Structural Estimation of a Model of School Choices: the Boston Mechanism vs. Its Alternatives (Online Appendix)

Caterina Calsamiglia,[*] Chao Fu[†] and Maia Güell[‡]

November 12, 2018

1 Backward Induction Solution Method

The backward induction solution method is applicable to a broad class of mechanisms, which includes BM, constrained and unconstrained DA, first preference first, Chinese parallel, and variants or hybrids of the above.[1] In the following, we refer to this class as the class under consideration. This section is organized as follows. First, we prove that the optimization problem for strategic households can be fully solved via backward induction. Second, we show how one can apply this method to mechanisms other than BM in the class under consideration. Then, we formally describe the dimensionality involved in the solution.

1.1 Proof

In the following, we prove that our solution method solves a strategic household's problem fully, using the standard BM as an example. The proof is easily adaptable to other mechanisms in the class under consideration. Proofs for these other cases are available upon request.

Proposition 1 *Consider a standard Boston Mechanism where a household can choose up to R schools out of J, who is given a vector of school-specific priorities $\left(S_i \equiv \{s_{ij}\}_{j=1}^J\right)$ and a random lottery number that is used in all rounds. Let v_{ij} be the value of being assigned to School j for Household i, let χ_i be the vector of preference-related variables (e.g., household characteristics, location and tastes), let $\pi(S_i, \chi_i, A)$ be the expected payoff from listing A.*

[*]ICREA, IPEg and CEPR.

[†]Corresponding author: Chao Fu, University of Wisconsin and NBER. cfu@ssc.wisc.edu.

[‡]University of Edinburgh, CEPR, FEDEA and IZA.

[1]Particular variants of the discussed algorithms can be characterized in a similar fashion, e.g., the mechanism in Cambridge, serial dictatorship and the Pan London Admissions.

Then, the solution to $\max_{A\in\mathbf{P}(J;R)}\pi\left(S_i,\chi_i,A\right)$ *is equivalent to the solution to the following recursive problem, with value functions given by*

$$V^r\left(S_i,\chi_i,\overline{\xi}_i^r\right)=\max_{j\in J}\left\{p_j^r\left(S_i|\overline{\xi}_i^r\right)v_{ij}+(1-p_j^r\left(S_i|\overline{\xi}_i^r\right))V^{r+1}\left(S_i,\chi_i,\overline{\xi}_i^{r+1}\right)\right\}, \tag{1}$$

$$s.t.\ \overline{\xi}_i^{r+1}=\begin{cases}\min\left\{cut_j,\overline{\xi}_i^r\right\} & \text{if } s_{ij}=\overline{s}_j \text{ and } r=\overline{r}_j,\\ \overline{\xi}_i^r \ \text{otherwise,}\end{cases}$$

$$p_j^r\left(S_i|\overline{\xi}_i^r\right)=\begin{cases}1 \text{ if } r<\overline{r}_j \text{ or } (r=\overline{r}_j \text{ and } s_{ij}>\overline{s}_j),\\ \max\left\{0,\frac{\overline{\xi}_i^r-cut_j}{\overline{\xi}_i^r}\right\} \text{ if } r=\overline{r}_j \text{ and } s_{ij}=\overline{s}_j,\\ 0 \text{ otherwise.}\end{cases}$$

$$V^{R+1}(\cdot)=v_{i0},\ \overline{\xi}_i^1=1.$$

Proof. Step 1: show recursive representation. An optimal strategy is a list A of ranked schools of length $n\le R$. When $n<R$, there is a list of length R that yields the same expected payoff and hence is also optimal (e.g., by adding a zero-probability school for positions $n+1,...,R$). Therefore, without loss of generality, consider lists of length R. Let $p_j^r(S_i,A)$ be the probability of being admitted to school j when j is ranked the r^{th} on A for someone with priority score vector S_i. As a common feature of mechanisms in the class under consideration, the ranked schools on an application list are considered sequentially in the procedure, and the r^{th}-listed school (a_r) is relevant only if one is rejected by all previously listed schools. In particular, one's probability of being admitted to a_r does not depend on which schools are listed after a_r, i.e.,

$$p_j^r(S_i,A)=p_j^r(S_i,\{a_{r'}\}_{r'=1}^{r-1}).$$

Therefore, the expected payoff from submitting A is given by
$\pi\left(S_i,\chi_i,A\right)$

$$\equiv p_{a_1}^1(S_i,A)v_{i,a_1}+(1-p_{a_1}^1(S_i,A))\left(\begin{array}{c}p_{a_2}^2(S_i,A)v_{i,a_2}+\\ \left(1-p_{a_2}^2(S_i,A)\right)\left(p_{a_3}^3(S_i,A)v_{i,a_3}+\ldots(1-p_{a_R}^R(S_i,A))v_{i0})...\right)\end{array}\right)$$

$=$

$$p_{a_1}^1(S_i)v_{i,a_1}+(1-p_{a_1}^1(S_i))\left(\begin{array}{c}p_{a_2}^2(S_i,a_1)v_{i,a_2}+\\ \left(1-p_{a_2}^2(S_i,a_1)\right)\left(p_{a_3}^3(S_i,\{a_{r'}\}_{r'=1}^2)v_{i,a_3}+\ldots(1-p_{a_R}^R(S_i,\{a_{r'}\}_{r'=1}^{R-1}))v_{i0})...\right)\end{array}\right).$$

It is easy to see that

$$\max_{A\in P(J,R)}\pi\left(S_i,\chi_i,A\right) \tag{2}$$

is equivalent to

$$V^1(S_i,\chi_i) = \max_j\{p_j^1(S_i)v_{ij} + (1-p_j^1(S_i))V^2(\cdot,a_1)\},$$

with

$$V^{R+1}(S_i,\chi_i,\{a_{r'}\}_{r'=1}^R) = v_{i0},$$
$$V^R(S_i,\chi_i,\{a_{r'}\}_{r'=1}^{R-1}) = \max_j \left\{p_{a_R}^R(S_i,\{a_{r'}\}_{r'=1}^{R-1})v_{ij} + (1-p_{a_R}^R(S_i,\{a_{r'}\}_{r'=1}^{R-1}))V^{R+1}(\cdot,\{a_{r'}\}_{r'=1}^R)\right\},$$
$$...,$$
$$V^r(S_i,\chi_i,\{a_{r'}\}_{r'=1}^{r-1}) = \max_j\{p_j^r(S_i,\{a_{r'}\}_{r'=1}^{r-1})v_{ij} + (1-p_j^r(S_i,\{a_{r'}\}_{r'=1}^{r-1}))V^{r+1}(\cdot,\{a_{r'}\}_{r'=1}^r)\}.$$

That is, the best-permutation problem (2) can always be expressed in a recursive manner, where the state space in value function $V^r(\cdot)$ contains the ordered schools listed so far $\{a_{r'}\}_{r'=1}^{r-1}$. The recursive problem with $\{a_{r'}\}_{r'=1}^{r-1}$ as state variables in $V^r(\cdot)$ has the same dimensionality as in (2).

Step 2: show that the relevant information contained in $\{a_{r'}\}_{r'=1}^{r-1}$ is fully captured by $\overline{\xi}_i^r$. From the BM algorithm, it follows that admission probabilities to each school j can be characterized by a triplet $(\overline{r}_j,\overline{s}_j,cut_j)$, where $\overline{r}_j$ is the round at which j is filled up, $\overline{s}_j$ is the priority score for which lottery numbers are used to break ties for j's slots, cut_j is the cutoff of the lottery number for admission to j. School j will admit any r^{th}-round applicant before $\overline{r}_j$, any $\overline{r}_j^{th}$-round applicant with $s_{ij} > \overline{s}_j$, and any $\overline{r}_j^{th}$-round applicant with score $\overline{s}_j$ and random lottery higher than cut_j; and it will reject any other applicant. That is, after being rejected by $(a_1,...a_{r-1})$, i will be rejected by a_r if a_r has been filled up in some previous round; if a_r still has available seats, whether or not a_r will admit her at Round r is *fully* determined by her score s_{ia_r} and her random lottery number. Among the two factors, one's priority score is fully determined by pre-determined characteristics (e.g., applicant characteristics, school characteristics and the interaction between the two) and is independent of one's application list. The lottery number is drawn after the application is submitted and hence unknown to the applicant when making her decisions. Because a household has a single lottery number across all tie-breaking cases, correlation arises between admissions probabilities across rounds: the probability of being allocated in Round r conditional on being rejected by $(a_1,...a_{r-1})$ is (weakly) lower than the unconditional probability. In particular, if one is rejected by a_1 due to losing the lottery in Round 1 reveals that one's lottery number is below cut_{a_1}, if one is rejected again by a_2 due to losing the lottery in Round 2 reveals that one's lottery number is below $\min\{cut_{a_1},cut_{a_2}\}$, and so on. However, other than this, it follows from the algorithm that $(a_1,...a_{r-1})$ bears no information that is payoff relevant for one's decision on a_r. Therefore, the relevant information in $(a_1, ...a_{r-1})$ can be fully captured in an updated upper bound on one's lottery number $\overline{\xi}_i^r$ and the problem is fully described as in (1). ■

Remark 1 *Step 1 in the proof above holds in all mechanisms within the broad class under consideration. Step 2 differs across specific mechanisms. For example, in some mechanisms, one's priority for a school also depends on the rank position of the school on her list, however, it does not depend on other schools listed. One exception is the BM in Barcelona and Spain in general, where one's priorities for all schools depend on the school one listed first, which makes these cases more complicated than regular cases, which can still be efficiently solved via backward induction, as shown in the main text. We show in the following that the solution can be easily adapted to other mechanisms.*

1.2 Backward Induction Solution Method under Various Mechanisms

The two properties, i.e., sequentiality and reducible history, are common across the class of mechanisms under consideration.[2] Therefore, our backward induction solution method is applicable to all of them. In this class, mechanisms differ in two aspects. First, whether or not the assignment in a given round is final (e.g., BM versus DA). Second, whether or not the lottery numbers used to break ties for a given households for different schools are correlated.[3] In other words, the implementation of our solution method differs across these mechanisms in the way a household computes and updates their beliefs about their admissions probabilities when rejected by a school listed at position r. In the main text, we have described the case of BM in detail. In the following, we show how these probabilities may be constructed for some of the other best known mechanisms.[4]

1.2.1 DA (both constrained and unconstrained)

As in Abdulkadiroğlu, Chen and Yasuda (2015), we can characterize admissions probabilities under DA through cutoffs $(\bar{s_j}, cut_j)$. In particular, the probability of being assigned to school a_r when rejected from the previously listed schools, $p_j(\cdot)$, is independent of r and depends only on whether one's priority and/or lottery number are above or below the cutoffs $(\bar{s_j}, cut_j)$. The probabilities can be characterized by $(\bar{s_j}, cut_j)$. Let $\overline{\xi}_i^r$ be the upper bound

[2] Across all mechanisms, it is commonly assumed in the literature that households are small players, who take the equilibirum admissions probabilities as given.

[3] The most common practice is to use the same lottery number across all rounds for a given household.

[4] Particular variants of the discussed algorithms can be characterized in a similar fashion, e.g., the mechanism in Cambridge, serial dictatorship and the Pan London Admissions.

of i's lottery number after being rejected by first $r-1$ choices $(a_1, ...a_{r-1})$, then

$$p_j\left(S_i|\overline{\xi}_i^r\right) = \begin{cases} 1 \text{ if } s_{ij} > \overline{s}_j, \\ \max\left\{0, \frac{\overline{\xi}_i^r - cut_j}{\overline{\xi}_i^r}\right\} \text{ if } s_{ij} = \overline{s}_j, \\ 0 \text{ if } s_{ij} < \overline{s}_j. \end{cases} \tag{3}$$

$$\overline{\xi}_i^{r+1} = \begin{cases} \min\left\{cut_j, \overline{\xi}_i^r\right\} \text{ if } s_{ij} = \overline{s}_j, \\ \overline{\xi}_i^r \text{ otherwise.} \end{cases}$$

1.2.2 The parallel Chinese system

This system is carefully described in Chen and Kesten (2017) as a hybrid of BM and DA. In the parallel mechanism, rounds (r) and choice bands (b) shall be distinguished. Each choice band consists of a number of rounds. Let $l_1 < l_2 < ...l_B$ be the cumulative numbers of rounds at the end of each band. Allocation starts from Band 1, applicants are pre-accepted in every round until Round l_1 is reached (the end of the first choice band), when the assignment is final. The algorithm proceeds into the rounds of the following choice band, starting from Round $l_1 + 1$, by pre-accepting applicants for the seats that are still available, until Round l_2 is reached. The procedure continues up until everybody is assigned a seat. Probability of admissions can be characterized for each school by $(\bar{b}_j, \overline{s}_j, cut_j)$, where $\bar{b}_j$ is the choice band at which school j is filled up, and the remaining is as in BM or DA.

$$\overline{\xi}_i^{r+1} = \begin{cases} \min\left\{cut_j, \overline{\xi}_i^r\right\} \text{ if } s_{ij} = \overline{s}_j \text{ and } l_{\bar{b}_j-1} + 1 \leq r \leq l_{\bar{b}_j}, \\ \overline{\xi}_i^r \text{ otherwise,} \end{cases} \tag{4}$$

$$p_j^r\left(S_i|\overline{\xi}_i^r\right) = \begin{cases} 1 \text{ if } r \leq l_{\bar{b}_j-1} \text{ or } (l_{\bar{b}_j-1} + 1 \leq r \leq l_{\bar{b}_j} \text{ and } s_{ij} > \overline{s}_j) \\ \max\left\{0, \frac{\overline{\xi}_i^r - cut_j}{\overline{\xi}_i^r}\right\} \text{ if } l_{\bar{b}_j-1} + 1 \leq r \leq l_{\bar{b}_j} \text{ and } s_{ij} = \overline{s}_j, \\ 0 \text{ otherwise.} \end{cases}$$

1.2.3 First preference first systems

These systems include any mechanism where priorities are determined by the rank a particular school is given on the list. In cases where assignments are final in each round, the implementation of the backward induction is similar to the BM case. In cases where students are only pre-accepted in each round, the implementation is similar to the DA case.

1.3 Dimensionality of Strategic Household's Problem under BM

Case 1) School-Specific Priorities and Lottery Numbers: the dimensionality is $J \times R$, i.e., choosing the best school out of J for R times.

Case 2) School-Specific Priorities and Single Lottery Number

Let $O^r \subseteq J$ be the subset of schools that have been filled up by the beginning of Round r, the size of which is given by $|O^r| \equiv \sum_{j=1}^{J} I(r > \overline{r}_j)$. Let $J^r(S_i) \subseteq O^{r+1}$ be the subset of schools for which a household with priority score vector S_i can possibly be subject to lotteries in round r, the size of which is given by

$$|J^r(S_i)| \equiv \sum_{j=1}^{J} I(s_{ij} = \overline{s}_j, r = \overline{r}_j). \tag{5}$$

Let $NA^r(S_i) \subseteq O^{r+1}$ be the subset of schools that would reject i for sure in Round r (NA for not available).

In Round 1, $\overline{\xi}_i^1 \in \{1\}$. For Round $r > 1$, including the unconditional upper bound of 1, the maximum number of different values the state variable $\overline{\xi}_i^r$ can take is $1 + \sum_{r'=1}^{r-1} \left|J^{r'}(S_i)\right|$, which happens when the cut_j's are all different and those occur in Round r are uniformly higher than those occurring in Round $r+1$.[5] Notice 1) $\overline{\xi}_i^r = 1$ for $r > 1$ is possible only if the school listed $a_{r'} \in NA^{r'}(S_i)$ for all $r' < r$, and 2) $\left\{\cup_{r'<r} J^{r'}(S_i), \cup_{r'<r} NA^{r'}(S_i)\right\} \subseteq O^r$ hence $1 + \sum_{r'=1}^{r-1} \left|J^{r'}(S_i)\right| \leq |O^r|$; and the inequality is strict if $|\cup_{r'<r} NA^{r'}(S_i)| > 1$. Therefore, given there are J schools to choose from in each round (including those with zero admissions probabilities), the dimension of the problem in Case 2) cannot be larger than $J\left(1 + \sum_{r=2}^{R}\left(1 + \sum_{r'=1}^{r-1}\left|J^{r'}(S_i)\right|\right)\right) \leq J\left(1 + \sum_{r=2}^{R}|O^r|\right)$, which is much smaller than $|\mathbf{P}(J; R)|$.

Case 3) Constant Priority Score and Single Lottery Number

In Round 1, a household's state variable is again $\overline{\xi}_i^1 \in \{1\}$. In Round $r > 1$, besides the state variable $\overline{\xi}_i^r$, there is also an additional state variable, the priority score of one's top-listed school. Let Ω_i be the support of Household i's priority scores and $|\Omega_i|$ be the size of Ω_i. For a given s in Ω_i, the number of schools for which Household i can possibly be subject to lotteries in round $r > 1$ is given by $|J^r(s\mathbf{1})| \equiv \sum_{j=1}^{J} I(s = \overline{s}_j, r = \overline{r}_j)$. Therefore given s, the maximum number of different values the state variable $\overline{\xi}_i^r$ can take is at most $1 + \sum_{r'=1}^{r-1}\left|J^{r'}(s\mathbf{1})\right|$. Notice that $\cup_{r'<r, s\in\Omega_i} J^{r'}(s\mathbf{1}) \subseteq O^r$, hence $\sum_{s\in\Omega_i}\left(1 + \sum_{r'=1}^{r-1}\left|J^{r'}(s\mathbf{1})\right|\right) \leq |\Omega_i| + |O^r|$.

[5] For example, suppose there is one cut_1 in Round 1 and one cut_2 in Round 2, so that $\overline{\xi}_i^1 \in \{1, cut_1\}$. Given the rule that $\overline{\xi}_i^2 = \min\left\{cut_2, \overline{\xi}_i^1\right\}$, if $cut_1 \leq cut_2$ then the $\overline{\xi}_i^2 \in \{1, cut_1\}$, while if $cut_1 > cut_2$, $\overline{\xi}_i^2 \in \{1, cut_1, cut_2\}$.

Therefore, the dimension of the problem in Case 3) cannot be larger than
$J\left(1+\sum_{s\in\Omega_i}\sum_{r=2}^{R}\left(1+\sum_{r'=1}^{r-1}\left|J^{r'}(s\mathbf{1})\right|\right)\right)=J\left(1+\sum_{r=2}^{R}\sum_{s\in\Omega_i}\left(1+\sum_{r'=1}^{r-1}\left|J^{r'}(s\mathbf{1})\right|\right)\right)$
$\leq J\left(1+\sum_{r=2}^{R}\left(|\Omega_i|+|O^r|\right)\right)=J\left(1+(R-1)|\Omega_i|+\sum_{r=2}^{R}|O^r|\right)$, which is in turn smaller than $|\mathbf{P}(J;R)|$.

Example Consider the case with $J=30$, $R=3$ and $|\Omega_i|=2$, i.e., one has 2 different priority levels. Among the 30 schools, 5 are leftover schools, 10 are filled up in the first round, 10 in the second, 5 in the third. In this case, $O^1=0$, $O^2=10$ and $O^3=20$. The dimensionality (D) under the best-permutation solution is $D=|\mathbf{P}(J;R)|=25,260$. For each of the three BM cases, we have
Case *1)* $D<JR=90$.
Case *2)* $D<J\left(1+\sum_{r=2}^{R}|O^r|\right)=30(1+10+20)=990$.
Case *3)* $D<J\left(1+(R-1)|\Omega_i|+\sum_{r=2}^{R}|O^r|\right)=30(1+(2-1)\times 3+(10+20))=1,020$.

2 Model Details

2.1 Properties of the optimal list for a strategic household

Consider an optimal list $A_i^1=\left\{a_1^1,...,a_r^1,...a_{R'}^1\right\}$ derived via backward induction, if the student does not face 100% admissions rate for any of the first $r-1$ listed schools, and she does for the r^{th} listed school $\left(p^r_{a^1_r}\left(s\mathbf{1},\overline{\xi}^r_i\right)=1\right)$, then the following lists all generate the same value for the household as A_i^1 does, and hence are all optimal:
1) a list that shares the same first r elements of A_i^1.[6]
2) a list of length n ($r<n\leq R$), which shares the same first $r-1$ elements of A_i and the last (n^{th}) element is a_r^1 with 100% admissions probability for Household i at Round n, and for all elements $r'\in\{r,...,n-1\}$, i faces 0 admissions probability.
3) Furthermore, if this r^{th} listed school is one's backup school with $p_{a_r}(:)=1$, then any list of length n ($r-1\leq n\leq R$) is also optimal if it has the same first $r-1$ elements of A_i^1 and the admissions probabilities to the other elements are all 0.

Consider an optimal list $A_i^1=\left\{a_1^1,...,a_r^1,...a_{R'}^1\right\}$ derived via backward induction, if the student does not face 100% admissions rate for any of the first $r-1$ listed schools, and she does for the r^{th} listed school $\left(p^r_{a^1_r}\left(s\mathbf{1},\overline{\xi}^r_i\right)=1\right)$, then the following lists all generate the same value for the household as A_i^1 does, and hence are all optimal:
1) a list that shares the same first r elements of A_i^1.[7]

[6] In particular, one optimal list may have the same school j listed in two different rounds $r<r'$, with $p_j^r\left(s\mathbf{1}|\overline{\xi}^r_i\right)=1$.

[7] In particular, one optimal list may have the same school j listed in two different rounds $r<r'$, with $p_j^r\left(s\mathbf{1}|\overline{\xi}^r_i\right)=1$.

2) a list of length n ($r < n \leq R$), which shares the same first $r-1$ elements of A_i and the last (n^{th}) element is a_r^1 with 100% admissions probability for Household i at Round n, and for all elements $r' \in \{r, ..., n-1\}$, i faces 0 admissions probability.
3) Furthermore, if this r^{th} listed school is one's backup school with $p_{a_r}(:) = 1$, then any list of length n ($r-1 \leq n \leq R$) is also optimal if it has the same first $r-1$ elements of A_i^1 and the admissions probabilities to the other elements are all 0.

2.2 Proof for Claim 1

An application list with the following features reveals that the household must be non-strategic: 1) for some r^{th} ($r > 1$) element a_r on the list $p_{a_r}^r(\cdot) = 0$, and 2) $p_{a_{r'}}^{r'}(\cdot) < 1$ for all $r' < r$, and 3) for some $r^{''} \geq r+1$, $0 < p_{a_{r^{''}}}^{r^{''}}(\cdot) < 1$ and $p_{a_{r^{'''}}}^{r^{'''}}(\cdot) < 1$ for any $r < r''' < r^{''}$.

The conditions characterize a dominated strategy: the applicant ranks a 0 probability school when he is not guaranteed a slot in any previously-ranked school, and his chances at at least one school listed immediately below are reduced by this act. Without Feature 2), the list can still be strategically optimal due to Remark 2. Without Feature 3) a household may still be strategic if it prefers some sure-to-get-in school listed later over any of the schools listed after a_r, including a_r. All three features guarantee that the household is non-strategic. The formal proof is as follows.[8]

Proof. Take a given list that satisfies all three features in Claim 1: $A = \left\{a_1, ..., a_r, ..., a_{r''}, ...\right\}$, where $a_{r''}$ is the *first* school that satisfies Feature 3). This implies that $p_{a_{r^{'''}}}^{r^{'''}}(s\mathbf{1}|\overline{\xi}_i^{r^{'''}}) = 0$ for all $r < r^{'''} < r^{''}$, since Feature 3) ensures that they be smaller than 1 and $r^{''}$ is the first to be strictly positive. Let $W_i^r(A)$ be the residual value of this list starting from the r^{th} element. Given $p_{a_{r^{'''}}}^{r^{'''}}(s1|\overline{\xi}_i^{r^{'''}}) = 0$ for all $r < r^{'''} < r^{''}$, the continuation value at Round r is

[8] Notice that the proof, as consistent with the model and the allocation rule, uses the fact that $v_{ij} = E\max\{u_{ij}, \eta\}$, i.e., a household can either choose to attend the school it is assigned to or the outside option. If a household manages to give up its assigned seat in exchange for a seat in a leftover school that is not on its application list, the proof still holds as long as there is a cost, even if tiny, involved in such post-assignment re-arrangement.

the same as that in Round r''. That is,

$$\begin{aligned}
W_i^r(A) &= p_{a_r}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_r} + (1-p_{a_r}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r+1}(A)\\
&= W_i^{r+1}(A)\\
&= p_{a_{r+1}}^{r+1}\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_{r+1}} + (1-p_{a_{r+1}}^{r+1}\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r+2}(A)\\
&= W_i^{r+2}(A)\\
&= \ldots\\
&= p_{a_{r''-1}}^{r''-1}\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_{r''-1}} + (1-p_{a_{r''-1}}^{r''-1}\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r''}(A)\\
&= W_i^{r''}(A).
\end{aligned}$$

Consider an alternative (not necessarily optimal) application list $B = \left\{a_1, ..., a_{r''}, ..., a_{r''}, ...\right\}$, which differs from A only in that a_r is replaced by $a_{r''}$. Note that $p_{a_{r''}}^r(.) = 1$ since the school is filled up in round $r'' > r$.[9] The residual value of this list at its r^{th} element (now $a_{r''}$) is given by

$$\begin{aligned}
W_i^r(B) &= p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_{r''}} + (1-p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r+1}(B)\\
&= p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_{r''}} + (1-p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r+1}(A)\\
&= p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right)v_{ia_{r''}} + (1-p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right))W_i^{r''}(A)\\
&> p_{a_{r''}}^{r''}\left(s\mathbf{1}|\overline{\xi}_i^{r''}\right)v_{ia_{r''}} + (1-p_{a_{r''}}^{r''}\left(s\mathbf{1}|\overline{\xi}_i^{r''}\right))W_i^{r''}(A)\\
&= W_i^{r''}(A)
\end{aligned}$$

The inequality holds because $p_{a_{r''}}^{r''}\left(s\mathbf{1}|\overline{\xi}_i^{r''}\right) \in (0,1)$ implies $p_{a_{r''}}^r\left(s\mathbf{1}|\overline{\xi}_i^r\right) = 1$ for $r < r''$, and

$$v_{ia_{r''}} = E\max\left\{u_{ia_{r''}}, \eta\right\} > E(\eta) = 0.$$

Given that the first $r-1$ elements are also unchanged, it is immediate that the value of the whole list $W_i^1(B) > W_i^1(A)$. ■

2.3 Discussion: Size Condition on Non-Strategic Households Application

A non-strategic household lists schools according to its true preferences $\{v_{ij}\}_j$. Suppose household i ranks its backup school as its n_i^*-th favorite, then the length of i's application

[9] Note that for both list A and list B, the value of $\overline{\xi}_i^r$ in every round is the same because there is no updating for rounds when $p_{a_{r'''}}^{r'''}(s\mathbf{1}|\overline{\xi}_i^{r'''}) = 0$ for all $r < r''' < r''$. As a result, $W_i^{r+1}(B) = W_i^{r+1}(A)$.

list n_i is such that

$$n_i \geq \min \{n_i^*, R\}. \tag{6}$$

That is, when there are still slots left on its application form, a non-strategic household will list at least up to its backup school.

In order to predict allocation outcomes and to calculate welfare, we need to predict the content of a household's application list at least up to the point beyond which listing any additional schools will not affect the allocation outcome. Consider an application list of full length $A_i^0 = \{a_1^0, a_2^0, ..a_R^0\}$, if none of the R schools listed admits the household for sure in the round it is listed, then the entire list is outcome-relevant. If some elements in A_i^0 are such that i's admissions probability to a_r^0 is 1 in Round r, then the list is outcome-relevant only up to its r^*-th element $a_{r^*}^0$, where $a_{r^*}^0$ is the foremost-listed school that admits the household for sure. To predict the outcome, we could impose a different condition labeled Condition S (S for strong) that, when the list is incomplete, a non-strategic household list at least up to a_{r*}^0. However, Condition S implicitly requires that a non-strategic household knows that its admissions probability to School a_{r*}^0 in Round r^* is 1, which involves a substantial amount of sophistication. In comparison, Condition (6) is a much weaker requirement that a non-strategic household know which schools will be leftovers and list at least up to its backup school. It requires far less sophistication than to know the admissions probabilities by school and by round.[10] One reason is the high persistence in whether or not a school was left over, which was true even between the two years before and after the drastic re-definition of priority zones: 265 out of the 317 schools were either left over twice or never left over in the years of 2006 and 2007. Such high persistence makes it easy to predict the set of leftover schools. Therefore, it may be reasonable to believe that even the non-strategic households may have this (minimal) level of sophistication. With this weak requirement, Condition (6) achieves the same goal as Condition S.[11]

[10] In this model, we have assumed that it is free to fill in the application and, if failing to be assigned within R rounds, to propose a leftover school. Given the knowledge of the set of leftover schools and that leftover schools have 100% admissions probabilities, a non-strategic household would be indifferent between adding or not adding its backup school to an incomplete list. Condition (6) specifies that, if indifferent, a non-strategic household will add its backup school. It is also consistent with a situation where the cost of proposing a leftover school after being rejected in all rounds is higher than listing one more school to one's list.

[11] To see why Condition (6) achieves the same goal as its much stronger counterpart, consider the following exhaustive cases. Case 1: None of its R favorite schools admits the household for sure. Both Condition S and Condition (6) require the same full list of length R. Case 2: At least one of its R favorite schools admits the household for sure. By definition, a backup school admits the student for sure and therefore $r^* \leq n_i^*$. If $r^* < n_i^*$, i.e., the first sure-to-get school is preferrable to the backup: lengthening the list to n_i^* will not change the outcome, because only the first r^* elements are outcome-relevant. If $r^* = n_i^*$, then both conditions lead to the same list.

3 Identification

In this section, we will first prove the identification of our model in its simplified version. Then, we will give more intuition by contrasting the original model with a re-estimated model, the latter imposing that a lower fraction of households be strategic. Finally, we will discuss the exclusion restriction.

3.1 Proof

Since the dispersion of post-application shocks is mainly identified from the enrollment decisions, to ease the illustration, we show the identification of the model without post-application shocks. A household has observables (x_i, l_i) and can be one of two types $T = 0, 1$. Home-school distance is given by $d_{ji} = d(l_i, l_j)$ and z_{l_i} is the zone that l_i belongs to. Let the taste for school be $\epsilon_{ij} \sim i.i.d.\ N(0,1)$.[12] In line with (A2) and (A3) in the paper, assume that d is independent of T conditional on (x, z_l) and ϵ is independent of (x, l, T). To give the idea, consider the case where a household can apply only to one school from the choice set of schools 1 and 2, and where all households face the same admissions probabilities. Household-specific admissions probabilities provide more variations, which will provide more identification power.

Let $\overline{u}_{ij}$ be the utility net of individual taste, $\overline{u}_{ij} = g(\kappa_j, w_j, x_i) - C(d_{ij})$, where $g(\cdot)$ is the reduced form function given by

$$g(\kappa_j, w_j, x_i) = \tau_1 x_{i4} + \tau_2 \left[I(x_{i5} = j) - I(x_{i5} = 0) \right] + \sum_e (\delta_{0e} + \delta_{1e}\kappa_j + \rho_e w_j) I(edu_i = e).$$

Let $p_j > 0$ be the probability of admission to school j and $p_1 \neq p_2$ (A1). Let y be the decision to list School 1. y is related to the latent variable y^* in the following way

$$y(x_i, l_i, \epsilon_i, T) = 1 \text{ if only if}$$

$$y^*(x_i, l_i, \epsilon_i, T) = T(p_{i1}u_{i1} - p_{i2}u_{i2}) + (1 - T)(u_{i1} - u_{i2}) > 0.$$

Hence the probability of observing the decision to list 1 by someone with (x_i, l_i) is

$$H(x_i, l_i) = \lambda(x_i, z_{l_i}) \Phi\left(\frac{p_1 \overline{u}_{i1} - p_2 \overline{u}_{i2}}{\sqrt{p_1^2 + p_2^2}} \right) + (1 - \lambda(x_i, z_{l_i})) \Phi\left(\frac{\overline{u}_{i1} - \overline{u}_{i2}}{\sqrt{2}} \right).$$

[12] Given that the linear distance enters the utility function with coefficient of minus one, the standard deviation of ϵ is identified from the variation in distance within (x, z_l) group. To simplify the notation, we will present the case where σ_ϵ is normalized to 1.

Fix (x, z_l), $H(\cdot)$ only varies with d, so we can suppress the dependence on (x, z_l) and let $g(\kappa_j, w_j, x_i) = g_j$ such that

$$H(d) = \lambda \Phi\left(\frac{(p_1 g_1 - p_2 g_2) - (p_1 C(d_1) - p_2 C(d_2))}{\sqrt{p_1^2 + p_2^2}}\right) + (1-\lambda)\Phi\left(\frac{(g_1 - g_2) - (C(d_1) - C(d_2))}{\sqrt{2}}\right). \tag{7}$$

3.1.1 Identification of $g(\cdot)$ and $\lambda(\cdot)$

The following theorem shows that for any fixed (x, z_l), $g(\kappa_j, w_j, x)$ and $\lambda(x, z_l)$ are identified.

Theorem 1 *Assume that 1) $\lambda \in (0,1)$, 2) there exists an open set $D^* \subseteq D$ such that for $d_{ij} \in D^*$, $C'(d_{ij}) \neq 0$. Then the parameters $\theta = [g_1, g_2, \lambda]'$ in (7) are locally identified from the observed application decisions.*

Proof. The proof draws on the well-known equivalence of local identification with positive definiteness of the information matrix. In the following, I will show the positive definiteness of the information matrix for model (7).
Step 1. Claim: *The information matrix $I(\theta)$ is positive definite if and only if there exist no $\omega \neq 0$, such that $\omega' \frac{\partial H(d)}{\partial \theta} = 0$ for all d.*
The log likelihood of an observation (y, d) is

$$L(\theta) = y \ln(H(d)) + (1-y)\ln(1 - H(d)).$$

The score function is given by

$$\frac{\partial L}{\partial \theta} = \frac{y - H(d)}{H(d)(1 - H(d))}\frac{\partial H(d)}{\partial \theta}.$$

Hence, the information matrix is

$$I(\theta|d) = E\left[\frac{\partial L}{\partial \theta}\frac{\partial L}{\partial \theta'}|d\right] = \frac{1}{H(d)(1 - H(d))}\frac{\partial H(d)}{\partial \theta}\frac{\partial H(d)}{\partial \theta'}.$$

Given $H(d) \in (0,1)$, it is easy to show that the claim holds.
Step 2. Show $\omega' \frac{\partial H(d)}{\partial \theta} = 0$ for all $d \Longrightarrow \omega = 0$.
Define $p_j^* = \frac{p_j}{\sqrt{p_1^2 + p_2^2}}$, $B_1(d) = (p_1^* g_1 - p_2^* g_2) - (p_1^* C(d_1) - p_2^* C(d_2))$, and $B_0(d) = \left(\frac{(g_1 - g_2) - (C(d_1) - C(d_2))}{\sqrt{2}}\right)$,

$\frac{\partial H(d)}{\partial \theta}$ is given by:

$$\frac{\partial H(d)}{\partial \lambda} = \Phi\left(B_1\left(d\right)\right) - \Phi\left(B_0\left(d\right)\right)$$
$$\frac{\partial H(d)}{\partial g_1} = \lambda \phi(B_1\left(d\right))p_1^* + \left(1-\lambda\right)\phi(B_0\left(d\right))\frac{1}{\sqrt{2}}$$
$$\frac{\partial H(d)}{\partial g_2} = -\lambda \phi(B_1)p_2^* - \left(1-\lambda\right)\phi(B_0)\frac{1}{\sqrt{2}}.$$

Suppose for some ω, $\omega' \frac{\partial H(d)}{\partial \theta} = 0$ for all d :

$$\omega_1[\Phi(B_1) - \Phi(B_0)] + \omega_2\left(\lambda\phi(B_1)p_1^* + \left(1-\lambda\right)\phi(B_0)\frac{1}{\sqrt{2}}\right)$$
$$-\omega_3\left(\lambda\phi(B_1)p_2^* + \left(1-\lambda\right)\phi(B_0)\frac{1}{\sqrt{2}}\right) = 0$$

Take derivative with respect to d_2 evaluated at some $d_2 \in D^*$

$$\omega_1[\phi(B_1)p_2^* - \frac{\phi(B_0)}{\sqrt{2}}]C'\left(d_2\right) + \omega_2\left(\lambda\phi'(B_1)p_1^*p_2^* + \left(1-\lambda\right)\phi'(B_0)\frac{1}{2}\right)C'\left(d_2\right) \tag{8}$$
$$-\omega_3\left(\lambda\phi'(B_1)\left(p_2^*\right)^2 + \left(1-\lambda\right)\phi'(B_0)\frac{1}{2}\right)C'\left(d_2\right) = 0.$$

Define $\gamma\left(d\right) = \frac{\phi(B_1)}{\phi(B_0)}$, divide (8) by $\phi(B_0)$:

$$\omega_1[\gamma\left(d\right)p_2^* - \frac{1}{\sqrt{2}}] - \omega_2\left(\lambda B_1\gamma\left(d\right)p_1^*p_2^* + \left(1-\lambda\right)B_0\frac{1}{2}\right)$$
$$+\omega_3\left(\lambda B_1\gamma\left(d\right)\left(p_2^*\right)^2 + \left(1-\lambda\right)B_0\frac{1}{2}\right) = 0$$

$$\gamma\left(d\right)[\omega_1p_2^* - \lambda B_1p_2^*(\omega_2p_1^* - \omega_3p_2^*)] - [\frac{\omega_1}{\sqrt{2}} + \left(\omega_2 - \omega_3\right)\left(1-\lambda\right)B_0\frac{1}{2}] = 0 \tag{9}$$

Since $\gamma(d)$ is a nontrivial exponential function of d, (9) hold for all $d \in D^*$ only if both terms in brackets are zero for each $d \in D^*$, i.e.

$$\omega_1p_2^* - \lambda B_1\left(d\right)p_2^*(\omega_2p_1^* - \omega_3p_2^*) = 0 \tag{10}$$
$$\frac{\omega_1}{\sqrt{2}} + \left(\omega_2 - \omega_3\right)\left(1-\lambda\right)B_0\left(d\right)\frac{1}{2} = 0.$$

Take derivative of (10) again with respect to d_2, evaluated at $d_2 \in D^*$:

$$-\lambda C'(d_2)(p_2^*)^2(\omega_2 p_1^* - \omega_3 p_2^*) = 0$$
$$(\omega_2 - \omega_3)(1-\lambda) C'(d_2) \frac{1}{2\sqrt{2}} = 0.$$

Since $\lambda \in (0,1)$, $p_j > 0$ (hence $p_2^{*2} > 0$) and $C'(d_2) \neq 0$ for some d, we have

$$\omega_2 p_1^* - \omega_3 p_2^* = 0$$
$$\omega_2 - \omega_3 = 0.$$

Given $p_1 \neq p_2$ (hence $p_1^* \neq p_2^*$), follows that $\omega = 0$. ■

3.1.2 Identification of $C(d_{ij})$ and $(\tau, \delta, \kappa, \rho)$

1) Given the identification result from B2.1, and given that $C(d_{ij})$ is common across (x, z_l)'s, the parameters in $C(d_{ij})$ solves for the system of equations (7), where one equation corresponds to one (x, z_l).
2) Once the value of $g(\kappa_j, w_j, x_i)$ is identified for each (j, x_i), the parameters governing $g(\cdot)$ are identified using the variations in (w_j, x_i). In particular, for the middle-education group $\delta_{02} = 0$, $\delta_{12} = 1, \alpha_2 = 0$ are normalized, which means $g(\kappa_j, w_j, x_i) = \tau_1 x_{i4} + \tau_2[I(x_{i5} = j) - I(x_{i5} = 0)] + \kappa_j$. As a result, $\left(\tau, \{\kappa_j\}_j\right)$ are identified using the variation of (x_{i4}, x_{i5}) within the middle-education group, so the value $(\delta_{0e} + \delta_{1e}\zeta_j + w_j\alpha_e)$ is known for $e = 1, 3$. The variation of w_j thus identifies (δ_e, α_e).

3.2 A Model with Fewer Strategic Households

To obtain further insights into our identification, we have re-estimated our model imposing that a lower fraction of households be strategic. In particular, we have re-estimated all model parameters other than the type distribution parameters, which are restricted such that 80% of households are strategic.[13] The new estimates are reported in Tables B5 and B6. The differences between the original parameter estimates and those from this new model are intuitive. For example, the cost of distance becomes more convex and the taste dispersion increases. These findings are intuitive, schools with shorter distance are more likely to be in-zone (high-priority) schools; when fewer households are strategic, preferences for shorter distance need to be stronger in order to explain why households listed in-zone schools. Similarly, larger taste dispersion helps explain why they left out schools that look

[13] In particular, we adjusted the constant in type distribution $\lambda(x_i, l_i)$ such that 80% of all households are strategic. The fractions by education level from low to high are 72%, 77% and 87% respectively.

better than those on their application lists. Table B7 shows the in-sample (2006) fit of this model alongside with the original model fit. Given that the new model has fewer degrees of freedom, the fit is naturally worse. However, the comparison reveals some intuition behind the identification of our model. For example, with fewer strategic households, the new model over-predicts the fraction of households top-listing schools that are not in their zones (Row 2); similarly, it underpredicts the fraction of households assigned in Round 1 and overpredicts the fraction unassigned.

3.3 The Exclusion Restriction

Although types are usually unobservable, they are observed for the group of obviously naive households. We utilize this data feature to provide some supporting evidence for our assumption that distance enters preferences but not type distribution (conditional on zone characteristics). In particular, we test whether or not obvious naivety is correlated with distance measures. At least, the two sets of regressions suggest that our exclusion restriction may not be unreasonable.

First, we run the following regression

$$naive_i = \alpha_1 d_i^{near} + \alpha_2 q_i^{near} + \alpha_3 n_i^{500m} + \alpha_4 n_i^{1km} + \alpha_5 n_i^{2km} + \alpha_6 d_i^{ave} + X_i B_1 + Z_i B_2,$$

where $naive_i = 1$ if i is obviously naive, d_i^{near} is the distance to the school that is closest to i, q_i^{near} is the quality of this school. $n_i^{500m}, n_i^{1km}, n_i^{2km}$ are the number of schools within 500 meters, 1 km and 2 km, respectively. d_i^{ave} is the average school-home distance across all in-zone schools for i. X_i is the vector of household observable characteristics, and Z_i is the vector of characteristics of the zone i lives in. As Table B8 shows, none of the α coefficients are significant, nor can one reject the hypothesis that they are jointly insignificant (Prob > F = 0.7825). That is, naive households are not more or less likely to live closer to a school (α_1) or closer to a better school (α_2) , to have more schools nearby (α_3 to α_5), or to have lower average distance to in-zone schools (α_6).

Second, we run three regressions of the following form

$$y_i = \beta_1 I\left(naive_i = 1\right) + X_i A_1 + Z_i A_2.$$

The dependent variables in the three regressions are 1) whether or not i top-listed the closest school, 2) whether or not i included the closest school in the application list, and 3) whether or not i top-listed an in-zone school. The Results are shown in Table B9 in the appendix. β_1 is insignificant in 1) and 2), but significantly negative in 3). That is, naive households are less likely to top-list an in-zone school, for which one has higher priorities, as is consistent with their naive labels. However, they do not seem to exhibit different tastes for distance.

4 Data Details

4.1 Priority Score Structure

Case 1: Those who do not have a sibling in school have two levels: $x_i a$ ($x_i a + b_1$) for out-of-zone (in-zone) schools.
Case 2: Those whose sibling(s) is (are) in in-zone schools have 3 levels: $x_i a$ ($x_i a + b_1$) for out-of-zone (in-zone) non-sibling schools, and $x_i a + b_1 + b_2$ for sibling schools.
Case 3: Those whose sibling(s) is (are) in out-of-zone schools have 3 levels: $x_i a$ ($x_i a + b_1$) for out-of-zone (in-zone) non-sibling schools, and $x_i a + b_2$ for sibling schools.
Case 4: Those with sibling(s) in some in-zone school and sibling(s) in some out-of-zone school have 4 levels: $x_i a$ ($x_i a + b_1$) for out-of-zone (in-zone) non-sibling schools, and $x_i a + b_2$ ($x_i a + b_1 + b_2$) for out-of-zone (in-zone) sibling schools.

4.2 Data Sources

The final data set consists of merged data sets from five different administrative units: the *Consorci d'Educacio de Barcelona* (local authority handling the choice procedure in Barcelona), *Department d'Ensenyament de Catalunya* (Department of Education of Catalunya), the *Consell d'Avaluacio de Catalunya* (public agency in charge of evaluating the Catalunya educational system), the *Instituto Nacional de Estadistica* (national institute of statistics) and the *Institut Catala d'Estadistica* (statistics institute of Catalunya).[14]

From the *Consorci d'Educacio de Barcelona,* we obtain access to every applicant's application form, as well as the information on the school assignment and enrollment outcomes. An application form contains the entire list of ranked schools a family submitted. In addition, it records family information that was used to determine the priority the family had for various schools (e.g., family address, the existence of a sibling in the first-ranked school and other relevant family and child characteristics). The geocode in this data set allows us to compute a family's distance to each school in the city.

From the Census and local register data, we obtain information on the applicant's family background, including parental education and whether or not both parents were registered in the applicant's household. Since information on siblings who were not enrolled in the school the family ranked first is irrelevant in the school assignment procedure, it is not available from the application data. However, such information is relevant for family's

[14] These five different data sources were merged and anonimized by the Institut Catala d'Estadistica (IDESCAT).

application decisions. From the Department of Education, we obtained the enrollment data for children aged 3 to 18 in Catalunya. This data set is then merged with the local register, which provides us with the ID of the schools enrolled by each of the applicant's siblings at the time of the application.

To measure the quality of schools, we use the external evaluation of students conducted by *Consell d'Avaluacio de Catalunya.* Since 2009, such external evaluations have been imposed on all schools in Catalunya, in which students enrolled in the last year of primary school are tested on math and language subjects.[15] From the 2009 test results that we obtained, we calculated the average test score across subjects for each student, then use the average across students in each school as a measure of the school's quality.[16] Finally, to obtain information on the fees charged by semi-public schools (public schools are free to attend), we use the survey data collected by the *Instituto Nacional de Estadistica.*[17]

4.3 Data Evidence on Strategic Behavior via Regression Analyses

Using each household-school pair as an observation, we run regressions of the following form

$$y_{ij} = X_i\alpha + sib_{ij}\beta_1 + d_{ij}\beta_2 + d_{ij}^2\beta_3 + d_{ij}^3\beta_4 + I\left(l_i \in z_{l_j}\right)\beta_5 + \mu_j + \varepsilon_{ij},$$

where $y_{ij} = 1$ if j was i's top choice, X_i are household characteristics, $sib_{ij} = 1$ if i has a sibling in School j, d_{ij} is school-home distance, $I\left(l_i \in z_{l_j}\right) = 1$ if i lives in School j's zone, μ_j is a school fixed effect, and ε_{ij} is the error term. The results are shown in Table B3. Conditional on a third-order polynomial of distance, in-zone status still significantly increases the choice probability.[18]

[15] As mentioned in the background section, a student has the priority to continue her primary-school education in the same school (with the same capacity) she enrolled for preschool education, which makes it very unlikely that one can transfer to a better school between preschool-primary school transition. For example, at least 94% of the 2010 preschool cohort were still enrolled in the same school for primary school education in 2013.

[16] Following the same rule used in Spanish college admissions, we use unweighted average of scores across subjects for each student.

[17] See http://www.idescat.cat/cat/idescat/publicacions/cataleg/pdfdocs/dossier13.pdf for a summary of the survey data.

[18] There are 317 observations per household in these regressions. The results are robust when we restrict attention to a smaller number of schools for each household (which always include the applied schools), e.g., the closest N schools, N randomly drawn schools, all in-zone schools plus N randomly drawn out-of-zone schools ($N = 30, 50, 100$).

5 The DA and TTC Algorithms

5.1 DA

The DA algorithm assigns students as follows.
Round 1: Each school j tentatively assigns its seats to students who top-listed it, one at a time following their priority order. If school j is over-demanded, lower-ranked applicants are rejected.
In general, at Round r: Each school j considers the students it has been holding, *together with* students who were rejected in the previous round but listed j as their r^{th} choice. Seats in school j are tentatively assigned to these students, one at a time following their priority order. If school j is over-demanded, lower-ranked applicants are rejected.
The algorithm terminates when no student is rejected and each student is assigned her final tentative assignment.

The key differences between DA and BM are 1) in each round, students are only temporarily assigned to a school until the whole procedure ends; and 2) temporarily held students are considered based only on priorities along with students who were rejected from their choices in previous rounds and added into a school's student pool in the current round. As such, a previously held student can be crowded out by a newly-added student who has higher priority. That is, top-listing a school does not improve one's chance of being finally admitted to this school, which makes truth-telling a (weakly) dominant strategy for households under DA. Moreover, DA eliminates justified envy. The appealing properties of DA, however, may conflict with Pareto efficiency, as shown by Abdulkadiroğlu and Sönmez (2003).

5.2 TTC

The TTC algorithm assigns students as follows.
Round 1: Assign a counter for each school which keeps track of how many seats are still available at the school, initially set to equal the school capacity. Each school points to the student who has the highest priority for the school. Each student points to her favorite school *under her announced preferences*.[19] This will create ordered lists of distinct schools (j) and distinct students (i) : $(j_1, i_1, j_2, i_2, ...)$, where j_1 points at i_1, i_1 points at j_2, and j_2 points at i_2, etc. Because there are finite number of schools, at least one cycle will be formed, where i_k ($k \geq 1$) points at j_1. Although there may be multiple cycles formed in a round, each school can be part of at most one cycle and each student can be part of at most one cycle. Every student in a cycle is assigned a seat at the school she points to and is

[19] A student announces her entire list of schools before the assignment starts. As such, the "pointing" by a student is mechanically following her announced list.

removed. The counter of each school in a cycle is reduced by one and if it reduces to zero, the school is also removed. Counters of all schools that are not in any cycle stay put.
In general, at Round r: Each remaining school points to the student with highest priority among the remaining students and each remaining student points to her favorite school among the remaining schools. Every student in a cycle is assigned a seat at the school that she points to and is removed. The counter of each school in a cycle is reduced by one and if it reduces to zero the school is also removed. Counters of all other schools stay put.
The algorithm terminates when all students are assigned a seat.

Intuitively, in each round TTC creates cycles of trade between individuals. Each individual in a cycle trades off a seat in her highest-priority school for a seat in her most preferred school among those that still have seats. Whenever such a cycle is formed the allocation is final. Hence, the only way for an individual to improve her allocation is through "stealing" another individual's school assignment, which will in turn make this other individual worse off. As such, TTC is Pareto efficient as shown by Abdulkadiroğlu and Sönmez (2003), who also prove that TTC is truth-revealing. However, TTC does not eliminate justified envy because student-school priorities are ignored in the TTC trade between individuals

5.3 BM may respect cardinal preference better than truth-revealing mechanisms

The intuition can be explained by the following simple example with equal priorities. Consider three schools and a set of households who share the same ordinal but different cardinal preferences for these schools, where the schools are ranked from high to low as Schools 1, 2 and 3. Under BM, the strategic decision is whether to take the high risk and top-list School 1 or to play it safe and top-list School 2. Given the same evaluation for School 1, a household whose evaluations for Schools 2 and 3 are similar is more likely to choose the risky strategy because it has less to lose from the gamble. Given the same evaluation for School 3, a household that values School 1 much higher than School 2 is more likely to choose the risky strategy because it has more to gain from the gamble.

6 Welfare under a Standard BM

Following a referee's suggestion, we have simulated an equilibrium under a standard BM, i.e., BM with school-specific priorities and a single lottery number. The welfare comparison between the standard BM, DA and TTC is shown in Table B10. The results are similar to those shown in Table 11 of the paper. The welfare calculation under a standard BM carries two caveats. C1: in computing a standard BM, we hold the distribution of household (non)strategic types invariant. C2: BM may admit multiple equilibria. We search for an equilibrium numerically, with initial guesses set at the current Barcelona equilibrium.

The counterfactuals in the main text are free from these caveats. First, when we replace the mechanism in Barcelona with truth-revealing mechanisms, all households, strategic or not, will rank schools according to their true preferences. We can compare the current regime with truth-revealing alternatives without worrying about whether or not (non)strategic type distribution is policy variant because type does not matter when all households tell the truth. However, switching to another manipulable mechanism, e.g., a standard BM, we have to make assumptions like C1. Second, although BM may admit multiple equilibria, we observe the equilibrium in the data. We take the realized equilibrium as given and estimate household preferences and type distribution. Then we can make a very clean comparison between the counterfactual truth-revealing mechanisms and the current Barcelona equilibrium.

7 Additional Tables

7.1 Data

Table B1 Prob of Admission to one's First Choice $p^1_{ia_1}(S_i)$

	Enrolled	Opted out
All Households	91.8%	75.3%
Assigned within 10 Rounds	92.7%	86.8%
Unassigned within 10 Rounds	48.5%	44.8%

Admission prob in Round 1, averaged for each group of households.

Table B2 "Better" Schools Than the Top-Listed One

	% Households	# Better Sch	%Better w/ Higher p
All Households (6836)			
Have Sch. Better in Quality, Fees	97.5%	66.6 (44.3)	9.1%
Have Sch. Better in Quality, Dist	41.1%	4.7 (10.6)	14.5%
Have Sch. Better in Fees, Dist	62.2%	14.3 (30.6)	24.3%
Sib School not Top-listed (4025)			
Have Sch. Better in Quality, Fees	98.2%	69.4 (44.2)	15.4%
Have Sch. Better in Quality, Dist	39.6%	4.3 (9.1)	25.7%
Have Sch. Better in Fees, Dist	60.2%	12.4 (27.0)	42.7%

% Households: % of households that satisfy the condition specified in each row.

#Better Sch: average (std.dev.) num. of better schools for households with such schools.

%Better w/ higher p: % of better sch with higher admission prob. than one's top choice.

Table B3 Evidence on Strategic Behavior: Regression

In-zone	Distance	Distance^2	Distance^3
0.0165	-0.0187	0.0038	-0.0002
(0.0001)	(0.0001)	(0.00003)	(0.00001)

Regression controls for school fixed-effect, sibling school, and household characteristics.

Table B4 School Filled Round*(%)

	2006	2007
1	44.2	46.4
2	7.6	11.4
≥ 3	8.2	6.0
Leftover	40.0	36.3

*A school is filled in Round r if it has open seats in rounds 1 to r, but not in later rounds.

7.2 Model Contrast: Original vs 80% Strategic

Table B5 Structural Preference Parameters

	Model^a	80% strategic^b
Distance^2	-0.05	-0.06
Distance>5 (100m)	-55.3	-59.8
Distance>10 (100m)	-46.5	-45.4
Sibling School	1339.0	1164.9
Single Parent	-404.3	-253.6
σ_ϵ(taste dispersion)	66.3	70.4
σ_η(post-app shock)	1937.8	1965.7

^{a} original model, ^{b}80% households are strategic

Table B6 Preference Parameters

Summarize School FE[b]						
By Edu	Edu < HS		Edu = HS		Edu > HS	
	Model	80% Strategic	Model	80% Strategic	Model	80% Strategic
Constant	2766.3	2846.1	2783.3	2749.1	2423.0	2272.6
Quality	152.0	163.2	176.4	187.1	187.8	201.3
Fee	-1.0	-1.0	-0.6	-0.6	-0.3	-0.4
Semi-Public	-0.6	-1.4	6.5	6.4	0.8	0.5
Common for all		Model			80% Strategic	
Capacity		0.7			0.8	
Capacity^2		-0.001			-0.001	
Quality^2		-9.9			-10.7	

[b]OLS regression of the estimated school value parameters on observables.

Table B7 Model vs. 80% Strategic (2006)

	Data	Model	80% Strategic*
Top-Listed Schools			
In Zone 06 Only (%)	24.1	24.1	23.1
In Zone 07 Only (%)	3.0	4.5	5.1
Quality	7.9	7.9	7.9
Distance (100m)	7.1	7.2	7.4
Fee (100 Euros)	8.1	8.1	8.0
Assignment Round (%)			
1	93.0	91.3	88.2
2	2.8	4.0	3.9
≥ 3	1.5	0.9	1.2
Unassigned	2.7	3.8	6.7
Enrollment in the Public System (%)			
All	96.7	96.5	94.9
Assigned in R1	97.8	97.1	96.7

*Adjust the fraction of strategic household to 80% by changing the constant term in the type distribution function, all parameters are re-estimated other than the type distribution parameters..

7.3 The Exclusion Restriction

Table B8 Naivety and Distance: Part 1

	Naive=1
Dist_nearest	0.006 (0.016)
Q_nearest	-0.0015 (0.0035)
$N\_500m * 100$	0.007 (0.114)
$N\_1km * 100$	0.049 (0.072)
$N\_2km * 100$	-0.032 (0.032)
Dist_ave	-0.013 (0.011)

Other controls: education, single parent, I(have sibling), zone characteristics

Table B9 Naivety and Distance: Part 2

	Top-Closest	Apply-Closest	Top-Inzone
Naive	0.033 (0.040)	-0.003 (0.044)	-0.15 (0.03)
Edu=mid	-0.047 (0.013)	-0.012 (0.015)	-0.024 (0.010)
Edu=high	-0.073 (0.013)	-0.027 (0.015)	-0.015 (0.010)

Other controls: single parent, I(have sibling), zone characteristics

7.4 Standard BM

Table B10 Household Welfare: Standard BM vs. DA vs. TTC

%	DA-Std BM[a]		TTC-Std BM[b]	
	Δutils	Δ100 euros	Δutils	Δ100 euros
All	-5.7	-11.6	1.6	3.2
Edu< HS	-10.3	-10.4	0.1	0.1
Edu= HS	-7.5	-24.2	0.7	2.2
Edu> HS	-1.2	-2.2	3.6	6.4

[a]change from Std BM to DA, [b]change from Std BM to TTC

Δutils: welfare change in utils. Δ100 euros: welfare change in 100 euros.

References

Chen, Y. and O. Kesten (2017) "Chinese College Admissions and School Choice Reforms: A Theoretical Analysis," *Journal of Political Economy,* 125 (1): 99-139.

Abdulkadiroğlu, A., Y. Chen and Y. Yasuda (2015) “Expanding “Choice” in School Choice”, *American Economic Journal: Microeconomics* 7, 1-42.