\eta^{*,n}) - D(p|\eta^*)\| &= \sup_j |D_j(p|\eta^{*,n}) - D_j(p|\eta^*)| \\
&= \sup_j |\eta^{*,n}(v_{p,j}) - \eta^*(v_{p,j})| \\
&= \sup_j \left| \sum_{(R,t)\in\mathcal{R}\times T} (m^{\sigma^{*,n}}(R,t) - m^{\sigma^*}(R,t))\gamma_\nu(\{\nu : f(R,t,\nu) \in v_{p,j}\}) \right| \\
&= \sup_j \left| \sum_{(R,t)\in\mathcal{R}\times T} \left(\int (\sigma_R^{*,n}(v,t) - \sigma_R^*(v,t)) \mathrm{d}F_{V,T} \right) \gamma_\nu(\{\nu : f(R,t,\nu) \in v_{p,j}\}) \right| \\
&\leq \|\sigma^{*,n} - \sigma^*\|_F \sup_j \left| \sum_{(R,t)\in\mathcal{R}\times T} \gamma_\nu(\{\nu : f(R,t,\nu) \in v_{p,j}\}) \right| \leq \|\sigma^{*,n} - \sigma^*\|_F
\end{aligned}$$

The right-hand side converges to 0 by assumption. Therefore, we have that

$$\sup_p \|D(p|\eta^{*,n}) - D(p|\eta^*)\| \xrightarrow{p} 0.$$

If η^n is a sequence of empirical measures constructed draws from $\eta^{*,n}$, we have that

$$\begin{aligned}
\sup_p \|D(p|\eta^n) - D(p|\eta^*)\| &\leq \sup_p \|D(p|\eta^n) - D(p|\eta^{*,n})\| + \sup_p \|D(p|\eta^{*,n}) - D(p|\eta^*)\| \\
&\leq \sup_{p,j} J|\eta^n(v_{p,j}) - \eta^{*,n}(v_{p,j})| + \sup_p \|D(p|\eta^{*,n}) - D(p|\eta^*)\| \xrightarrow{p} 0,
\end{aligned}$$

since $\mathcal{V} = \{v_{p,j} : p \in [0,1]^J, j \in J\}$ is a uniform Glivenko-Cantelli class.

By arguments identical to those made in Part 1 of Theorem A.3, if p^n is a market clearing cutoff for $D(p|\eta^n)$ and q^n, then $p^n \xrightarrow{p} p^*$ where p^* is the unique market clearing cutoff for $D(p|\eta^*)$ and q. By the continuous mapping theorem, for each (R,t), we have that

$$\int D^{(R,t,\nu)}(p^n)\mathrm{d}\gamma_\nu \xrightarrow{p} \int D^{(R,t,\nu)}(p^*)\mathrm{d}\gamma_\nu.$$

Since $D^{(R,t,\nu)}(p^n)$ is bounded, we have that

$$\mathbb{E}_{\sigma^{*,n}}\left[\int D^{(R,t,\nu)}(p^n)\mathrm{d}\gamma_\nu \middle| R,t \right] \to \int D^{(R,t,\nu)}(p^*)\mathrm{d}\gamma_\nu. \tag{C.9}$$

By a similar argument, we have that

$$\mathbb{E}_{\sigma^*}\left[\int D^{(R,t,\nu)}(p^n)\mathrm{d}\gamma_\nu \middle| R,t\right] \to \int D^{(R,t,\nu)}(p^*)\mathrm{d}\gamma_\nu. \tag{C.10}$$

Equations (C.9) and (C.10) imply that the right hand side of equation (C.8) converges to 0.

Part 2: Consider a sequence of equilibrium strategies $\sigma^{*,n}$ such that $\|\sigma^{*,n} - \sigma^*\|_F \to 0$. We will show that $\sigma^*_R(v,t) > 0$ for all $(v,t) \in \text{int}(\text{supp}F_{V,T})$ only if $v \cdot (L^{\infty,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R',t}) \geq 0$ for all $R' \in \mathcal{R}$.

Fix $(v,t) \in \text{int}(\text{supp}F_{V,T})$. Towards a contradiction, suppose that $\sigma^*_R(v;t) > 0$, and $v \cdot (L^{\infty,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R',t}) < -2\varepsilon$ for some $R' \in \mathcal{R}$ and $\varepsilon > 0$. Since $(v,t) \in \text{int}(\text{supp}F_{V,T})$, there exists a $\delta > 0$, such that for all v' with $\|v - v'\| < \delta$, we have $v' \in \text{int}(\text{supp}F_{V,T})$, and $v' \cdot (L^{\infty,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R',t}) < -\varepsilon$.

By Part 1, $\left\|L^{n,\sigma^{*,n}}_{R',t} - L^{\infty,\sigma^*}_{R',t}\right\| \to 0$. Since $L^{n,\sigma^{*,n}}_{R,t}$ is bounded, there exists an N, such that for all $n > N$ and all $R' \in \mathcal{R}$,

$$\left\|L^{n,\sigma^{*,n}}_{R',t} - L^{\infty,\sigma^*}_{R',t}\right\| \leq \frac{\varepsilon}{2(\|v\| + \delta)}.$$

Hence, for all v' in the δ neighborhood of v, we have that

$$\begin{aligned} v' \cdot (L^{n,\sigma^{*,n}}_{R,t} - L^{n,\sigma^{*,n}}_{R',t}) &\leq v' \cdot (L^{\infty,\sigma^{*,n}}_{R,t} - L^{\infty,\sigma^{*,n}}_{R',t}) + 2\|v'\|\|L^{n,\sigma^{*,n}}_{R',t} - L^{\infty,\sigma^*}_{R',t}\| \\ &\leq v' \cdot (L^{\infty,\sigma^{*,n}}_{R,t} - L^{\infty,\sigma^{*,n}}_{R',t}) + \varepsilon < 0 \end{aligned}$$

Since $\sigma^{*,n}$ is a Bayesian Nash Equilibrium strategy, it must be that for all $n > N$ and v' such that $\|v - v'\| < \delta$, $\sigma^{*,n}_R(v',t) = 0$. Therefore, $\|\sigma^{*,n} - \sigma^*\|_F \to 0$ implies that $\sigma^*(v',t) = 0$ for all v' in the δ neighborhood of v. This conclusion contradicts the hypothesis that $\sigma^*_R(v,t) > 0$ for any R such that $v \cdot (L^{\infty,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R',t}) < 0$. Hence, σ^* is a limit equilibrium.

Part 3: Consider the constant sequence $\sigma^{*,n} = \sigma^*$. By the assumptions of the proposition, for each (R,t),

$$\|L^{n,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R,t}\| \to 0.$$

Moreover, this convergence is uniform in (R,t) since $\mathcal{R} \times T$ is a finite set. Fix $\varepsilon > 0$ and pick n_0 such that for all $n > n_0$,

$$\sup_{R,t} \|L^{n,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R,t}\| < \frac{\varepsilon}{2}.$$

Note that the choice of n_0 did not depend on v_i.

Since σ^* is a limit equilibrium, $\sigma^*_{R_i}(v_i, t_i) > 0$ implies that for all R'_i,

$$\begin{aligned} & v_i \cdot L^{\infty,\sigma^*}_{R_i,t_i} && \geq v_i \cdot L^{\infty,\sigma^*}_{R'_i,t_i} \\ \Rightarrow \quad & v_i \cdot L^{n,\sigma^*}_{R_i,t_i} && \geq v_i \cdot L^{n,\sigma^*}_{R'_i,t_i} - 2\sup_{R,t} |v_i \cdot (L^{n,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R,t})| \end{aligned}$$

for all $n > n_0$. By the Cauchy-Schwarz inequality, $\sup_{R,t} |v_i \cdot (L^{n,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R,t})| \leq \|v_i\| \sup_{R,t} \|L^{n,\sigma^*}_{R,t} - L^{\infty,\sigma^*}_{R,t}\|$. Therefore,

$$v_i \cdot L^{n,\sigma^*}_{R_i,t_i} \geq v_i \cdot L^{n,\sigma^*}_{R'_i,t_i} - \varepsilon\|v_i\|.$$

□

D Auxilliary Results on Identification

D.1 Characterization of Partially Identified Set

Consider the collection of markets

$$\mathcal{T}(\xi, z) = \{\Gamma_{ib} = (\xi_b, z_{ib}, t_{ib}, \mathcal{L}_b) : (\xi_b, z_{ib}) = (\xi, z)\}.$$

The dependence of the set of lotteries $\mathcal{L}$ on the market index b indicates that we allow variation in this dimension to be useful in the present exercise. We will consider results that fix (ξ, z) and therefore drop this from the notation. As a reminder, conditioning on z is without loss since it is observed, but this implies that the researcher assumes that the variation considered holds school unobservables ξ fixed.

The next result characterizes what can be learned about the distribution of utilities from observing data from several markets in $\mathcal{T}$. Let $N_{\mathcal{L}_\Gamma}(L) = \{v \in \mathbb{R}^J : v \cdot (L - L') \geq 0 \text{ for all } L' \in \mathcal{L}_\Gamma\}$ be the normal cone to $L \in \mathcal{L}_\Gamma$ corresponding to the set $\mathcal{L}_\Gamma$. (We switch notation from using C_R for lottery L_R for clarity since this section uses different sets $\mathcal{L}_\Gamma$, which are not explicitly referred to in the relatively compact notation, C_R.) Further, let $\mathcal{N} = \{\text{int}(N_{\mathcal{L}_\Gamma}(L))\}_{\Gamma \in \mathcal{T}, L \in \mathcal{L}_\Gamma}$ be the collection of (the interiors of) normal cones to lotteries faced by agents in the markets $\mathcal{T}$. For a collection of sets $\mathcal{N}$, let $\mathcal{D}(\mathcal{N})$ be the smallest collection of subsets of $\mathbb{R}^J$ such that

1. $\mathbb{R}^J \in \mathcal{D}(\mathcal{N})$ and $\mathcal{N} \subset \mathcal{D}(\mathcal{N})$
2. For all $N \in \mathcal{D}(\mathcal{N})$, $N^c \in \mathcal{D}(\mathcal{N})$
3. For all countable sequences of sets $N_k \in \mathcal{D}(\mathcal{N})$ such that $N_{k_1} \cap N_{k_2} = \emptyset$, $\bigcup_k N_k \in \mathcal{D}(\mathcal{N})$

The collection $\mathcal{D}(\mathcal{N})$ is sometimes called the minimal Dynkin system containing $\mathcal{N}$.

Theorem D.1. *Given $P(L \in \mathcal{L}_\Gamma|\Gamma)$ for each $\Gamma \in \mathcal{T}$ and $L \in \mathcal{L}_\Gamma$, the quantity*

$$h_D = \int 1\{v \in D\} \mathrm{d}F_V(v)$$

is identified for each $D \in \mathcal{D}(\mathcal{N})$.

Proof. The identified set of conditional distributions $F_V(v)$ is given by

$$\mathscr{F}_I = \left\{ F_V \in \mathscr{F} : \forall L \in \mathcal{L}_\Gamma \text{ and } \Gamma \in \mathcal{T}, P(L \in \mathcal{L}_\Gamma|\Gamma) = \int 1\{v \in N_{\mathcal{L}_\Gamma}(L)\} \mathrm{d}F_V(v) \right\}.$$

Note that for any two distributions F_V and $\tilde{F}_V$ in $\mathscr{F}$, the collection of sets

$$\mathscr{L}(F_V, \tilde{F}_V) = \left\{ A \in \mathcal{F} : \int 1\{v \in A\} \mathrm{d}F_V(v) = \int 1\{v \in A\} \mathrm{d}\tilde{F}_V(v) \right\}$$

is a Dynkin system for the Borel σ-algebra $\mathcal{F}$. Since $\mathcal{D}(\mathcal{N})$ is the minimal Dynkin system where all elements of $\mathscr{F}_I$ agree, $\mathcal{D}(\mathcal{N}) \subseteq \mathscr{L}(F_V, \tilde{F}_V)$ for any two elements F_V and $\tilde{F}_V$. Hence, for all $D \in \mathcal{D}(\mathcal{N})$, we have that

$$h_D = \int 1\{v \in D\} \mathrm{d}F_V(v) = \int 1\{v \in D\} \mathrm{d}\tilde{F}_V(v)$$

is therefore identified. $\square$

The result follows from basic measure theory and characterizes the features of $F_V(v)$ that are identified under such variation in choice environments without any further restrictions. In particular, with the free normalization $\|v_i\| = 1$, the result implies that the mass accumulated on the projection of the sets in $\mathcal{D}(\mathcal{N})$ on the $J-1$ dimensional sphere, $\mathbb{S}^J$, is identified. Typically, this implies only partial identification of $F_V(v)$, but extensive variation in the lotteries could result in point identification.[2]

D.2 Preliminaries for Theorem A.2

Lemma D.1. *Let $f_{\varepsilon,C}(x) = 1\{x \in C\} e^{-2\pi\langle \varepsilon, x\rangle}$ for some polygonal, full-dimensional convex cone C and let $\hat{f}_{\varepsilon,C}(\xi)$ be its Fourier Transform. If C is salient and $\varepsilon \in \operatorname{int}(C^*)$ where C^* is the dual of C, then $\hat{f}_{\varepsilon,C}$ is an entire function. Further, there is no non-empty open subset of $\mathbb{R}^J$ where $\hat{f}_{\varepsilon,C}$ is zero.*

[2] Specifically, the $\pi - \lambda$ theorem implies that $F_V(v)$ is identified if and only if the Dynkin-system $\mathcal{D}(\mathcal{N})$ contains a π-system that generates the Borel σ-algebra.

Proof of Lemma D.1. Note that $\exists \varepsilon \in \operatorname{int}(C^*)$ because C is a salient cone. Let $\{C_1, \ldots, C_Q\}$ be a simplicial triangulation of C. Let A_q be a matrix $[a_{q1}, a_{q2}, ..., a_{qn}]$ with the linear independent vectors that span cone C_q arranged as column vectors. $x \in C_q \iff x = A_q\alpha$ for some $0 \leq \alpha \in \mathbb{R}^J \iff A_q^{-1}x \geq 0$. Normalize A_q so that $|\det A_q| = 1$. Let $f_{\varepsilon,C}(x) = 1\{x \in C\}e^{-2\pi\langle\varepsilon,x\rangle}$. This is an integrable function (if ε is in the dual of the cone C). Consider its Fourier transform:

$$
\begin{aligned}
\hat{f}_{\varepsilon,C}(\xi) &= \int_C \exp\left(-2\pi i \langle \xi - i\varepsilon, x\rangle\right) dx \\
&= \sum_Q \int_{C_q} \exp\left(-2\pi i \langle \xi - i\varepsilon, x\rangle\right) dx \\
&= \sum_Q \int_{\mathbb{R}^J} 1\{x : A_q^{-1}x \geq 0\} \exp\left(-2\pi i \langle \xi - i\varepsilon, x\rangle\right) dx \\
&= \sum_Q \int_{\mathbb{R}^J_+} \exp\left(-2\pi i \langle \xi - i\varepsilon, A_q y\rangle\right) dy \\
&= \sum_Q \int_{\mathbb{R}^J_+} \exp\left(-2\pi i \left\langle A_q'\xi - iA_q'\varepsilon, y\right\rangle\right) dy \\
&= \sum_{q=1..Q} \prod_{j=1..J} \int_{\mathbb{R}_+} \exp\left(-2\pi i \left(a_{qj}'\xi - ia_{qj}'\varepsilon\right) y\right) dy \\
&= \sum_{q=1..Q} \prod_{j=1..J} \int_{\mathbb{R}_+} \exp\left(-y\left[2\pi\left(a_{qj}'\varepsilon\right) + 2\pi i\left(a_{qj}'\xi\right)\right]\right) dy \\
&= \sum_{q=1..Q} \prod_{j=1..J} \frac{1}{2\pi} \frac{1}{\left[\left(a_{qj}'\varepsilon\right) + i\left(a_{qj}'\xi\right)\right]},
\end{aligned}
$$

where the last equality follows from the fact that $-y2\pi(a_{qj}'\varepsilon) < 0$. Note that the closed-form expression implies that $\hat{f}_{\varepsilon,C}(\xi)$ is an entire function for every $\varepsilon \in C\setminus\{0\}$. Therefore, if it is zero in an open subset of $\mathbb{R}^J$ is zero everywhere.

We now show that $\hat{f}_{\varepsilon,C}(\xi)$ is non-zero on a non-empty open set. Let K be a full-dimensional simplicial convex cone such that $C \subset K$. K exists because C is salient. Let A_K be the corresponding matrix for K. $\kappa_{qj} = A_K^{-1}a_{qj} > 0$ for all $q \in \{1, \ldots, Q\}$ and $j \in \{1, \ldots, J\}$. Consider $\xi = \left(A_K^{-1}\right)'\alpha$,

$$
\begin{aligned}
\hat{f}_{\varepsilon,C}\left(\left(A_K^{-1}\right)'\alpha\right) &= \left(\frac{1}{2\pi i}\right)^J \sum_{q=1,\ldots,Q} \prod_{j=1,\ldots,J} \frac{1}{\left[\left(\kappa_{qj}'\alpha\right) - i\left(a_{qj}'\varepsilon\right)\right]} \\
&= \left(\frac{1}{2\pi i}\right)^J \sum_{q=1,\ldots,Q} \prod_{j=1,\ldots,J} \frac{\left(\kappa_{qj}'\alpha\right) + \left(a_{qj}'\varepsilon\right) i}{\left[\left(\kappa_{qj}'\alpha\right)^2 + \left(a_{qj}'\varepsilon\right)^2\right]}
\end{aligned}
$$

Each term in the summation has a positive denominator and a numerator that is a polynomial function of α with positive coefficients. It follows that it is not zero everywhere, and therefore there is no open subset of $\mathbb{R}^J$ where $\hat{f}_{\varepsilon,C}$ is zero. $\square$

E Estimation Details

E.1 Gibbs' Sampler: Implementation Details

E.1.1 Optimal Responses

We adapt the Gibbs' Sampler for a standard discrete choice model from McCulloch and Rossi (1994) to our case. The main difference is that we have to draw latent utility vectors satisfying the restrictions $v_i \cdot (L_{R_i} - L_{R'}) \geq 0$ for all $R' \in \mathcal{R}$ instead of restrictions of the form $v_{ij} \geq v_{ij'}$ for all choices j' where j is the chosen option.

Let Z_i be a $J \times (K \times J)$ block-diagonal matrix that is constructed placing the K-row vector covariates $z_{ij} = [z_{ijk}]_{k=1}^K$ in each of the J blocks; $\beta = vec(\{\beta_{jk}\})$, a KJ-column vector; and D_i a $J \times J$ diagonal matrix with d_{ij} in the j-th position. The system in equation (2) can be compactly written as:

$$v_i = Z_i\beta - D_i + \varepsilon_i$$

The unobserved utilities v_i are treated as unknown parameters along with β and Σ. We specify independent prior distributions for β and Σ:

$$\begin{aligned} p(\beta, \Sigma) &= p(\beta)p(\Sigma), \\ \beta &\sim \mathcal{N}(\overline{\beta}, A^{-1}), \\ \Sigma &\sim IW(\nu_0, V_0), \end{aligned}$$

where IW is the inverse Wishart distribution.

The Gibbs sampler proceeds as follows:

0. Start with initial values Σ^0 and $v^0 = \{v_i^0\}_{i=1}^N$ so that $v_i^0 \in C_{R_i}$ for all $i = 1, \ldots, N$ where R_i is the report of student i.

   Since $C_{R_i} = \left\{v \in \mathbb{R}^J : \Gamma_i v \geq 0\right\}$ where $\Gamma_i = (L'_{R_i} - L'_{R_1}, \ldots, L'_{R_i} - L_{R'_{|\mathcal{R}|}})'$,[3] v_i^0 can be

[3]For the specification that assumes truthful reporting, Γ_i, is a matrix that encodes the inequalities implied by the rank order list $R_i = (R_i(1), \ldots, R_i(K))$. Hence, $\Gamma_i v_i > 0$ if and only if $v_{iR_i(1)} > v_{iR_i(2)} > \ldots > v_{iR_i(K)}$, $v_{i0} < v_{iR_i(K)}$ and $v_{ij} < v_{iR(K)}$ if $j \notin R_i$.

found by finding a solution to the inequalities

$$\Gamma_{ik} v_i \geq \varepsilon,$$

for each row k of Γ_i, and a small positive number ε. We implement this step using the Gurobi solver.

1. Draw $\beta^1 | v^0, \Sigma^0$ from a $N\left(\tilde{\beta}, V\right)$,

$$\begin{aligned} V &= \left(Z^{*\prime} Z^* + A\right)^{-1}, \tilde{\beta} = V\left(Z^{*\prime} v^* + A\overline{\beta}\right) \\ Z^* &= \begin{bmatrix} Z_1^* \\ \dots \\ Z_S^* \end{bmatrix} \\ Z_i^{*\prime} &= C' Z_i, v_i^* = C' v_i^0 \\ \Sigma^0 &= C'C, \end{aligned}$$

where $C'C$ results from a Cholesky decomposition of Σ^0.

2. Draw $\Sigma^1 | v^0, \beta^1$ from a $IW\left(\nu_0 + N, V_0 + S\right)$

$$\begin{aligned} S &= \sum_{i=1}^{n} \varepsilon_i \varepsilon_i', \\ \varepsilon_i &= v_i^0 - Z_i \beta^1 \end{aligned}$$

3. Draw $v^1 | \beta^1, \Sigma^1, R$ iterating over students and schools.

   For each school $j = 1 \dots J$, draw

$$v_{ij}^1 | \left\{v_{ik}^1\right\}_{k=1}^{j-1}, \left\{v_{ik}^0\right\}_{k=j+1}^{J}, \beta^1, \Sigma^1$$

   from a truncated normal $TN\left(\mu_{ij}, \sigma_{ij}^2, a_{ij}, b_{ij}\right)$, where

$$\begin{aligned} \mu_{ij} &= \sum_{k=1}^{K} \beta_{jk}^1 z_{ijk} - d_{ij} \\ \sigma_{ij}^2 &= \Sigma_{jj}^1 - \Sigma_{j(-j)}^1 \left[\Sigma_{(-j)(-j)}^1\right]^{-1} \Sigma_{(-j)j}^1 \end{aligned}$$

   and the truncation points a_{ij} and b_{ij} guarantee the draw v_{ij}^1 is such that

$$v = \left[\left\{v_{ik}^1\right\}_{k=1}^{j-1}, v_{ij}^1, \left\{v_{ik}^0\right\}_{k=j+1}^{J}\right]'$$

lies in the interior of C_{R_i}. To calculate these truncation points, define A^j_{ik} be the k-th row of Γ_i with its j-th column removed and let $v^j_i = \left[\{v^1_{ik}\}^{j-1}_{k=1}, \{v^0_{ik}\}^{J}_{k=j+1}\right]'$.[4]

$$
\begin{aligned}
a_{ij} &= \max_{k\in\left\{k:\Gamma_{ikj}>0\right\}} \frac{-A^j_{ik}v^j_i}{\Gamma_{ikj}} \\
b_{ij} &= \min_{k\in\left\{k:\Gamma_{ikj}<0\right\}} \frac{-A^j_{ik}v^j}{\Gamma_{ikj}}
\end{aligned}
$$

where Γ_{ikj} is the (k, j)-th element of Γ_i.

4. Set $\Sigma^0 = \Sigma^1$ and $v^0 = v^1$, store, and repeat the steps 1-3 to obtain (β^k, Σ^k, v^k) given $(\beta^{k-1}, \Sigma^{k-1}, v^{k-1})$ and the priors.

E.1.2 Naïve-Sophisticate Mixture Model

We extend the Gibbs' sampler described earlier to allow for two types of agents. The model assumes that naïve agents report truthfully while sophisticates pick the report that maximizes their expected utility. For a rank-order list $R = (R(1), R(2), \ldots, R(K))$ of length K, let $\tilde{C}_R$ be the region in utility space such that $v_i \in \tilde{C}_R \implies v_{iR(1)} > v_{iR(2)} > \ldots > v_{iR(K)} > v_{ij}$ for all $j \notin R_i$, and $v_{iR(K)} > v_{i0}$. Note that $\tilde{C}_R$ is a convex cone in $\mathbb{R}^J$. Let π_i be an indicator for whether a student is naïve. Therefore, the model specifies the observed report of the agent given v_i and π_i as follows:

$$
\begin{aligned}
R_i = R, \pi_i = 0 &\implies v_i \in C_R \\
R_i = R, \pi_i = 1 &\implies v_i \in \tilde{C}_R.
\end{aligned}
$$

The Gibbs' sampler for this model uses data augmentation on π_i in addition to v_i. Let $\bar{\pi}$ be the fraction of naӱe agents in the economy. We let $\bar{\pi}$ be a vector to allow for free-lunch and paid-lunch students to have differing proportions of naïve and sophisticated agents. We

[4]We pre-process the matrix Γ_i using Gurobi to eliminate redundant linear constraints to speed up this step. The k-th row is a redundant constraint if the solution to the problem

$$\min_v \Gamma_{ik} v \text{ subject to } \Gamma_i v \geq 0$$

is non-negative.

specify independent prior distributions for $\beta, \bar{\pi}$ and Σ:

$$
\begin{aligned}
p(\beta, \Sigma) &= p(\beta)p(\bar{\pi})p(\Sigma), \\
\beta &\sim \mathcal{N}(\overline{\beta}, \bar{\Sigma}^{-1}), \\
\bar{\pi}_l &\sim Beta\,(a_0, b_0) \\
\Sigma &\sim IW(\nu_0, V_0),
\end{aligned}
$$

where IW is the inverse Wishart distribution and $l \in \{\text{Paid Lunch, Free Lunch}\}$. The Gibbs' sampler proceeds as follows:

0. Start with initial values Σ^0, $\pi^0 = \{\pi_i^0\}_{i=1}^N$, and $v^0 = \{v_i^0\}_{i=1}^N$ so that $v_i^0 \in \tilde{C}_{R_i}$ for all $i = 1, \ldots, N$.

1-2. Update (Σ, β) according to steps 1-2 in Appendix E.1.

3. Update $\bar{\pi}^1|\pi^0$. For $l \in \{\text{Paid Lunch, Free Lunch}\}$, draw $\bar{\pi}_l$ from

$$
Beta\left(a_0 + |\mathcal{I}_l| - \sum_{i \in \mathcal{N}_l} \pi_i^0, b_0 + \sum_{i \in \mathcal{I}_l} \pi_i^0\right),
$$

where $\mathcal{I}_l$ is the set of students in paid/free-lunch group l.

4. Draw $v^1|\beta^1, \Sigma^1, \bar{\pi}^1, y$ iterating over students and schools. For the observed report R_i for student i, consider the cones

$$
\begin{aligned}
\tilde{C}_{R_i} &= \left\{v \in \mathbb{R}^J : v_{R_i(1)} > v_{R_i(2)} > \ldots > v_{R_i(K)} > v_{ij} \text{ for all } j \in \{0, \ldots, J\} \backslash R_i\right\} \\
C_{R_i} &= \left\{v \in \mathbb{R}^J : \Gamma_i v \geq 0\right\},
\end{aligned}
$$

where $\Gamma_i = (L'_{R_i} - L'_{R_1}, \ldots, L'_{R_i} - L_{R'_{|\mathcal{R}|}})'$. Let $\bar{\pi}_i^1 = \bar{\pi}_l^1$, for l equal to the paid lunch status of i. For each school $j = 1...J$, draw

$$
v_{ij}^1 |\left\{v_{ik}^1\right\}_{k=1}^{j-1}, \left\{v_{ik}^0\right\}_{k=j+1}^{J}, \beta^1, \Sigma^1, \bar{\pi}_i^1
$$

from a mixture of two truncated normals $\mathcal{TN}\left(\mu_{ij}, \sigma_{ij}^2, \tilde{a}_{ij}, \tilde{b}_{ij}\right)$ and $\mathcal{TN}\left(\mu_{ij}, \sigma_{ij}^2, a_{ij}, b_{ij}\right)$ with weights $\bar{\pi}_i^1$ and $(1 - \bar{\pi}_i^1)$. $\mu_{ij}, \sigma_{ij}^2, a_{ij}$ and b_{ij} are defined as in step 3 in Appendix E.1. The truncation points $\left(\tilde{a}_{ij}, \tilde{b}_{ij}\right)$ guarantee that draws from $\mathcal{TN}\left(\mu_{ij}, \sigma_{ij}^2, \tilde{a}_{ij}, \tilde{b}_{ij}\right)$ lay in the interior of $\tilde{C}_{R_i}$.

5. Update $\pi^1|v^1, \bar{\pi}^1$. For each student i, draw π_i^1 from a binomial distribution with parameter $\bar{\pi}_i^1$ if $v_i^1 \in C_{R_i} \cap \tilde{C}_{R_i}$. If $v_i^1 \in C_{R_i} \backslash \tilde{C}_{R_i}$, set $\pi_i^1 = 0$. If $v_i^1 \in \tilde{C}_{R_i} \backslash C_{R_i}$, set $\pi_i^1 = 1$.

6. Repeat steps 1-5 to obtain $(\beta^k, \Sigma^k, v_i^k, \pi_i^k, \bar{\pi}^k)$ given $(\beta^{k-1}, \Sigma^{k-1}, v_i^{k-1}, \pi_i^{k-1}, \bar{\pi}^{k-1})$.

We parametrize v_i as in Appendix E.1 and assume identical distributions for naïves are sophisticates.

E.1.3 Priors

We use very diffuse priors to minimize their influence on our estimates and as a reflection of our prior uncertainty about the values of the parameters of the model. We set the prior distribution of $\beta \sim \mathcal{N}(\overline{\beta}, \bar{\Sigma}^{-1})$

$$\begin{aligned} \overline{\beta} &= 0 \\ \bar{\Sigma}^{-1} &= 100 \times I \end{aligned}$$

and the prior of $\Sigma \sim IW(\nu_0, V_0)$

$$\begin{aligned} \nu_0 &= 100 \\ V_0 &= I. \end{aligned}$$

We experimented with more diffuse priors $\left(\bar{\Sigma}^{-1} = 200 \times I, \nu_0 = 50\right)$ without noticeable changes in our main results.

For the mixture model, we set the prior of $\bar{\pi}_l = Beta\,(a_0, b_0)$, with $a_0 = b_0 = 1$ for $l \in \{\text{Paid Lunch, Free Lunch}\}$.

E.1.4 Convergence Diagnostics

For each specification, the Gibbs' sampler produces a Markov chain with the posterior distribution of the parameters as the invariant distribution. Since the chain is ergodic, it ultimately converges to this distribution irrespective of the starting point. However, it is essential to burn-in a large set of initial draws since they are influenced by the starting point, and to check that the chains have converged. We simulate three chains of length 400,000 and burn-in the first half to ensure mixing. The three chains with different starting values were used to assure convergence to the same parameter value. We monitored convergence by examining the trace plots of the various co-efficients and use Geweke's means test across and within the chains to ensure mixing. Finally, we use the Raftery-Lewis Diagnosis Test to to check that the chain has been simulated for long enough. The test quantifies whether a low quantile has been estimated precisely in order to diagnose convergence of the distribution. We check that the 2.5th percentile of the vast majority of parameters are estimated within a tolerance of 0.005 with 95% probability.

E.2 Bootstrap

The standard errors for $\hat{L}$, $\hat{\theta}$, and counterfactuals were estimated by a bootstrap. To construct each of the S bootstrap samples we sampled n students with replacement from each year of our sample, where n is the number of students in that year. For each bootstrap sample $s \in \{1, \ldots, S\}$, we computed:

- Lottery estimate $\hat{L}^s$: For each of the five years in the data, we computed $\hat{L}^s$ using the bootstrap sample s using the same procedure used to obtain $\hat{L}$. i.e. we resampled $n-1$ individuals and generated $n-1$ draws of the tie-breaker $B = 1,000$ times. For each simulated sample b, we computed the market clearing cutoff $p_{b,s}^{n-1}$, and for each (R, t) calculated the vector of assignment probabilities averaging across the B simulated samples following equation (9). The standard errors for the lotteries presented in table E.I in the Appendix are the standard deviation of the $\hat{L}^s$ across $S = 1,000$ bootstrap samples.

- Parameter estimates $\hat{\beta}^s, \hat{\Sigma}^s$: We ran a Monte Carlo Markov Chain on the bootstrap sample s using the same procedure described in the paper and in Appendix E.1 using the bootstrap samples. We ran one chain of $100,000$ draws and burned-in the first $50,000$. The last $50,000$ draws were used to compute the mean of each parameter which we denote $\hat{\beta}^s, \hat{\Sigma}^s$. The standard errors in tables VII and E.III were estimated by the standard deviation of the mean utilities and $\hat{\beta}_s$ across the $S = 250$ bootstrap samples. We used a smaller number of bootstrap samples, S, in this step to reduce the computational burden of drawing a large number of Markov chains.

- Counterfactual: We simulated the deferred acceptance counterfactual assuming parameters $\hat{\beta}^s, \hat{\Sigma}^s$ and computed the difference in utility for each individual in the bootstrap sample s. For the Cambridge mechanism, we used $\hat{L}^s$. The standard errors reported in tables X and XII were estimated by the standard deviation of the difference in utilities across the $S = 250$ boostrap samples.

The same boostrap procedure was used to compute standard errors for the coarse beliefs, adaptative expectations and mixture specifications. However, the standard errors for the truthful specification were not obtained by bootstrap. They were estimated directly from the original MCMC chains.

References

Abdulkadiroglu, A., Pathak, P. A., and Roth, A. E. (2009). Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match. *American Economic Review*, 99(5):1954–1978.

Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., and Sonmez, T. (2006). Changing the Boston School Choice Mechanism. *NBER Working Papers 11963.*

Azevedo, E. M. and Budish, E. (2017). Strategyproofness in the large. *NBER Working Paper*, No. 23771.

Azevedo, E. M. and Leshno, J. (2016). A Supply and Demand Framework for Two-Sided Matching Markets. *Journal of Political Economy*, 124(5):1235–1268.

Berry, S. T., Gandhi, A., and Haile, P. A. (2013). Connected Substitutes and Invertibility of Demand. *Econometrica*, 81(5):2087–2111.

Berry, S. T. and Haile, P. A. (2010). Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis : The Kernel Approach with S-Plus Illustrations: The Kernel Approach with S-Plus Illustrations*. OUP Oxford.

Chen, Y. and Kesten, O. (2013). From Boston to Chinese parallel to deferred acceptance: Theory and experiments on a family of school choice mechanisms. *Discussion Papers, Research Unit: Market Behavior.*

Dubins, L. and Freedman, D. (1981). Machiavelli and the Gale-Shapley Algorithm. *The American Mathematical Monthly1*, 88(7):485–494.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(5):1921–1947.

Kalai, E. (2004). Large Robust Games. *Econometrica*, 72(6):1631–1665.

Kapor, A., Neilson, C., and Zimmerman, S. (2017). Heterogeneous Beliefs and School Choice.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

Pennell, H., West, A., and Hind, A. (2006). Secondary School Admissions in London.

Roth, A. E. (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7(4):617–628.

van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.

Table E.I: Estimated Assignment Probabilities w/ Standard Errors

Ranked	Graham Parks	Haggerty	Baldwin	Morse	Amigos	Amigos Bilingual	Cambridgeport	King Open	Peabody	Tobin	Tobin K4	Tobin K5	Fletcher Maynard	Kenn Long	MLK	King Open Ola
Panel A: All Students																
First	0.43	0.59	0.63	0.57	0.73	0.98	0.60	1.00	0.94	0.85	0.31	0.34	0.92	1.00	1.00	1.00
	[0.03]	[0.06]	[0.05]	[0.08]	[0.06]	[0.03]	[0.08]	[0.01]	[0.04]	[0.04]	[0.08]	[0.11]	[0.02]	[0.00]	[0.01]	[0.01]
Second	0.24	0.25	0.23	0.20	0.35	0.94	0.18	0.92	0.83	0.74	0.04	0.14	0.86	1.00	0.99	1.00
	[0.03]	[0.05]	[0.05]	[0.09]	[0.09]	[0.07]	[0.07]	[0.07]	[0.07]	[0.06]	[0.07]	[0.14]	[0.03]	[0.01]	[0.02]	[0.01]
Third	0.21	0.19	0.18	0.10	0.25	0.83	0.10	0.67	0.61	0.66	0.02	0.08	0.77	0.90	0.90	0.89
	[0.03]	[0.03]	[0.03]	[0.06]	[0.07]	[0.06]	[0.03]	[0.07]	[0.07]	[0.05]	[0.05]	[0.10]	[0.03]	[0.02]	[0.03]	[0.02]
Panel B: Paid Lunch																
First	0.22	0.45	0.49	0.54	0.73	1.00	0.51	1.00	0.94	0.93	0.32	0.36	1.00	1.00	1.00	1.00
	[0.03]	[0.07]	[0.06]	[0.08]	[0.06]	[0.00]	[0.09]	[0.02]	[0.04]	[0.07]	[0.08]	[0.12]	[0.00]	[0.00]	[0.00]	[0.01]
Second	0.00	0.05	0.03	0.16	0.35	1.00	0.08	0.89	0.82	0.76	0.03	0.16	1.00	1.00	1.00	1.00
	[0.00]	[0.05]	[0.04]	[0.08]	[0.09]	[0.02]	[0.07]	[0.08]	[0.07]	[0.10]	[0.07]	[0.15]	[0.00]	[0.00]	[0.00]	[0.01]
Third	0.00	0.01	0.00	0.06	0.24	0.85	0.01	0.56	0.56	0.64	0.01	0.09	0.89	0.89	0.89	0.87
	[0.00]	[0.02]	[0.01]	[0.05]	[0.07]	[0.04]	[0.03]	[0.09]	[0.08]	[0.08]	[0.05]	[0.11]	[0.02]	[0.02]	[0.02]	[0.02]
Panel C: Free/Reduced Lunch																
First	0.82	0.87	0.90	0.64	0.74	0.97	0.77	1.00	0.94	0.72	0.31	0.29	0.76	1.00	1.00	1.00
	[0.05]	[0.07]	[0.06]	[0.08]	[0.07]	[0.06]	[0.07]	[0.01]	[0.04]	[0.02]	[0.09]	[0.12]	[0.07]	[0.01]	[0.02]	[0.01]
Second	0.71	0.65	0.61	0.26	0.35	0.90	0.39	0.98	0.86	0.71	0.07	0.08	0.59	0.99	0.99	1.00
	[0.08]	[0.10]	[0.10]	[0.11]	[0.10]	[0.12]	[0.10]	[0.05]	[0.08]	[0.01]	[0.11]	[0.12]	[0.10]	[0.03]	[0.04]	[0.01]
Third	0.62	0.56	0.52	0.18	0.27	0.83	0.28	0.87	0.70	0.68	0.03	0.05	0.53	0.92	0.92	0.94
	[0.07]	[0.09]	[0.08]	[0.08]	[0.08]	[0.10]	[0.08]	[0.07]	[0.08]	[0.03]	[0.08]	[0.08]	[0.09]	[0.04]	[0.05]	[0.03]

Note: Average estimates weighted by number of students of each type. Probabilities estimated using B=1,000. Ranks and priority types of opposing students are drawn with replacement from the observed data. Second and third rank assignment probabilities are conditional on no assignment to higher ranked choices, averaged across feasible rank order lists. Standard errors computed using 1,000 subsamples from each of the 1,000 bootstrap samples in parentheses.

Table E.II: Estimated Preference Parameters: Truthful Reporting

	Constant	Paid Lunch	Sibling	Black	Asian	Hispanic	Other Ethn	Spanish	Portuguese	Other Lang	Unobs s.d.
Graham Parks	2.23	1.18	3.31	-0.68	0.07	-0.60	-0.46	-0.46	-2.87	-0.14	1.77
	[0.16]	[0.14]	[0.31]	[0.15]	[0.17]	[0.22]	[0.31]	[0.33]	[0.84]	[0.15]	[0.10]
Haggerty	2.59	0.91	4.70	-0.92	-0.21	-0.76	-0.27	-0.97	-2.09	0.13	1.98
	[0.18]	[0.16]	[0.45]	[0.18]	[0.20]	[0.26]	[0.35]	[0.43]	[1.01]	[0.17]	[0.10]
Baldwin	2.33	1.06	3.56	-0.58	0.15	-0.44	-0.11	-0.92	-1.59	-0.39	1.87
	[0.16]	[0.14]	[0.32]	[0.15]	[0.17]	[0.22]	[0.31]	[0.35]	[0.63]	[0.15]	[0.11]
Morse	1.92	0.65	3.82	0.03	0.35	-0.44	-0.01	-0.42	-2.21	0.19	1.98
	[0.18]	[0.14]	[0.34]	[0.16]	[0.20]	[0.24]	[0.36]	[0.36]	[0.77]	[0.16]	[0.09]
Amigos	1.19	0.77	11.99	-0.54	-0.32	0.67	-0.28	0.65	-0.85	-0.91	1.65
	[0.19]	[0.15]	[3.01]	[0.17]	[0.21]	[0.22]	[0.35]	[0.33]	[0.56]	[0.19]	[0.10]
Cambridgeport	1.94	0.99	4.71	-0.50	-0.35	-0.47	-0.39	-0.65	-2.61	-0.05	1.71
	[0.15]	[0.13]	[0.60]	[0.15]	[0.17]	[0.21]	[0.31]	[0.32]	[0.70]	[0.15]	[0.09]
King Open	1.94	0.74	4.52	-0.11	-0.06	-0.27	-0.35	-0.61	-0.65	-0.25	1.59
	[0.15]	[0.12]	[0.49]	[0.14]	[0.17]	[0.20]	[0.32]	[0.30]	[0.45]	[0.14]	[0.08]
Peabody	1.88	0.34	3.60	-0.02	0.28	-0.37	-0.19	-0.19	-1.90	-0.02	1.69
	[0.17]	[0.14]	[0.39]	[0.16]	[0.18]	[0.23]	[0.33]	[0.35]	[0.84]	[0.15]	[0.08]
Tobin	1.79	-0.31	4.41	0.01	0.37	-0.20	-0.21	0.23	-0.45	0.17	1.85
	[0.19]	[0.16]	[0.48]	[0.18]	[0.21]	[0.28]	[0.40]	[0.41]	[0.76]	[0.18]	[0.09]
Flet Mayn	1.12	-0.36	2.99	0.61	-0.08	0.09	0.31	-0.12	-10.35	0.05	1.65
	[0.18]	[0.16]	[0.40]	[0.17]	[0.21]	[0.24]	[0.36]	[0.32]	[5.25]	[0.16]	[0.10]
Kenn Long	1.86	0.04	2.70	-0.01	0.09	-0.10	-0.44	-0.16	-0.68	-0.15	1.44
	[0.16]	[0.13]	[0.26]	[0.15]	[0.18]	[0.21]	[0.37]	[0.30]	[0.43]	[0.15]	[0.08]
MLK	1.15	0.12	2.50	0.40	0.43	0.10	-0.13	0.27	-1.04	0.15	1.52
	[0.18]	[0.13]	[0.36]	[0.15]	[0.19]	[0.22]	[0.33]	[0.30]	[0.58]	[0.15]	[0.10]
King Open Ola	-0.43	0.19	12.98	0.20	-2.11	-2.97	-3.13	-3.02	4.72	-4.67	0.67
	[0.33]	[0.19]	[3.32]	[0.22]	[1.12]	[0.63]	[2.58]	[2.14]	[0.63]	[1.63]	[0.13]

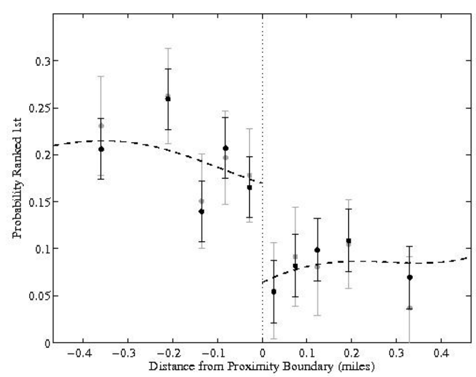
Notes: Demand estimates under the assumption of truthful reporting. N=2128. Excluded ethnicity is white and excluded language is english. The table reports means and standard deviations of the posterior distribution of each parameter. The distance coefficient is normalized to -1; therefore, all magnitudes are in equivalent miles

Table E.III: Estimated Preference Parameters: Rational Expectations

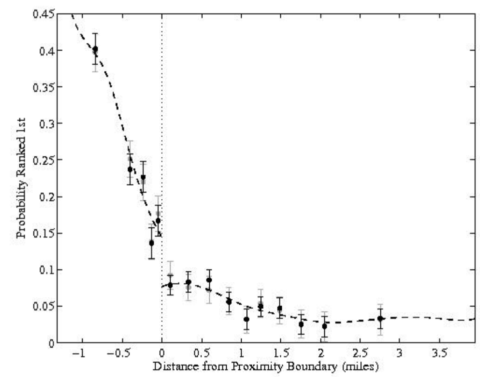
	Constant	Paid Lunch	Sibling	Black	Asian	Hispanic	Other Ethn	Spanish	Portuguese	Other Lang	Unobs. s.d.
Graham Parks	0.83	1.52	4.64	-0.21	0.17	-0.41	-0.40	-0.80	-2.38	-0.12	1.90
	[0.35]	[0.47]	[1.06]	[0.29]	[0.26]	[0.39]	[0.53]	[1.35]	[4.92]	[0.25]	[0.34]
Haggerty	1.68	0.22	4.68	-0.40	-0.38	-0.34	0.00	-1.04	-0.83	0.41	1.52
	[0.23]	[0.23]	[0.76]	[0.22]	[0.21]	[0.39]	[0.34]	[1.79]	[4.82]	[0.19]	[0.13]
Baldwin	1.65	0.15	3.45	-0.10	0.14	-0.08	0.51	-0.66	0.02	-0.27	1.06
	[0.17]	[0.17]	[0.83]	[0.17]	[0.16]	[0.20]	[0.31]	[0.83]	[3.04]	[0.15]	[0.11]
Morse	1.30	-0.31	3.85	0.42	0.46	-0.27	0.55	0.20	-0.43	0.43	1.18
	[0.21]	[0.16]	[0.82]	[0.18]	[0.18]	[0.27]	[0.39]	[0.61]	[4.25]	[0.15]	[0.09]
Amigos	0.35	0.04	13.36	0.11	-0.35	1.19	0.27	1.06	-0.92	-0.97	1.44
	[0.28]	[0.21]	[4.07]	[0.21]	[0.24]	[0.28]	[0.43]	[0.44]	[4.13]	[0.27]	[0.15]
Cambridgeport	1.30	-0.16	5.12	-0.05	-0.28	-0.11	-0.07	-0.57	-10.71	0.09	1.00
	[0.16]	[0.14]	[2.53]	[0.15]	[0.16]	[0.18]	[0.25]	[0.36]	[4.83]	[0.13]	[0.10]
King Open	0.96	-0.07	6.26	0.29	0.11	-0.01	0.00	-0.68	0.67	-0.16	1.20
	[0.17]	[0.14]	[0.73]	[0.16]	[0.14]	[0.20]	[0.25]	[0.42]	[1.81]	[0.15]	[0.09]
Peabody	0.92	-0.48	4.14	0.41	0.25	-0.31	-0.12	0.17	-0.46	-0.01	1.42
	[0.19]	[0.16]	[0.75]	[0.17]	[0.19]	[0.29]	[0.35]	[0.51]	[4.43]	[0.16]	[0.10]
Tobin	0.75	-1.28	5.20	0.59	0.52	0.13	-0.34	0.59	-12.73	0.30	1.87
	[0.33]	[0.27]	[0.92]	[0.33]	[0.30]	[0.36]	[0.96]	[0.65]	[4.73]	[0.24]	[0.17]
Flet Mayn	-0.01	-1.68	4.10	1.24	-0.18	0.71	1.12	0.06	-12.78	0.44	1.65
	[0.34]	[0.33]	[0.95]	[0.27]	[0.40]	[0.39]	[0.49]	[0.50]	[5.13]	[0.24]	[0.20]
Kenn Long	0.77	-0.71	3.12	0.08	0.36	0.32	-0.80	0.12	-5.79	0.17	1.29
	[0.20]	[0.20]	[0.43]	[0.22]	[0.22]	[0.28]	[3.23]	[0.39]	[5.14]	[0.19]	[0.13]
MLK	-0.17	-0.76	3.42	0.81	0.66	0.52	0.13	0.55	-0.13	0.45	1.56
	[0.31]	[0.20]	[1.80]	[0.24]	[0.24]	[0.36]	[1.03]	[0.47]	[5.39]	[0.20]	[0.16]
King Open Ola	-0.48	-0.36	13.89	0.27	-1.25	-2.22	-10.94	-2.74	4.46	-2.14	0.83
	[0.55]	[0.44]	[3.60]	[0.48]	[1.20]	[1.57]	[4.14]	[2.43]	[1.90]	[1.78]	[0.30]

Notes: Demand estimates under the rational expectations assumption. N=2071. Excluded ethnicity is white and excluded language is english. The table reports means and bootstrap standard errors of each parameter. The distance coefficient is normalized to -1; therefore, all magnitudes are in equivalent miles

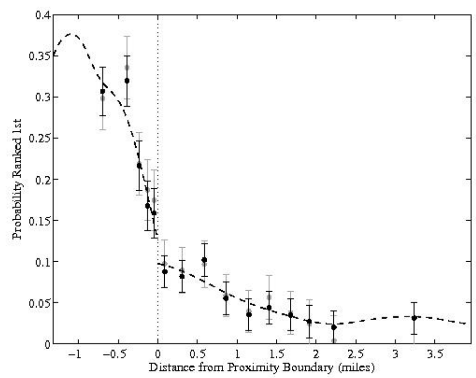
Figure E.I: Effect of Proximity Priority on Ranking Behavior



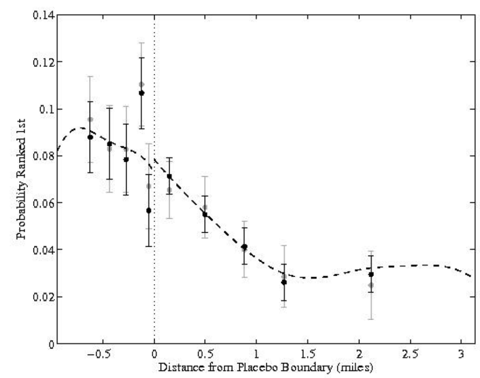
(i) Second and Third Closest Schools



(ii) First Rank: Paid Lunch Students



(iii) First Rank: Free Lunch Students



(iv) Pacebo at Unprioritized Schools

Notes: The graphs are bin-scatter plots (based on distance) with equally sized bins on either side of the boundary. For each student, we construct a boundary distance, $\bar{d}_i$, based on her distance to the schooling options. For a given school-student pair, the horizontal axis represents $d_{ij} - \bar{d}_i$. The vertical axis is the probability that a student ranks the school in the relevant distance bin. Range plots depict 95% confidence intervals. Black plot points are based on the raw data, while the grey points control for school fixed effects. Dashed lines represent local linear fits estimated on either side of the boundary based on bandwidth selection rules recommended in Bowman and Azzalini (1997) (page 50). Panels (a) through (c) use the average distance between the second and third closest schools as the boundary. A student is given proximity priority at the schools to the left of the boundary and does not receive priority at schools to the right. Panel (a) only considers the two closest schools. Panel (d) drops the two closest schools and considers a placebo boundary at the mid-point of the fourth and fifth closest schools. All panels plot the probability that a school is ranked first. Distances as calculated using ArcGIS. Proximity priority recorded by Cambridge differs from these calculations in about 20% of the cases. Graphs are qualitatively similar when using only students with consistent calculated and recorded priorities. Details in data appendix.