

College choice, selection, and allocation mechanisms: A structural empirical analysis

JOSÉ RAIMUNDO CARVALHO
CAEN, Universidade Federal do Ceará

THIERRY MAGNAC
Toulouse School of Economics, University of Toulouse Capitole

QIZHOU XIONG
Faculty of Economics and Management, OVGU Magdeburg and Financial Markets, IWH

We use rich microeconomic data on performance and choices of students at college entry to analyze interactions between the selection mechanism, eliciting college preferences through exams, and the allocation mechanism. We set up a framework in which success probabilities and student preferences are shown to be identified from data on their choices and their exam grades under exclusion restrictions and support conditions. The counterfactuals we consider balance the severity of congestion and the quality of the match between schools and students. Moving to deferred acceptance or inverting the timing of choices and exams are shown to increase welfare. Redistribution among students and among schools is also sizeable in all counterfactual experiments.

KEYWORDS. Education, two-sided matching, school allocation mechanism, policy evaluation.

JEL CLASSIFICATION. C57, D47, I21.

1. INTRODUCTION

The *matching* literature provides analyses of mechanisms allocating goods or relationships between many parties in the absence of a price mechanism, and examples range from kidney exchange and marriage to school choice (see Roth and Sotomayor (1992),

José Raimundo Carvalho: josecarv@ufc.br
Thierry Magnac: thierry.magnac@tse-fr.eu
Qizhou Xiong: qizhou.xiong@iwh-halle.de

This is a much revised version of a previous paper entitled “College Choice and Entry Exams” by two of the coauthors that has been circulated since 2009. We thank the two anonymous referees for their insightful suggestions. Useful talks and interactions with Yinghua He, Philipp Heller, Jean-Marc Robin, Bernard Salanié, and comments by participants at conferences in Brown, Bristol, Atlanta, Northwestern, Shanghai, Rio de Janeiro, IAAE’15, and San Francisco as well as seminars at Oxford, CREST, CEMMAP, Cambridge, Amsterdam, Barcelona, Manchester, and Bern are gratefully acknowledged. This research has received financial support from CNPq (Project 21207) and the European Research Council under the European Community’s Seventh Framework Program FP7/2007–2013 grant agreement N°295298. The usual disclaimer applies.

Roth (2008)). The analysis of centralized mechanisms in school choice as a many-to-one match has been very popular in the recent theoretical and empirical literature (for instance, Abdulkadiroğlu, Agarwal, and Pathak (2017), Agarwal (2015), Azevedo and Leshno (2016), Budish and Cantillon (2012), Calsamiglia, Fu, and Güell (2018), Chen and Kesten (2017), He (2017), Agarwal and Somaini (2018) among others) and has had practical value for policy implemented in primary or high schools in various countries.

College choice adds the new dimension of elicitation of college preferences over students which is of secondary importance in primary and high school choice. This elicitation process costs time and money because of congestion if application costs are low. This market friction is large when the allocation is decentralized as in the US (Che and Koh (2016) and Chade, Lewis, and Smith (2014)) but not only. Even with centralized mechanisms, for instance, used by universities in China (Chen and Kesten (2017)) or Turkey (Balinski and Sönmez (1999)), it is costly to organize a general exam whose results determine students' ranking while keeping up with the quality of selection in decentralized systems. This is why these exams are generally composed of proofs in different fields (maths, literature, etc.) and at times consist of two stages. The first stage selects out students at minimal costs while the second stage allows for a costly but more precise evaluation (Hafalir, Hakimov, Kübler, and Kurino (2018)).

In this paper, we analyze the interactions between allocation mechanisms and selection when college choice is centralized. College preferences are not taken as granted as in the matching literature and they are costly to elicit. We exploit an admittedly specific college choice experiment in order to quantify some of the trade-offs that matching and selection involve. Our observational "experiment" uses observed entry exam grades and choices between schools within a Federal university in Brazil in 2004.

A mechanism called *Vestibular* was in place at this university and worked as follows. During their last high school year, students chose a single specialization field or "school"¹ before taking a two-stage exam at the end of high school. The first stage is a cost-minimizing multiple question exam common to all fields and selects but the top-ranked students for a more in-depth and specialized second-stage exam. Aggregating scores of both exams yield the final rankings and admissions into each school.

This paper aims at evaluating the effects, on student allocations and their welfare, of changing the allocation mechanism and the selection device with respect to the existing *Vestibular*. In the absence of experiments (Calsamiglia, Haeringer, and Klijn (2010)) or quasi-experiments (Pathak and Sönmez (2013)), estimating a structural model is key to our empirical strategy. We construct such a model of college choice, exhibit conditions under which parameters are identified, and derive empirical counterfactual results on outcomes and welfare.

The paper makes three original contributions.

Our first contribution is to adopt a two-step empirical strategy that uses first information on performance at the two-stage exams to estimate success probabilities at each school. Second, we estimate preference parameters from observed school choices when students play strategically by taking into account their expected probabilities of success

¹We use the terms "college" and "school" interchangeably for these fields in the following.

(Arcidiacono (2005), Epple, Romano, and Sieg (2006)). As far as we know, the previous empirical literature does not estimate school choice models in which students face uncertainty about their entrance exam scores. This is permitted by our rich data on exam scores as well as on an initial measure of ability obtained a year before the exams are taken.

Our second original contribution is to derive conditions under which expected success probabilities of entry and preferences are identified from observing the distribution of grades and college choices. Students play a “congestion” game in which choices of other students affect their own success probabilities. We adopt specific and admittedly strong assumptions to solve this game. The solution concept we use is a Nash equilibrium. Students have symmetric information about random shocks on grades, that is, they know their distribution functions only and the information set of students and econometricians is the same. We also assume that expectations are perfect in the sense that they can be obtained by infinitely repeating the game with the same players. We justify these assumptions in the specific context of our empirical application.

We show that the distribution of success probabilities can be obtained by resampling in our observed sample and by using Nash equilibrium conditions. We derive from the latter, grade thresholds for being admitted in a specific college at each exam stage and show that success probabilities are identified. We then provide a proof of nonparametric identification of preference parameters by using, as in Matzkin (1993), exclusion restrictions and conditions that success probabilities fully vary over the simplex. This proof of identification specifically deals with two prevalent issues in college choice. First, data are likely to be choice-based. Second, outside options play a much more important role than in school choice (Agarwal and Somaini (2018)) since the number of candidates is well above the number of seats, by a factor of 15 in our data.

Our third original contribution is to analyze the aggregate and distributive effects on the allocation and welfare of students and schools of three different counterfactual mechanisms that play with the trade-offs between congestion costs, the adequacy of student selection, and the quality of the resulting match. These three experiments aim at analyzing salient policy issues in the current debates on school choice (Roth (2018)).

In the first experiment, we restrict the number of seats available at the second-stage exam. We argue that it reduces screening costs for schools at the risk of degrading selection of adequate students. The counterfactual effect on matching quality we obtain is, however, small. In the second experiment, students are allowed to submit a list of two choices instead of a single one in order to get closer to a Gale–Shapley deferred acceptance mechanism. It indeed results in a positive aggregate effect in terms of utilitarian social welfare though it also has distributive effects. Strategic effects in the original mechanism are shown to be sizeable. Interchanging the timing of choices and the first exam is the basis of our third counterfactual experiment. We allow students to choose colleges after passing the first-stage exam instead of having them to choose before this exam. This allocation mechanism is quite popular, as in Japan for instance (Hafalir et al. (2018)). As expected, it has strong redistributive effects between schools and between students since it favors more opportunistic behavior.

Related literature

This paper touches different strands of the matching and school choice literature.

Analyzing the matching of students to schools has a long history and a brief survey of the recent literature in which differences between school choice, college admission, and student placement are rigorously defined is available in [Sönmez and Ünver \(2011\)](#). Prominently in this literature, Gale–Shapley deferred acceptance mechanisms that satisfy both properties of stability and strategy-proofness (on the student side if students propose) if preferences are strict (e.g., [Abdulkadiroğlu and Sönmez \(2003\)](#)). If such a mechanism is used, the elicitation process through which schools decide on their ranking of students has very little impact on the preference lists submitted by students. The use of deferred acceptance mechanisms, however, could involve larger congestion costs than other nonstable mechanisms ([He and Magnac \(2018\)](#)) such as deferred acceptance with a truncated list of preferences as is the case with the mechanism we study in this paper. The truncation is severe since the list of schools is of length one.

The seminal analysis of college admissions by [Balinski and Sönmez \(1999\)](#) was theoretical albeit oriented toward the analysis of a specific mechanism. They studied the optimality of student placement in Turkish universities in which selection and competition among students are nationwide unlike our case. Students choose a rank-ordered list of colleges prior to writing exams in various subjects. Student rankings are constructed using exam grades, and are allowed to differ across colleges by weighting subjects differently. Grades in mathematics are given more weight by math schools.

Most of the empirical literature on matching, however, is concerned by primary or high school choices. [Abdulkadiroğlu, Pathak, and Roth \(2009\)](#) studied the mechanisms used in the New York high school system and focused on the trade-offs between efficiency, strategy-proofness, and stability. This research line on primary and secondary schools questions the relative standing of the Gale–Shapley and the Boston mechanisms ([Abdulkadiroğlu, Pathak, Roth, and Sönmez \(2006\)](#)). Others analyze the Boston mechanism as [He \(2017\)](#) who uses school allocation data from Beijing and evaluates the cost of strategizing for sophisticated and naive agents. The question of the importance of truncated lists of preferences used in practice in deferred acceptance mechanisms is high on the agenda in recent research about middle or high school choices ([Calsamiglia, Fu, and Güell \(2018\)](#), [Fack, Grenet, and He \(2017\)](#)).

School and college choice, however, differ in a number of dimensions and the questions set out in this paper are more specific to college choice. College preferences over students depend on their past investments in human capital and abilities and not only on priorities given by residence and siblings. This implies in particular that colleges have strict preferences over students and the arguments underpinning the debate between the choice of allocation mechanisms such as deferred acceptance and Boston can be misleading for college choice. Furthermore, demand for colleges is not localized and is much larger than supply.

In the most recent literature, demands for colleges are estimated in [Hastings, Kane, and Staiger \(2009\)](#) to study how enlarging choice sets might have unintended consequences for minority students. In [Agarwal \(2015\)](#), medical schools and medical residents

preferences are estimated using a two-sided school choice model. Fu (2014) estimates demand and supply equations when students have heterogeneous abilities and preferences and when college applications are costly and uncertain. Akyol and Krishna (2017) also estimated a structural model of high school choice using Turkish data in order to understand whether the higher standing of elite schools is due to selection or to value added. As the allocation mechanism in place is deferred acceptance, preferences can be directly estimated from rank-ordered lists. It is remarkable that they find that estimates of value added are small.

To our knowledge, there is no comparative survey of college admission procedures in different countries. There exist empirical papers about the “parallel” mechanism used in China (Chen and Kesten (2017) or Zhu (2014)) or descriptive analyses in Turkey (Dogan and Yuret (2013)) or in Egypt (Selim and Salem (2009)). Abizada and Chen (2011) analyzed the eligibility restrictions to college access that gives a way of reducing costs of evaluation of students by colleges. A descriptive analysis of the mechanism centralized at the level of the country, which has been used in Brazil since 2010, is provided by Aygün and Bo (2017) and Machado and Szerman (2017).

The most obvious distinction between college admission procedures is their degree of centralization. Decentralized models of college choices, as in the United States, are studied by Chade, Lewis, and Smith (2014), Che and Koh (2016), and Hafalir et al. (2018) among others. In the last paper, low and high ability students are shown to have different preferences over centralized and decentralized mechanisms and a small amount of literature about centralization is surveyed there. Congestion is reduced by either making students pay an application cost or by making them choose only one college. In Chade, Lewis, and Smith (2014), school preferences are noisy signals of students’ abilities and college strategizing can lead to inefficient sorting of students. The use of waiting lists might lead to unstable mechanisms. In Che and Koh (2016), the uncertainty of student preferences makes schools play strategically and this leads to inefficient and unfair assignments because the management of offers and acceptance of offers is uncertain and takes time.

Centralization may avoid the costs of congestion if colleges do not have to deal with all student files. It also streamlines the competition between colleges. Yet, centralization assumes that college preferences are adequately translated by the information revealed at a general exam (Hafalir et al. (2018)). In a decentralized system like in the US, many other elements than the SAT score are evaluated and the selection is multidimensional. The two-stage exam set-up tries to mitigate the reduction in selection quality. The selection mechanism used in our empirical illustration is broadly akin to the Japanese experience in which a first stage centralized exam is followed by a second-stage exam decentralized at the level of each university *on the same day* which effectively avoids congestion (see Hafalir et al. (2018)). The choice of college and the sequential exams are also akin to the system now in place in South Korea (Avery, Lee, and Roth (2014)). As a matter of fact, these two-step revelation procedures of school preferences are rather common (job market for PhDs, “grandes écoles” in France) although their interaction with the allocation mechanism is seldom studied in the literature (although see Lee and Schwarz (2017)).

Last, Agarwal and Somaini (2018) developed independently after us a proof of non-parametric identification of preferences in a school choice model. It either relies on exogenous variation in the environment, that is, in expected success probabilities, as in our case, or on the existence of a special regressor, such as distance to school and quasi-linearity of preferences. Their more-in-depth analysis of the latter case is specifically suited to school choice in primary and secondary schools, while our results bear on college choice. It is indeed more credible there that success probabilities continuously vary, for instance because of grades, than in the case of school choice in which only discrete priorities matter. Conversely, a special regressor such as distance is likely to be irrelevant in college choice. Overall, the intuition for both results is based on Matzkin (1992, 1993). We investigate more in-depth which preference functionals are identified when there is exogenous variation in the environment and specifically because of the presence of outside options and choice-based sampling.

The paper is organized in the following way. Section 2 describes our modeling assumptions for college choices, the formation of expectations, and the conditions under which preferences are identified. Section 3 presents the particulars of our empirical application, explains the estimation and computation of success probabilities and the estimation procedure of preference parameters. It also summarizes results from the estimated coefficients of grade and preference shifters. Section 4 details the results of the three counterfactual experiments. A Supplementary Appendix, available upon request, gathers the details and results of our many procedures.

2. THEORETICAL SET-UP

We start by describing a framework, encompassing our empirical application, in which we provide modeling tools and identification results. We abstract from some aspects of the empirical application, such as the two-stage nature of exams, that do not bear on general results and are clarified in the empirical Section 3.

The first subsection defines notation, formalizes the timing of events for students and describes the primitives of the decision problem and the observed variables. Students are assumed to play an imperfect information game in which information on future grades is imperfect but symmetric and its distribution known by agents. Students have no private information and we assume that the solution concept is Nash. In particular, observed characteristics and preference shocks of students are common knowledge. The construction of this set-up in terms of information sets and expectations is presented in the second subsection. We also derive the necessary conditions for a Nash equilibrium.

The final subsection provides conditions under which student preferences are identified.

2.1 *Timing for the decision maker*

First, we adopt a simplifying framework in which students choose, according to their preferences, one and only one school among many within the university to apply to, as

in our empirical application. Rank-ordered lists submitted by students are thus highly truncated and more so than in the empirical application of [Agarwal and Somaini \(2018\)](#) in which rank-ordered lists are of length three. The main reason for adopting such a setting is that it does not change the list of identified objects. We return to this point after stating our identification results. It is worth mentioning that this is akin to the identification results of [Agarwal and Somaini \(2018\)](#) which are insensitive to the allocation mechanism in place provided that certain conditions are satisfied (Definitions 1–3, pp. 407–408) excluding top-trading cycles. In this sense, having rank-ordered lists of length one is the minimal observational requirement for preferences to be identified. This also calls to mind that observing the ranking of alternatives in multinomial choices does not enlarge the set of identified objects but allows them to be more precisely estimated.

Second, student preferences can be monetary or nonmonetary and describe the consumption value of education ([Alstadsæter \(2011\)](#), [Jacob, McCall, and Stange \(2012\)](#)) as well as its investment value. The latter is derived from earnings that a degree from a specific school raises in the labor market.

We omit the individual index for readability. A random variable, say D , describes school choice and takes as realization, a specific school, j . The set of available schools is denoted by a discrete set of indices, \mathcal{J} , to which we add an outside option, $D = \emptyset$. We denote $J = \text{card}(\mathcal{J})$ the number of available schools. Observed student characteristics which affect preferences (resp., performance or grades) are denoted X (resp., Z) and variables X and Z can be overlapping.

We describe the assignment mechanism by a simple sequence of four steps. At each step, students obtain information or make decisions.

- *School capacities*: Every school announces the number of seats available or its capacity, n^j .

- *Choice of school*: Students apply to one and only one school among available options, $j \in \{\emptyset\} \cup \mathcal{J}$. The outside option $j = \emptyset$ means that one forfeits the opportunity to get into one of these schools and either chooses another university, searches for a job, or any other alternative (waiting until next year, staying at home). After that stage, students are allocated, according to their school choices, to J subsamples which are observed in our empirical application. We do not observe students who choose an external option and in this sense we have a choice-based sample.

- *Exam stage*: All students take a single exam or multiple exams, identical across schools, and exam grades are aggregated into a single grade denoted m and written as a function of characteristics, Z , as

$$m = m(Z, u; \beta)$$

in which u are random individual circumstances that affect results at these exams and β is an unknown parameter. College preferences are formed using these heterogeneous grades.

- *College entry*: In each subsample, defined by $D = j$, students are ranked according to their values of grade m and the first n^j students are accepted in school j that they have chosen previously. This selection can be expressed using a threshold, T^j , that describes the set of successful students by the condition, $m \geq T^j$ (as in [Azevedo and Leshno](#)

(2016)). Those who succeed receive a value, V^j , describing their preferences. Those who fail, get the value of their outside option that we normalize to 0. Individual rationality implies that j is never chosen if $V^j < 0$.

There could be additional decision nodes to consider if the value of outside options evolves over time because of the selection process. Students could leave the game after taking or passing exams because grades could give students a way to signal their ability to potential employers or other universities. This would modify the value of the outside option after the exam stage. This is why, in our empirical application, we select elite schools which are so attractive that almost no students quit after taking or passing exams.

Determining choices is now easy. Define the probability of success in school D as

$$P^D = \Pr(m(Z, u; \beta) \geq T^D),$$

in which we delay until next section the precise definition of the probability measure for random thresholds T^D since it depends on the definition of information sets and expectations. The expected value of choosing school D at the time of the choice is given by

$$\mathbb{E}V^D = P^D V^D,$$

and, as the outside option has value zero, choosing $j \in \mathcal{J}$ is described by the choice-based condition $\max_{k \in \mathcal{J}} (V^k) > 0$. Moreover, maximizing expected utility leads, for any $j \in \mathcal{J}$, to

$$D = j \quad \text{iff} \quad \max_{k \in \mathcal{J}} (V^k) > 0 \quad \text{and} \quad \forall k \in \mathcal{J}/\{j\}; \quad P^j V^j > P^k V^k. \quad (1)$$

We shall specify later on, values as functions $V^j(X, \varepsilon; \zeta)$ in which X are observed characteristics, ε is an unobservable preference random term and ζ are preference parameters. It is enough at this stage to define choices as $D(X, \varepsilon, \zeta, \{P^j\}_{j \in \mathcal{J}})$.

2.2 Expectations and Nash equilibrium

We now state our main assumptions, formalize the timing, explain how student beliefs about success probabilities are formed and finish by the Nash equilibrium conditions.

2.2.1 Stochastic assumptions, information, and solution concept We first argue that the following assumptions are adapted to our empirical setting:

ASSUMPTIONS S(SETTING).

(S.i) *Preference shocks, ε , and grade shocks, u , are independent of (X, Z) and between each other, and both are continuously distributed.*

(S.ii) *The solution concept is a Nash equilibrium. Students have common knowledge of the sample-specific preferences, ε_i , characteristics, X_i and Z_i as well as common knowledge of grade equation parameters, β , and preference parameters, ζ .*

(S.iii) *The information of students and econometricians on the distribution of random grade shocks, u , and characteristics, X_i and Z_i is symmetric.*

(S.iv) *The distribution of grades is such that $\forall j \in \mathcal{J}, P^j > 0$ almost everywhere P_Z .*

In Assumption S.i, independence of shocks and (X, Z) is a standard exogeneity assumption while it is key in the following that preference shocks, ε and grade shocks, u , are independent. It is akin to the usual assumption in consumer studies that income and preference shocks are independent. A relaxation of this assumption would require an instrumental strategy that is beyond the scope of this paper.²

Assumption S.ii is a complete information set-up that we adopt for two reasons. First, school choice at this university is a game which had been repeated every year over a long time span and which had high stakes for students, families, high schools, and preparatory courses alike. The strategizing ability seems more acceptable in our set-up than in the case of primary or high schools (for instance, see He (2017)). The time period over which a student ability is assessed is much longer and many other agents like parents or teachers are ready to help out students to form expectations (see Manski (1993), for a critical appraisal of such assumptions). Second, a Bayesian–Nash solution concept would be appropriate when agents have private information about their preference shocks, ε_i . Yet, as this congestion game involves many players, it can be conjectured that strong laws of large numbers ensure that the two set-ups are close in terms of aggregate outcomes.

Assumption S.iii might be more controversial since it posits that students have no better knowledge of their own success probabilities than their fellow students or econometricians. First, school choices are shown below to ultimately depend on the ratio of success probabilities in the different schools. Any superior knowledge of an individual specific effect affecting success is partly wiped out by this nonlinear differencing. Second, we use an observable pre-exam national grade in the empirical application to control for superior knowledge. We will briefly return to the effect that the existence of superior information could have on our procedure at the end of this section.

Finally, Assumption S.iv makes sure that a pure strategy is optimal almost surely for all students and simplifies the analysis of the game.

2.2.2 Timing The timing of information revelation, described in the previous section, is formalized as follows. Before schools are chosen, the number of seats in each school, $\{n^j\}_{j \in \mathcal{J}}$ are announced and the total number of participants, say $n + 1$, is observed. We assume that $n + 1 \gg \sum_{j \in \mathcal{J}} n^j$ since the *Vestibular* exam is highly selective.

We distinguish one arbitrary applicant, indexed by 0, from all other applicants, $i = 1, \dots, n$, and we analyze her decision making. We can proceed this way because we are considering an independently and identically distributed (i.i.d.) setting and because the model is assumed symmetric between agents (although they differ ex ante in their observed characteristics and ex post in their unobserved shocks). Applicant 0

²We test and do not reject an implication of this assumption in the empirical section, conditional on admittedly specific auxiliary conditions.

faces the n other applicants and we shall construct her best response to other players' choices, $\{D_i\}_{i=1,\dots,n} \equiv D_{(n)}$ since we use a Nash solution concept (Assumption S.ii). The information set of student 0 at the initial stage comprises at least all elements of $W_0 = (X_0, Z_0, \varepsilon_0)$, $X_{(n)} = \{X_i\}_{i=1,\dots,n}$, $Z_{(n)} = \{Z_i\}_{i=1,\dots,n}$, and $D_{(n)}$.

Student 0 chooses her school ($D_0 \in \mathcal{J}$) as a function of her success probabilities, $\{P_0^j\}_{j \in \mathcal{J}}$, and her preferences as shown in equation (1). Because of Assumption S.iv, student 0 plays a pure strategy almost surely. This is her best response to the aggregate behavior of other students on which success probabilities depend. In this sense, this is an aggregative game (Jensen (2010)) and we will later make use of this characteristic.

After choosing one school, the exam is taken and students are selected in or out of each school, j , by retaining the best n^j students and this defines the thresholds as functions of observed grades. There are two types of risks that student 0 faces. First, the aggregate risks due to grade shocks affecting other students, $U_{(n)}$ whose elements are u_i , $i = 1, \dots, n$, second the individual risks due to her own grade shock, u_0 . Integrating out both risks allows success probabilities to be derived as the rational expectations of success of student 0.

2.2.3 Success probabilities and best responses Denote $Z_{(n)}^j$ the set of grade shifters of the sub-sample of students $i = 1, \dots, n$ applying to school $j \in \mathcal{J}$ that student 0 considers when she computes her best response to $D_{(n)}$. By construction, $Z_{(n)} = (Z_{(n)}^j)_{j \in \mathcal{J}}$. Similarly, we denote $U_{(n)}^j$ the corresponding components of $U_{(n)}$. We shall see in the next subsection how subsamples are derived from primitives. Denote $T = (T^j)_{j \in \mathcal{J}}$ the random vector of exam thresholds that determine entry into each school, $j \in \mathcal{J}$ and whose realizations are observed thresholds $(t^j)_{j \in \mathcal{J}}$. These thresholds are random unknowns at the initial stage since they depend on variables, u , that are random unknowns at the initial stage.³

Should school j be chosen by student 0, her success would be determined, considering the sample of other students, by the binary condition

$$\mathbf{1}\{m(Z_0, u_0, \beta) \geq T^j(Z_{(n)}^j, U_{(n)}^j)\}.$$

Given that the Nash solution concept, Assumption S.ii, fixes the sample of applicants to school j , threshold $T^j(\cdot)$ for school $j \in \mathcal{J}$ only depends on the characteristics of applicants to this school, $Z_{(n)}^j$, and on their grade shocks, $U_{(n)}^j$. Because grades are continuously distributed (Assumption S.i), we can also neglect ties. The existence of thresholds resembles what Azevedo and Leshno (2016) derived in a different context of a stable equilibrium with an infinite number of applicants.

The formal construction of these thresholds is explained below after having determined choices but the intuition is clear. School j threshold that student 0 considers is equal to the grade obtained by the n^j -ranked student in $i = 1, \dots, n$. These thresholds are not explicitly indexed by 0 although they refer to the thought experiment that student 0 performs when constructing her expectations as a function of characteristics and strategies of other students $i = 1, \dots, n$.

³We adopt the term random unknowns to signal that the distribution function of those unknowns are common knowledge. Measurability issues are dealt with below.

When student 0 decides upon a school to apply to, she formulates expected probabilities of success by integrating the condition of success with respect to the aggregate source of risk described by $U_{(n)}^j$ (remember that student 0 observes $Z_{(n)}$ and conditions on $D_{(n)}$) and with respect to the individual source of risk, u_0 .⁴

$$\begin{aligned}
 P_0^j &= P^j(Z_0, Z_{(n)}^j, \beta) = E_{U_{(n)}^j, u_0}[\mathbf{1}\{m(Z_0, u_0, \beta) \geq T^j \mid Z_0, Z_{(n)}^j\}] \\
 &= E_{U_{(n)}^j}[p^j(Z_0, T^j, \beta) \mid Z_0, Z_{(n)}^j],
 \end{aligned}
 \tag{2}$$

in which the following function results from integrating out the individual shock, u_0 , only

$$p^j(Z_0, T^j, \beta) = E_{u_0}[\mathbf{1}\{m(Z_0, u_0, \beta) \geq T^j \mid Z_0, T^j\}].
 \tag{3}$$

Note that the only influence of $U_{(n)}$ is through thresholds which are sufficient statistics. They do not depend on the determinants of student preferences, X_0 and X , except through revealed school choices and they depend on $Z_{(n)}^j$ only through T^j that are computed below. We use the exclusion of X s below for identification.

Denote $D_0(X_0, \varepsilon_0, \zeta, \{P_0^j\}_{j \in \mathcal{J}}) \in \mathcal{J}$ the best response of applicant 0 resulting from equation (1). Given that the sample is i.i.d. and that 0 is an arbitrary representative element of the sample, we can by substitution construct the samples of applicants to school j by using

$$Z_{(n)}^j = \{i \in \{1, \dots, n\}; D_i(X_i, \varepsilon_i, \zeta, \{P_i^k\}_{k \in \mathcal{J}}) = j\}.$$

It is thus clear that the application mapping $Z_{(n)}$ into $Z_{(n)}^j$ is measurable although it remains to be shown that the application mapping $Z_{(n)}$ into thresholds $(T^j)_{j \in \mathcal{J}}$ is measurable. That is what we do now.

2.2.4 The determination of the thresholds We can now return to the determination of thresholds $(T^j)_{j \in \mathcal{J}}$, considered by agent 0. For any realization of $U_{(n)}$, the J Nash equilibrium conditions yield a realization of the thresholds, $\{t^j\}_{j \in \mathcal{J}}$, as:

$$\sum_{i=1}^n [\mathbf{1}\{D_i = j\} \mathbf{1}\{m(Z_i, u_i, \beta) \geq t^j\}] = n^j.
 \tag{4}$$

As usual with empirical quantiles, this system has many solutions, t^j . We retain the solution corresponding to the grades of the less well-ranked applicant in each school and because ties are absent with probability one, this solution is unique, and a measurable function of $Z_{(n)}$ and $U_{(n)}$. This defines the random thresholds, $\{T^j\}_{j \in \mathcal{J}}$.

Equations (1) and (4) are necessary conditions for a Nash equilibrium.⁵ A sketch of proof of the existence of a Nash equilibrium is spelled out in Appendix A and builds upon

⁴All expectations exist since integrands are measurable and bounded.

⁵Because the number of applicants is very large with respect to the total capacity, we neglect the occurrence that seats remain unmatched. In other words, we assume that the probability that one subsample j contains less than n^j students is zero or negligible. At the university under consideration, the average rate of success is between 5% and 20% (Table S.i, Supplementary Appendix).

tools developed for potential games with weak strategic substitutes (Dubey, Haimanko, and Zapechelnyuk (2006)).

2.3 Identification of success probabilities and preferences

We now study the identification of success probabilities and preferences.

2.3.1 Success probabilities Success probabilities are expressed, using equation (2), as a function of known variables—characteristics Z_0 , $Z_{(n)}$, and decisions $D_{(n)}$ —and unknown variables—parameter β , the distribution of grade shocks, u , and the distribution of thresholds $\{T^j\}_{j \in \mathcal{J}}$. First, the grade equation, $m = m(Z, u; \beta)$, identifies parameter β and the distribution of u . Plugging these objects into the Nash conditions (4), given $Z_{(n)}$ and $D_{(n)}$ and computing thresholds identifies the distribution of $\{T^j\}_{j \in \mathcal{J}}$. In consequence, success probabilities are identified. In the following, we denote them, $\{P^j(Z)\}_{j \in \mathcal{J}}$.

In general, the identification of $P^j(Z)$ depends on the context which may be less simple than the one we used here. Yet, it is likely in general that exogenous variation in these probabilities could be given by various measures of ability, not only of an aggregate type as here, but also by field-specific grades. In Section 3, we return to the identification of success probabilities in our empirical application.

2.3.2 Choice-based sample and outside options We adopt a general random utility setup in which values are continuously distributed (Assumption S.i). By the probability integral transform, we can thus always adopt the representation in which each function V^j is monotonic in one unobservable, denoted ε^j , whose marginal distribution is uniform on $[0, 1]$:

$$\forall j \in \mathcal{J}, \quad V^j = V^j(X, \varepsilon^j) < \infty.$$

Dependence between ε_j s is left unrestricted and is described by any continuous copula.

There are two issues of concern for identification that distinguishes this proof from Agarwal and Somaini (2018). The first one regards choice-based sampling since our sample comprises students interested by at least one school, so that we condition the analysis on the event that $\max_{j \in \mathcal{J}} (V^j) > 0$. Second, we have to consider that only some schools could have positive value for students and we have to condition the analysis on unobservable latent sets $\mathcal{J}_+ \subset \mathcal{J}$ of schools that provide positive utility. Namely, other schools, in the complement of \mathcal{J}_+ in \mathcal{J} , $\mathcal{J}_+^c = \mathcal{J} / \mathcal{J}_+$, are strongly dominated by the outside option with probability one.

The finite set \mathcal{J}_+ whose number of elements is greater or equal to one because of choice based sampling is a random set whose distribution is induced by the distribution of random values, V^j ,

$$Q(\mathcal{J}_+ | X) = \Pr\left(\forall j \in \mathcal{J}_+, V_j > 0; \forall j \in \mathcal{J}_+^c, V_j \leq 0 \mid \max_{j \in \mathcal{J}} (V^j) > 0, X\right).$$

Let us first derive the optimal school choice conditional on \mathcal{J}_+ and integrate out \mathcal{J}_+ in a second step. If set \mathcal{J}_+ is a singleton, student's choice is its single element and success

probabilities do not matter. If set \mathcal{J}_+ has two or more elements, success probabilities affect choices through the relative values of $P^j V^j$ (equation (1)). Students may disguise their true preferences and act strategically. These relative values are positive because set \mathcal{J}_+ is defined as such and because $P^j > 0$ by Assumption S.iv. We can rewrite the decision model when $j \in \mathcal{J}_+$ by taking the logarithm of equation (1):

$$D = j \text{ if } \log(P^j) + \log(V^j) > \max_{k \in \mathcal{J}_+ / \{j\}} (\log(P^k) + \log(V^k)), \tag{5}$$

in which we kept the dependence of V^k on X , ε^k and of P^j on Z implicit and in which ties are of probability zero because of Assumption S.i. Denote $\Delta^{jk}(Z) = \log(P^j(Z)) - \log(P^k(Z))$ in the following and express choice probabilities, by integrating out sets \mathcal{J}_+ , as

$$\begin{aligned} & \Pr(D = j \mid Z, X) \\ &= \sum_{\mathcal{J}_+ : \mathcal{J}_+ \supset \{j\}} Q(\mathcal{J}_+ \mid X) \Pr(\forall k \in \mathcal{J}_+, \Delta^{jk}(Z) > \log(V^k) - \log(V^j) \mid X, Z, \mathcal{J}_+, \mathcal{J}_+^c). \end{aligned} \tag{6}$$

It is useful to consider the two-school example to understand the sequence of proofs below.

The two-school example When the choice set is reduced to two elements, $\mathcal{J} = \{S, F\}$ as in our empirical application,⁶ Figure 1 exhibits how we solve the decision problem in each of four quadrants.

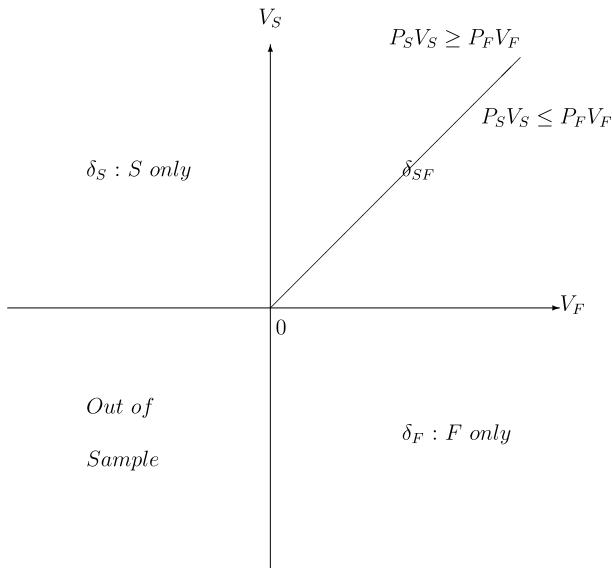


FIGURE 1. Choice space.

⁶Our empirical application deals with two schools in two cities, Fortaleza and Sobral, and we use their initials, F and S , to make easier the recollection of which school we are talking about.

First, the southwest quadrant is composed of individuals who are excluded from the choice-based sample and its probability measure is not identified. Second, in the northwest quadrant, $V^S > 0$ and $V^F \leq 0$, $\mathcal{J}_+ = \{S\}$ and school S is necessarily chosen. The probability measure of this quadrant is

$$\delta^S(X) = Q(\{S\} | X) = \Pr\{V^S > 0, V^F \leq 0 \mid \max(V^S, V^F) > 0, X\}.$$

Similarly, in the southeast quadrant, $V^F > 0$ and $V^S \leq 0$, $\mathcal{J}_+ = \{F\}$ and school F is necessarily chosen. Its probability measure is $\delta^F(X) = Q(\{F\} | X)$. In both regions, students reveal their true preferences and do not act strategically. Note that identification is ordinal only in these two quadrants.

This is different in the northeast quadrant since choices can change if success probabilities P^S and P^F change. Specifically, school S is chosen if and only if

$$\log(P^S) + \log(V^S) > \log(P^F) + \log(V^F).$$

Denoting $\delta^{SF}(X)$ as the probability of the northeast quadrant, the choice probability regarding the first school is derived from equation (6):

$$\begin{aligned} \Pr(D = S | X, Z) \\ = \delta^S(X) + \delta^{SF}(X) \Pr\{\log(P^S) - \log(P^F) > \log(V^F) - \log(V^S) \mid X, Z\}. \end{aligned} \quad (7)$$

Returning to the general equation (6), we now study the identification of the two following structural objects; first, the probability measure of each quadrant, $Q(\mathcal{J}_+ | X)$; second, the joint distribution of log-value differences, $\log(V^j) - \log(V^k)$, in each quadrant \mathcal{J}_+ .

2.3.3 Identification of preferences As is well known in discrete models since [Manski \(1988\)](#) and [Matzkin \(1993\)](#), a necessary condition for identification is the full variation of some regressors, conditionally on others. Those regressors are here the success probabilities:

ASSUMPTION CV (Complete Variation). *Almost everywhere (a.e.) P_X , the support of $P(Z) = (P^j(Z))_{j \in \mathcal{J}}$, conditional on X , is the set $(0, 1)^J$.*

Assumption CV requires first that the set of covariates Z is at least of dimension J and that their variation induces that the support of success probabilities is the full unit hypercube. Success probabilities $P(Z)$ act as prices ([Azevedo and Leshno \(2016\)](#)) and the effects of preference shifters cannot be identified from success probabilities absent exclusion restrictions. This is why this assumption requires that a sufficient number of grade shifters, Z , should be excluded from the list of preference shifters, X . This is akin to the exclusion of school priorities from preferences in [Agarwal and Somaini \(2018\)](#).

We use equation (6) and make success probabilities vary in the unit hypercube. We adopt a two-step strategy. First, we show that the probability measures of quadrants, $Q(\mathcal{J}_+ | X)$, are identified.

PROPOSITION 1. *Under Assumption CV, for any nonempty $\mathcal{J}_+ \subset \mathcal{J}$, $Q(\mathcal{J}_+ | X)$ is identified.*

PROOF. See Appendix B.1 □

The intuition for this proof is better gained by using again the two-school example. The structural probabilities of each quadrant in Figure 1 are

$$\{\delta^S(X), \delta^{SF}(X), \delta^F(X)\},$$

and these appear in equation (7). By Assumption CV, the support of $\Delta^{SF}(Z) = \log(P^S) - \log(P^F)$ is the full real line and we can identify δ^S by using the limit of equation (7):

$$\delta^S(X) = \lim_{\Delta^{SF}(Z) \rightarrow -\infty} \Pr(D = S | \Delta^{SF}(Z), X).$$

Interchanging S and F identifies δ^F and $\delta^{SF}(X) = 1 - (\delta^S(X) + \delta^F(X))$.

Returning to the general case, we now prove identification of the distribution function of log-value differences, $\log(V^j) - \log(V^k)$, in each quadrant \mathcal{J}_+ .

PROPOSITION 2. *Under Assumption CV and $\forall \mathcal{J}_+ \subset \mathcal{J}$, $Q(\mathcal{J}_+ | X) > 0$ a.e. P_X , the joint distribution of $\Pr((\log(V^k) - \log(V^j))_{k \in \mathcal{J}_+ / \{j\}}, | X, \mathcal{J}_+, \mathcal{J}_+^c)$ is identified a.e. P_X , for any $\mathcal{J}_+ \subset \mathcal{J}$, and fixing any specific $j \in \mathcal{J}_+$ as the “reference” alternative.*

PROOF. See Appendix B.2. □

We added the condition that $Q(\mathcal{J}_+ | X) > 0$ for simplicity. In the case, $Q(\mathcal{J}_+ | X) = 0$, preferences cannot be identified in set \mathcal{J}_+ but it has no importance.

The proof of the proposition is by induction over the total number of schools and we thus deal with the two-school example again to provide the intuition.

A TWO-SCHOOL EXAMPLE (CT’D). From equation (7) and assuming $Q(\{S, F\} | X) = \delta^{SF}(X) > 0$, we can form the expression that

$$\frac{\Pr(D = S | \Delta^{SF}(Z), X) - \delta^S(X)}{\delta^{SF}(X)} = \Pr(\Delta^{SF}(Z) > \log V^F - \log V^S | X, Z). \tag{8}$$

All terms on the left-hand side are identified and standard arguments (Matzkin (1993)) show that the distribution of $\log V^S - \log V^F$ conditional on X is identified under the condition that the support of $\Delta^{SF}(Z)$, conditional on X , is the full real line.

Returning to the main argument, it is to be emphasized that Proposition 2 states identification of a joint distribution of preferences within a quadrant. This implies that Propositions 1 and 2 have the corollary that counterfactuals, investigating alternative mechanisms, are identified and this is what we use in the empirical application. Expected utilities of a rank-ordered list of any length are derived from the success probabilities and the joint distribution of differences of values in each set \mathcal{J}_+ of alternatives

with positive values. Generally speaking, what matters is that expected utilities are bilinear functions of the underlying values, V^k , and of the success probabilities (equation (7), Agarwal and Somaini (2018)). The corollary thus applies to all mechanisms described by Definitions 1–3 of Agarwal and Somaini (2018).

We finish by a set of remarks about extensions.

REMARK 1. Most importantly, this identification proof is obtained under the restrictive condition that one school only is chosen by students. When a more informative rank-ordered list comprising several schools can be submitted by students, identified objects in Propositions 1 and 2 remain the same.⁷ We proceed by providing a counterexample of a result that would state that other objects can be identified in the two-school case.

Recall that in the two-school case, observing rank-ordered lists of length one identifies the probability of school, say S , to be positively valued and of school, say F , of being negatively valued, as a limit result by varying success probabilities in such a way that school, F , always dominates S if both are positively valued (Proposition 1). If we can now observe rank-ordered lists of length 2, the same probability can be identified by the probability of observing a rank-ordered list which ranks S first and the empty set second.

Using length-2 rank-ordered lists, we cannot identify, however, more than this probability in the quadrant in which $V_S > 0$ and $V_F \leq 0$ and this is true as well in the other quadrant $V_S \leq 0$ and $V_F > 0$. In the quadrant $V_S > 0$ and $V_F > 0$ in which the log differences of values are identified (Proposition 2), this holds true as well. Admittedly, success probabilities change if we change the length of the rank-ordered lists but their identification still relies on using the continuous variation of grades and the cutoffs are still determined by equations similar to equation (4). In conclusion, observing longer rank-ordered lists leads to overidentification that could help increasing the precision of preference estimates although this issue is out of the scope of this paper.

REMARK 2. Differences of log-values are nonparametrically identified but levels are not identified. In Section 4, the evaluation of counterfactual welfare is achieved by completing identifying conditions with additional assumptions.

REMARK 3. We could further adopt a linear median restriction for differences between logarithms of values such as, in the two-school example,

$$\log V^S - \log V^F = X\gamma + \varepsilon$$

in which the distribution of ε , $F(\cdot | X)$ is restricted as

$$F(0 | X) = \frac{1}{2}. \quad (9)$$

Parameter γ and $F(\varepsilon | X)$ are identified.

⁷This result requires that students never rank negatively-valued schools in their rank-ordered lists because, for instance, they face an infinitesimal cost of refusing an offer (that they could receive from a negatively-valued school if they rank it).

REMARK 4. It is possible to weaken Assumption CV and admit that the support of the conditional distribution of $\Delta^{jk}(Z)$ conditional on X might not be the full real line. If we keep the two-school example to make the point in a simple setting, assume for convenience that the support of Δ^{SF} for any value taken by X includes the value 0. Then as developed in Manski (1988), identification becomes partial under the median restriction (9) written above. Parameter γ is identified using the median restriction and $F(\cdot | X)$ is identified in the restricted support in which $\Delta(Z) + X\gamma$ varies. Our data exhibit limited variation and this is why we adopt, in the empirical application, a parametric assumption for $F(\cdot | X)$. What nonparametric identification arguments above prove is that this parametric assumption is a testable assumption at least in the support in which $\Delta^{SF}(Z) + X\gamma$ varies.

REMARK 5. We can now briefly return to the issue of superior information that students could have with respect to econometricians. To discuss this point, suppose that each student receives a signal, before choosing the school to apply to, about her ability, say σ_i , and which is correlated with exam grades. If we keep the complete information structure, signals are fully observed by agents. Using the same model of belief about success in each school but conditioning now on the vector of signals σ , agents use success probabilities that can be written as $\pi(Z, \sigma)$ instead of $P(Z)$. Because of the law of iterated expectations, we have that $E(\pi(Z, \sigma)|Z) = P(Z)$. Denote $W = \pi(Z, \sigma)/P(Z)$ the positive random variable standing for superior information and which is mean independent of Z by construction.⁸ This is not, however, a sufficient condition to recover log-value differences and it shall be additionally assumed that W and X, Z are independent to prove that log value differences are identified up to an additive independent “measurement” error term. A common prior assumption for agents and econometricians alike is thus a strong assumption but absent any other observed decision variable that might help recover or proxy σ (see Campbell (1987) for instance), dealing with the general case seems out of reach.

3. THE EMPIRICAL APPLICATION

We begin with describing our empirical application and with adapting the general model described in the previous section to the particulars that Universidade Federal do Ceará (UFC from here on out) in Northeastern Brazil used to select students in 2004. We then turn to the computation of success probabilities and give a summary of our empirical strategy. We finish by reviewing our estimation results.

We restrict, for various reasons, the empirical application to two medical schools only. They are respectively located in Sobral (denoted S), the second most populated city in the state of Ceará and Fortaleza (denoted F), the state capital. First, the content of second-stage exams differs if schools are in different fields (for instance, medicine or law) and this would introduce substantial heterogeneity between schools. Second, it

⁸We take the ratio between those probabilities because the decision model in set \mathcal{J}_+ is written in logarithms.

enables us to choose the best schools in the university for which our assumptions on information and outside options are the most likely to be satisfied. Third, the more schools are analyzed, the stronger the requirement of complete variation of success probabilities (Assumption CV) for identification is.

We chose the two best medical schools because (1) they are the schools which attract the best students among all candidates within UFC (see Table S.ii in the Supplementary Appendix); (2) on prior grounds, the best substitutes are a slightly lower quality medical school in the countryside (Barbalha) or pharmacy and related fields with a lower standing in terms of cut-off grades and (3) other schools of excellence are schools of law which are presumably bad substitutes. As a matter of fact, the best substitutes are outside the university: a state university and three private medical colleges in Fortaleza and Sobral; outside the state, medical schools in Recife at the closest or in Sao Paulo or Campinas further away. All substitutes are dealt within the model as an aggregate outside option.⁹

We focus on medical schools also because of their attractiveness for the best students. Almost no students desist between the two-stage exams if they pass the first stage. Being accepted in those schools is extremely valuable and the care and attention of students, parents, and teachers are certainly at their highest for those two schools. The school in Sobral is small and offers 40 positions only while Fortaleza is much larger since it offers 150 seats. As shown in the empirical analysis below, this asymmetry turns out to be key for evincing strategic effects.

3.1 *Timing and the two-stage exams of the Vestibular*

The timing of the real mechanism is enriched in two ways with respect to the stylized setting that we described in Section 2.

First, students take a standardized national exam, known as *ENEM* and measuring students' ability in different subjects (math, etc.) before college choices are made and about one year before *Vestibular* exams begin. *ENEM* results are used by the university when computing the passing thresholds at the *Vestibular* exams. It is also a very convenient measure of ability that all students know when they choose their preferred school.

Second, exams are taken in two stages. The first-stage exam is identical across schools and denoted as

$$m_1 = m_1(Z, u_1; \beta_1)$$

in which u_1 are random individual circumstances that affect results at this exam. After this first exam, students are ranked according to a weighted combination of grades *ENEM* and m_1 . Those weights are common knowledge *ex-ante* and measure the relative interest of schools in selecting students using the national and the local exam grades. The thresholds of success at the first-stage exam are given by the rule that the number of available slots is equal to four times the number of final seats offered by the school. Given that schools have respectively 40 and 150 seats, the number of students passing the first stage is 160 and 600 out of a total of 542 and 2325 candidates.

⁹See Sections S.1 and S.2 in the Supplementary Appendix which justify these arguments and complement the empirical analysis presented here.

We write the selection rule after the first exam as

$$m_1 \geq t_1^j(ENEM) = \tau_1^j - a_1 ENEM,$$

in which τ_1^j is determined by the number of candidates and positions available in the school. Threshold t_1^j depends on $ENEM$ because students are ranked according to a weighted sum of m_1 and $ENEM$ whose weights are $(1, a_1)$ but we make this dependence implicit in the following.

Students who do not pass the first exam get their outside option $D = \emptyset$, with utility, V_\emptyset . Other students take the second-stage exam and get a second-stage grade, denoted m_2 :

$$m_2 = m_2(Z, u_2; \beta_2),$$

where u_2 is an error term whose interpretation is similar to u_1 and u_2 and is possibly correlated with u_1 . These students are ranked according to a known weighted linear aggregator of $ENEM$, m_1 and m_2 , and this again stands for the relative importance given to each of these dimensions by schools. Students are accepted in the order of their ranks until completion of the positions available for each school. As before, we write the selection rule as

$$m_2 \geq t_2^j(ENEM, m_1),$$

as a function of a second threshold which also depends on previous exam grades since a linear aggregator is used to rank students. Students who fail the second-stage exam get the same outside utility as students who fail the first-stage exam.

We can then extend the definition of the probability of success in school D to

$$P^D = \Pr(m_1(Z, u_1; \beta_1) \geq T_1^D(ENEM), m_2(Z, u_2; \beta_2) \geq T_2^D(ENEM, m_1)).$$

3.2 Identification of grade equations and success probabilities

Only students who pass the first-stage exam can write the second-stage exam. Therefore, in our data, the second-stage grades, m_2 , are censored when first-stage grades, m_1 , are not large enough that is, $m_1 < T_1^j$ and in the absence of any restriction, the distribution of m_2 is not identified.

3.2.1 A control function approach To proceed, we shall specify that $(m_1(Z, u_1; \beta_1), m_2(Z, u_2; \beta_2))$ are linear indices of covariates with respective parameters β_1 and β_2 . The estimation of β_1 proceeds under the restriction that $E(u_1|Z) = 0$. In the second-stage grade equation, we use a control function approach to describe the influence of the unobservable factor derived from the first grade equation (Blundell and Powell (2003)). We assume that

$$u_2 = g(u_1) + u_2^*$$

in which u_2^* is mean independent of u_1 , $E(u_2^* | u_1, Z) = 0$.

By doing this, we are now also able to control the selection bias since u_2^* is supposed to be mean independent of u_1 and, therefore, $E(u_2^* | m_1 \geq T_1^j, Z) = 0$. This would identify

parameters and the control function $g(\cdot)$. Nonetheless, our goal is not only to estimate these parameters but also to estimate the joint distribution of (u_1, u_2) . This is why in the following we assume that u_1 and u_2^* are independent of each other and of variables Z and simply use the estimated empirical distributions of u_1 and u_2 when estimating success probabilities.

3.2.2 Simulated success probabilities To predict success probabilities, two important elements are needed: the joint distribution of random terms u_1 and u_2 and the admission thresholds for the first- and second-stage grades. We already stated assumptions under which we can recover the former. The latter are derived from the definition of the final admission in each school as described by two inequalities as functions of linear combinations of initial grades and first- and second-stage grades fixed by the university:

$$\begin{aligned} m_1 + 120 * ENEM/63 &\geq \tau_1^j, \\ 0.4 * (m_1 + 120 * ENEM/63) + 0.6 * m_2 &\geq \tau_2^j. \end{aligned} \quad (10)$$

Thresholds (τ_1^j, τ_2^j) are taken here as any possible realization and we construct equation (3) from the distribution of random grade shocks. Integrating out thresholds T_1^j and T_2^j comes in a second step.

Conditional success probabilities We first transcribe the inequalities (10) as functions of unobserved heterogeneity terms u_1 and u_2 . For every student, passing the two exams means that the two random terms in the grade equations should be large enough as described by

$$\begin{aligned} u_1 &\geq \tau_1^j - 120 * ENEM/63 - Z\beta_1, \\ u_2^* &\geq \frac{\tau_2^j}{0.6} - \frac{2}{3}(Z\beta_1 + u_1 + 120 * ENEM/63) - Z\beta_2 - g(u_1). \end{aligned}$$

Notice that the second inequality depends on first-stage grade shocks, u_1 , because of the correlation between grades. Therefore, the success probability in a school j , as defined by a function of thresholds in equation (3), can be expressed as

$$\begin{aligned} p^j(Z, \beta, \tau_1^j, \tau_2^j) &= \Pr\left\{u_1 \geq m_1^j - Z\beta_1, u_2^* \geq m_2^j - \frac{2}{3}Z\beta_1 - Z\beta_2 - \frac{2}{3}u_1 - g(u_1)\right\}, \\ &= \int_{m_1^j - Z\beta_1}^{\infty} f_{u_1}(x) \left(\Pr\left\{u_2^* \geq m_2^j - \frac{2}{3}Z\beta_1 - Z\beta_2 - \frac{2}{3}x - g(x)\right\} \right) dx, \\ &= \int_{m_1^j - Z\beta_1}^{\infty} f_{u_1}(x) \left[1 - F_{u_2^*}\left(m_2^j - \frac{2}{3}Z\beta_1 - Z\beta_2 - \frac{2}{3}x - g(x)\right) \right] dx, \end{aligned} \quad (11)$$

in which m_1^j and m_2^j are functions of thresholds:

$$\begin{cases} m_1^j = \tau_1^j - 120 * ENEM/63, \\ m_2^j = \frac{\tau_2^j}{0.6} - \frac{2}{3}(120 * ENEM/63). \end{cases}$$

Unconditional success probabilities As those are derived from an expectation taken over thresholds T in equation (2), we use a simulated sample analog and compute the distribution function of T at an arbitrary level of precision using equilibrium conditions (4)¹⁰ by simulation of $U_{(n)}$. By construction, T depends on observation 0, and thus its distribution has to be computed for every single observation. For simplicity and because this dependence matters less and less when n grows, we compute those thresholds in the empirical application using equation (4) in which the sums are taken over the full sample $i = 0, 1, \dots, n$ and success probabilities are estimated only once instead of $n + 1$ leave-one-out estimates.

3.3 Empirical strategy: Summary

We first estimate parameters of the grade equations and denote them $\hat{\beta}_n$. This, in turn, allows us to compute the expectation of the success probabilities conditional on thresholds t_k^j , $k = 1, 2$, $j = S, F$ as in equation (11) using the estimated distribution functions for errors in the grade equations. We then compute unconditional success probabilities by integrating out by simulation conditional success probabilities as in equation (2). Namely, for any simulation $c = 1, \dots, C$, draw in the distribution of $U^{(n)}$ and derive realizations of T , say t_c in the C samples of size n by fixing choices $\mathbf{1}\{D_i(Z_i, \varepsilon_i, \zeta, P_i^S, P_i^F) = S\}$, characteristics X_i , and by solving the equilibrium conditions (4). Equation (2) can then be computed by integration as

$$\hat{P}_{0,C}^j = \frac{1}{C} \sum_{c=1}^C P^j(Z_0, \hat{\beta}_n, t_{1,c}^j, t_{2,c}^j). \tag{12}$$

Preferences are described by the probabilities of each quadrant in Figure 1, $\{\delta^S(X), \delta^{SF}(X), \delta^F(X)\}$ and by the following parametric specification of log-value differences:

$$\log V^S - \log V^F = X\gamma + \varepsilon, \quad \varepsilon \sim N(0, 1).$$

Preference parameters $\zeta = (\delta, \gamma)$ are estimated using a conditional maximum likelihood approach:

$$\hat{\zeta}_n = \arg \max_{\zeta} l(\zeta | \hat{P}_{0,C}^S, \hat{P}_{0,C}^F).$$

This is a conditional likelihood function since $\hat{P}_{0,C}^S, \hat{P}_{0,C}^F$ depend on the first-step estimate, $\hat{\beta}_n$. Standard asymptotic arguments yield

$$\hat{\zeta}_n \xrightarrow[n \rightarrow \infty]{P} \zeta.$$

We used bootstrap to obtain the covariance matrix of those estimates by replicating the complete estimation procedure as a mixture of nonparametric (grade equations) and parametric bootstrap (choice equations).

¹⁰Generalizing them to the two-stage exam setting is straightforward; see equation (13) below.

TABLE 1. Descriptive statistics in the two medical majors.

Variable	Sobral: 40 positions					
	Mean	Median	Std. Dev.	Min.	Max.	<i>N</i>
Grade: National exam (m_0)	50.43	52.00	7.29	18.00	61.00	527
Grade: First stage	71.67	73.00	15.74	20.00	103.00	527
Grade: Second stage	240.0	246.5	33.98	94.3	296.6	160
Female	0.47	0	0.50	0	1	527
Age	19.58	21.50	2.48	16.00	25.00	527
Private high school	0.87	1	0.33	0	1	527
Repetitions	0.99	1	0.88	0	2	527
Preparatory course	0.71	1	0.45	0	1	527
Father's education	2.09	2	1.03	0	3	527
Mother's education	2.21	3	0.98	0	3	527

Variable	Fortaleza: 150 positions					
	Mean	Median	Std. Dev.	Min.	Max.	<i>N</i>
Grade: National exam (m_0)	49.16	52.00	10.03	12.00	63.00	2340
Grade: First stage	70.06	72.00	20.01	20.01	110.00	2340
Grade: Second stage	240.0	245.1	34.37	48.3	311.1	600
Female	0.54	1	0.50	0	1	2340
Age	19.13	17.50	2.43	16.00	25.00	2340
Private high school	0.77	1	0.41	0	1	2340
Repetitions	0.69	1	0.83	0	2	2340
Preparatory course	0.59	1	0.49	0	1	2340
Father's education	2.13	2	1.00	0	3	2340
Mother's education	2.15	2	0.98	0	3	2340

Note: Source: Vestibular cross section data in 2004.

3.4 A brief description of estimation results

The list of variables and descriptive statistics in the pool of applicants to the two schools we consider appear in Table 1. Looking at admission rates, one can see that Sobral admitted $40/527 = 7.6\%$ and Fortaleza $150/2340 = 6.4\%$ and this makes Fortaleza more competitive. Comparing the mean and median of initial and first-stage grades, Sobral has nonetheless better applications than Fortaleza. As to the second-stage grades, the group selected for Sobral has a slightly higher median than the one selected for Fortaleza although both groups have the same mean.

Because empirical results in this article are focused on counterfactuals, estimates from our empirical analysis are shown and analyzed in Section S.2 in the Supplementary Appendix. We fully report and comment therein estimates of grade equations, predictions of success probabilities and estimates of preference parameters. We now discuss only briefly our most important modeling choices and our main results.

As described in Table 1, explanatory variables are those that affect exam performance or school preferences. For grade equations, all potential explanatory variables are included: a proxy for ability which is the initial grade m_0 obtained at the national

exam (ENEM), age, gender, educational history, repetitions, parents' education, and the undertaking of a preparatory course. Our guide for selecting variables is that a better fit of grade equations leads to a better prediction of success probabilities in the further steps of our empirical strategy.

Second, as developed in Section 2.3.3, one exclusion restriction at least is needed to identify preferences. We chose to exclude from preference shifters all variables related to past educational history. Indeed, preferences are related to the forward looking value of the schools (e.g., wages) which, conditional on the proxy for ability, is unlikely to depend on the precise educational history of the student (e.g., private/public sector history and undertaking a preparatory course). This is even more likely since we condition on ability m_0 which is assessed in the ENEM after educational history. This dynamic exclusion restriction is akin to what is assumed in panel data and posits that m_0 is a sufficient statistic for educational history. As a consequence, preferences are specified as a function of ability, gender, age, education levels of father and mother, and the number of repetitions of the entry exam. The inclusion of gender, age, and education of parents is standard in this literature. The number of repetitions reveals either the determination of a student through her strong preference for the schools or the lack of good outside options. We performed a thorough specification search and tested for overidentifying restrictions.

Third, the second-stage exam has a different format (writing essays) than the first-stage multiple choice exam and the second-stage grade equation has a much lower R^2 . An interesting economic interpretation is that the first-stage exam is designed to skim out the weaker students and this multiple question exam is quite predictable (large R^2). In the second stage, the examiners can be selective in many more dimensions and try to pick out students using unobserved traits which are predictive of future behavior (success in the field of studies, drop outs, etc.) and that the econometrician cannot observe. This justifies the double stage nature of the exam as trying to minimize screening costs. We will return to this point below.

Fourth, Table 2 reports descriptive results on predicted probabilities of success. Means and medians of first-stage success probabilities are around 20–30% in both schools. This is close to what is observed in the sample but not exactly identical since these probabilities are partly counterfactual objects, for instance, success probabilities in Sobral for those who chose Fortaleza. The second-stage success probabilities are close to what is observed and as expected roughly four times lower than the first-stage ones.

Finally, students heavily favor Fortaleza over Sobral and this confirms that Fortaleza is the most popular medical school in the state. The ratio of those probabilities is 10 which is approximately the ratio between the populations of the two cities albeit much larger than the ratio of final seats in the two schools (150/40). Nonetheless, there is a substantial fraction of students whose utilities for both schools are positive (more than 40%).¹¹

¹¹Full details and comments of our empirical analysis appear in Section S.2 of the Supplementary Appendix.

TABLE 2. Simulated success probabilities.

	Sobral		Fortaleza	
	Stage 1	Final Success	Stage 1	Final Success
Min.	0.000	0.000	0.000	0.000
25%	0.001	0.001	0.000	0.000
Median	0.088	0.011	0.012	0.004
Mean	0.314	0.076	0.203	0.062
75%	0.676	0.103	0.360	0.071
Max.	1.000	0.934	1.000	0.920

Note: Success probabilities are constructed using 1000 Monte Carlo simulations.

4. EVALUATION OF THE IMPACT OF CHANGES OF MECHANISMS

We now investigate the impact of various changes in the allocation and selection mechanisms that are discussed in academic and policy debates. To organize the presentation, some preliminary discussion of school preferences over the information obtained at the different exams is in order. We also return to the issue of substitutes.

School preferences are revealed by the type of exams and selection rules that are used for admission as described by equations (10). In the following, we will evaluate outcomes and welfare in each counterfactual by conditioning on the expected final scores used for admission.¹² The first-stage selection sums two multiple-choice exam scores—ENEM and first stage—in a roughly equivalent way.¹³ Final selection however overweighs the second-stage grade (0.6) with respect to the compound ENEM and first-stage grade (0.4). As the weight of the latter is not zero, first-stage and ENEM scores provide valuable information in addition to the selection role they are used for. Scores at the two stage exams presumably measure two different cognitive dimensions affecting the future career of students in the schools. Note however that there is no bottom grade requirement at the second stage as there is at the first stage and the second stage cannot be considered as more informative even if its weight is larger.

Second, counterfactual analyses are conducted under the assumption that the choice-based sample remains the same and this implicitly means that the value of outside options does not change. We also assume that the change in predetermined variables, for instance, the take-up of preparatory courses or the exit and entry flows of students applying to these two schools because of the change in the admission rules is second order.

The first counterfactual experiment that we implement is to cut slots proposed at the second-stage exam by offering twice, instead of four times, the number of final seats. It is likely that a two-stage exam is used because schools want to avoid congestion and the tuning between the two stages is key. The first stage is very easy to grade since machines

¹²As expected, success probabilities are an increasing function of this score. More interestingly, the ratio between success probabilities at Sobral relative to Fortaleza is also increasing with this score except at the very top.

¹³In equation (10), the coefficient in front of ENEM is a rescaling term that equalizes ranges of m_1 and ENEM.

can mark multiple-choice exams very quickly. The second stage is much deeper since it relies on open-ended questions and is more costly to grade. The trade-off is therefore to balance these substantial screening costs with the depth of the first-stage selection that might select out good students because of the format (see also [He and Magnac \(2018\)](#)). There are other examples of this in other countries: in top engineering schools in France, selection is distinguished in an admissibility (written) and an admission stage (oral).

Second, we experiment with enlarging the choice set of students before taking exams. They would list two ordered choices instead of a single one so as to get closer to a Gale–Shapley mechanism. This means that even if students fail the first-stage qualification in one of the two schools they may still get into the second-stage exam for the other school. This implies that the average skill level of passing students increases and that the difference between the two schools is attenuated.

Third, since having two stages in the exam allows schools to cut costs and achieve a more in-depth selection at the second stage, another experiment consists in changing the timing of choice. In the third counterfactual experiment, students would choose their final school after taking the first exam and learning their grades. The experiment is different from the previous two since students have more information on their success probabilities when they choose. It generates however additional organization costs and delays due to the serial dictatorship mechanism that it induces after the first stage. It is also likely to generate more opportunistic behavior.

Before entering into the details of these counterfactual mechanisms, the identification of utilities from estimated preferences and success probabilities is key in these evaluations. We show that expected utilities are underidentified and we suggest how plausible bounds for counterfactual estimates can be constructed. We also explain how to compute counterfactual estimates conditional on observed choices.

4.1 Identifying counterfactual expected utilities

Taking expectations with respect to grades using success probabilities P_i^S, P_i^F of ex post utility levels, U_i , leads to

$$\begin{aligned} E(U_i | V_i^S, V_i^F) &= \mathbf{1}\{V_i^S \geq 0, V_i^F < 0\} P_i^S V_i^S + \mathbf{1}\{V_i^F \geq 0, V_i^S < 0\} P_i^F V_i^F \\ &\quad + \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} [\mathbf{1}\{D_i = S\} P_i^S V_i^S + \mathbf{1}\{D_i = F\} P_i^F V_i^F] \\ &= P_i^S V_i^S (\mathbf{1}\{V_i^S \geq 0, V_i^F < 0\} + \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = S\}) \\ &\quad + P_i^F V_i^F (\mathbf{1}\{V_i^F \geq 0, V_i^S < 0\} + \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = F\}). \end{aligned}$$

Even if the location parameter is fixed by the outside option, this expected utility can always be rescaled by any increasing function. This is why we choose the absolute value $|V_i^F|$ as the scale factor to set:

$$\begin{aligned} V_i^F &= 1 \quad \text{if } V_i^F > 0, \\ V_i^F &= -1 \quad \text{if } V_i^F < 0. \end{aligned}$$

Under this normalization,

$$\begin{aligned} E(U_i | V_i^S, V_i^F) &= P_i^S \left(V_i^S \mathbf{1}\{V_i^S \geq 0, V_i^F < 0\} + \frac{V_i^S}{V_i^F} V_i^F \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = S\} \right) \\ &\quad + P_i^F V_i^F (\mathbf{1}\{V_i^F \geq 0, V_i^S < 0\} + \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = F\}), \\ &= P_i^S \left(V_i^S \mathbf{1}\{V_i^S \geq 0, V_i^F < 0\} + \frac{V_i^S}{V_i^F} \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = S\} \right) \\ &\quad + P_i^F (\mathbf{1}\{V_i^F \geq 0, V_i^S < 0\} + \mathbf{1}\{V_i^F \geq 0, V_i^S \geq 0\} \mathbf{1}\{D_i = F\}), \end{aligned}$$

the only unknown is V_i^S when $V_i^S \geq 0, V_i^F < 0$ since $\frac{V_i^S}{V_i^F}$ when $V_i^F \geq 0, V_i^S \geq 0$ is identified (see Section 2.3.3). This partial identification issue comes from the fact that ordinal preferences only are recovered in the case in which only one of the value functions is positive and when both value functions are positive, relative cardinal utilities only can be identified.

Various assumptions are plausible. If there is some positive correlation between V_i^F and V_i^S , we would expect that

$$\begin{aligned} E(V_i^S | V_i^S \geq 0, V_i^F < 0) &< E(V_i^S | V_i^S \geq 0, V_i^F \geq 0) = E\left(\frac{V_i^S}{V_i^F} | V_i^S \geq 0, V_i^F \geq 0\right) \\ &< \exp(X_i \gamma) E(\exp(\varepsilon_i) | V_i^S \geq 0, V_i^F \geq 0) \\ &< \exp(X_i \gamma + 0.5), \end{aligned}$$

the last expression being obtained under normality of ε_i . This is why we assume that when $V_i^S > 0$,

$$\log V_i^S = \frac{\mu_0}{2} V_i^F + \left(\log \frac{V_i^S}{V_i^F} - \frac{\mu_0}{2} \right) |V_i^F| = \frac{\mu_0}{2} V_i^F + \left(X_i \gamma + \varepsilon_i - \frac{\mu_0}{2} \right) |V_i^F|,$$

where $\mu_0 > 0$ captures the positive dependence between V_i^S and V_i^F . This is coherent with the previous equation since

$$\begin{cases} V_i^S = \exp(X_i \gamma + \varepsilon_i) & \text{if } V_i^F = 1, \\ V_i^S = \exp(X_i \gamma + \varepsilon_i - \mu_0) & \text{if } V_i^F = -1. \end{cases}$$

We will thus evaluate $E(U_i | V_i^S, V_i^F)$ using bounds on $\mu = \exp(-\mu_0)$ that we make vary between 0 (the lower bound for V_i^S) and 1 (the case in which V^S and V^F are uncorrelated).

We use this measure of welfare in relative terms among students to evaluate the amount of redistribution between them of changes in the allocation mechanisms.¹⁴

¹⁴These welfare measures could be translated back into changes of odd ratios of expected success probabilities using the preference equation (5) but this does not add much to our evaluation.

4.2 Computing equilibria

In every counterfactual experiment, we draw unknown random terms conditional on observed choices for simulation purposes. This ensures that simulated choices are compatible with observed choices in the data. In each simulation, let \bar{D}_i be the counterfactual choices of the students that depend on counterfactual expectations \bar{P}_i^S and \bar{P}_i^F . Denote $\bar{n}_S = 2n_S$ and $\bar{n}_F = 2n_F$ the new number of seats in the cutting-seat counterfactual. In other cases, $\bar{n}_S = 4n_S$ and $\bar{n}_F = 4n_F$ as in the original system.

Given that historical variables and outside option value do not change, the population of reference does not change in the counterfactual experiments since experiments affect success probabilities only. The pool of applicants remains the set of students whose utilities are such that $V^S > 0$ or $V^F > 0$ and, therefore, we consider the same sample $i = 0, \dots, n$. Consistency of choices and expectations require that the counterfactual random thresholds, \tilde{T}_0 , as defined as the solution $(\tilde{t}_1^S, \tilde{t}_2^S, \tilde{t}_1^F, \tilde{t}_2^F)$ to the counterfactual counterpart of equation (4):

$$\begin{cases} \sum_{i=1}^n [\mathbf{1}\{\bar{D}_i(\bar{P}_i^S, \bar{P}_i^F) = S\} \mathbf{1}\{m_1(X_i, \beta, u_i) \geq \tilde{t}_1^S\}] = \bar{n}_S, \\ \sum_{i=1}^n [\mathbf{1}\{\bar{D}_i(\bar{P}_i^S, \bar{P}_i^F) = F\} \mathbf{1}\{m_1(X_i, \beta, u_i) \geq \tilde{t}_1^F\}] = \bar{n}_F, \\ \sum_{i=1}^n [\mathbf{1}\{\bar{D}_i(\bar{P}_i^S, \bar{P}_i^F) = S\} \mathbf{1}\{m_1(X_i, \beta, u_i) \geq \tilde{t}_1^S, m_2(X_i, \beta, u_i) \geq \tilde{t}_2^S\}] = n_S, \\ \sum_{i=1}^n [\mathbf{1}\{\bar{D}_i(\bar{P}_i^S, \bar{P}_i^F) = F\} \mathbf{1}\{m_1(X_i, \beta, u_i) \geq \tilde{t}_1^F, m_2(X_i, \beta, u_i) \geq \tilde{t}_2^F\}] = n_F, \end{cases} \tag{13}$$

have a distribution function that leads to the counterparts of equation (12):¹⁵

$$\tilde{P}_0^j = \mathbb{E}(\mathbf{1}\{m_1(X_0, \beta, u_0) \geq \tilde{t}_1^j, m_2(X_0, \beta, u_0) \geq \tilde{t}_2^j\}). \tag{14}$$

We thus propose to iterate the following algorithm (we explain it for observation 0 and this extends to any index i):

1. Initialization:

- Draw $C = 499$ random preference shocks $\varepsilon_{(n),c}$ in their distributions conditional to observed choices, D_i , and using preference parameter estimates $\hat{\zeta}_n$. Fix those $\varepsilon_{(n),c}$ for the rest of the procedure (see Supplementary Appendix S.3.1.1 for details).
- Draw C random vectors $U_{(n),c}$ and fix them for the rest of the procedure.
- Set the initial $P_0^{S,0}, P_0^{F,0}$ values at their simulated values $\hat{P}_{0,C}^j$ derived from equation (12) in which we use $U_{(n),c}$ and the observed experiment to compute thresholds $t_{1,c}^j$ and $t_{2,c}^j$ using equations (4).

¹⁵Changing the timing of choices requires to acknowledge that there are no choices to make before the first stage. The first two equations in (13) do not depend on \bar{D}_i and P_i^S, P_i^F are the conditional expectations after the second stage. Those adaptations do not modify the main principles.

2. At step k , denote $P_i^{S,k}, P_i^{F,k}$ the expected success probabilities:

- (a) Compute counterfactual choices $D_i(Z_i, \varepsilon_{i,c}, \hat{\zeta}_n, P_i^{S,k}, P_i^{F,k})$.
- (b) Compute a sequence of \tilde{t}_c for $c = 1, \dots, C$ using $U_{(n),c}$ and equations (13).
- (c) Derive $\hat{P}_{0,C}^{j,k+1}$ from equation (14).

3. Repeat the previous step until a measure of distance $d(P^{(k+1)}, P^{(k)})$ is small enough.

If this algorithm converges, then this is the fixed point we are looking for. We study in Appendix A a simplified model in which we show that a Nash equilibrium is obtained in a finite number of steps.

4.3 Cutting seats at the second-stage exam

We start with the easiest policy change that reduces admission rates to the second stage. As said, the existing *Vestibular* system allows the number of students who take the second exam to be four times the number of final seats. In the experiment, the number of available positions is kept unchanged but the number of admissions after the first-stage exam is now twice the number of seats. We explore the possible consequences of this policy and investigate two main issues—who among students benefit from this policy change and whether schools lose good students.

Some discussion about the expected effects are in order. Cutting seats in the second exam reduces schools' screening costs although this comes with the risk of losing talented students. Students may not be always consistent in their exam performance and even the most gifted may have a strong negative shock in the first exam. Those students would be eliminated too early without being given a second chance. Nonetheless, it could also be that cutting seats protect the first-stage best achievers from competition, and thus from the risk of losing ranks at the second-stage exam. A formal argument is as follows. Subpopulations defined by a specific final weighted score are now composed by more top achievers at the first stage. The net result is however unclear theoretically because the distribution of the final weighted score changes itself and needs to be integrated out. This is why an empirical analysis is worthy of attention.

The simulation of the counterfactual and the computation of expected utility follow procedures described in Section 4.1 and Section 4.2.

4.3.1 Changes in thresholds In Table 3, we present estimates of the new threshold distributions at both stage exams in the three counterfactual experiments. In the cutting seat experiment, the counterfactual first-stage thresholds are much higher than in the original experiment since fewer students are admitted after the first-stage exam. In contrast, the thresholds of the second-stage exam are slightly lower than in the original system because there is now less competition in the second-stage exam when half as many students are admitted. In both first- and second-stage exams, estimated thresholds in Sobral are more volatile than the ones in Fortaleza because Sobral is a much smaller school.

TABLE 3. Thresholds of the counterfactuals.

School			Sobral	Fortaleza
Stage 1	Original system	Mean thresholds	184.48	189.88
		Standard errors	(1.257)	(0.401)
	Cutting seats	Mean thresholds	195.79	201.04
		Standard errors	(0.996)	(0.506)
	Two-choices	Mean thresholds	186.98	190.13
		Standard error	(0.564)	(0.458)
	Timing-change	Mean thresholds	183.05	190.11
		Standard error	(0.859)	(0.447)
Stage 2	Original system	Mean thresholds	235.41	241.44
		Standard errors	(1.669)	(0.898)
	Cutting seats	Mean thresholds	233.34	237.77
		Standard errors	(3.094)	(1.603)
	Two-choices	Mean thresholds	235.38	241.19
		Standard error	(2.589)	(1.302)
	Timing-change	Mean thresholds	239.07	244.30
		Standard error	(2.722)	(1.408)

Note: 1. The coefficients and their standard errors are computed by using the 499 bootstrapped estimates of preference and grade parameters and applying the procedure in the text. 2. The cutting seats counterfactual has a few cases in which the computation developed in Section 4.2 does not converge after many repetitions, and we have excluded those bootstrap values that do not converge after 500 iterations.

To evaluate how this counterfactual brings benefit to schools and students, we study in turn, changes in success probabilities and changes in students' utilities.

4.3.2 Changes in success probabilities Schools would find that the admittance procedure has improved if abler students would get a higher chance of admission and the less gifted students would have a lower chance. This is why we evaluate changes in success probabilities in relation to an index of students' abilities. As our ability index, we use the expected final grade which is, as already said, a combination of the initial, first- and second-stage grades. We also choose to focus on the top 50% of students because the lower 50% of the sample have almost no chance of getting admitted whether the original or counterfactual mechanisms are used.

We represent changes in success probabilities in Figure 2 for Sobral. Three vertical lines are drawn at the median of expected final grade and at the quantiles associated to the first and second-stage thresholds *in the original system* (averaged across schools). Changes in probabilities are very similar in the two schools.¹⁶ The dispersion of these changes, conditional on expected grade, is due to the heterogeneity of observed characteristics across students.

The very top students, who are above the second-stage admission quantile, have better chances in the counterfactual system since they are likely to face less competition in

¹⁶The corresponding figure for Fortaleza appears in Section S.3 of the Supplementary Appendix (Figure S.iv).

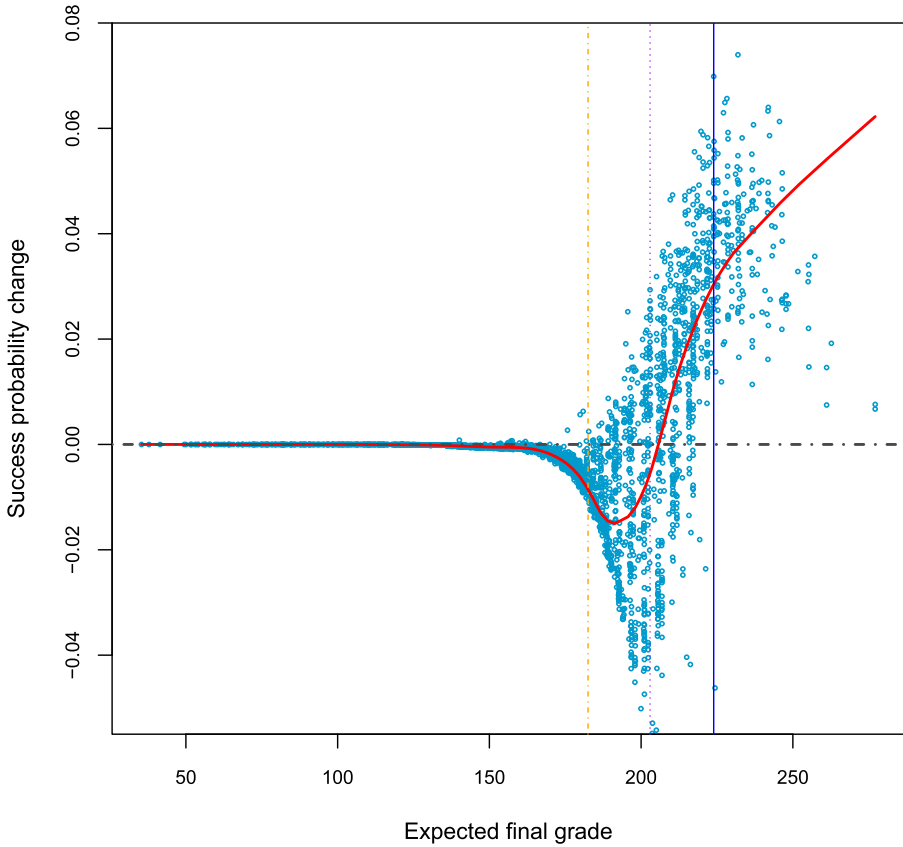


FIGURE 2. Cutting seats: changes of success probabilities in Sobral. [1] The circles plot individual success probability changes versus expected final grades; [2] From left to right, (1) the first vertical line is the median, (2) the second vertical line is the average of quantiles for 1st stage admission— $(1 - \frac{4(nos+nof)}{nobs}) \times 100\%$, and (3) the third line is the average of quantiles for 2nd stage admission— $(1 - \frac{(nos+nof)}{nobs}) \times 100\%$ in which *nos* is the number of final seats in Sobral, *nof* is the number of final seats in Fortaleza, and *nobs* is the number of total applicants; [3] The solid fitted curve is obtained by lowess smoothing.

the second-stage exam. Our estimates of grade equations show that second-stage grades have a much larger variance than first-stage grades. The risk of failing is thus lower when fewer students participate in the second-stage exam. In contrast, for students who are between the median and first-stage threshold in terms of expected final grades, this is the converse. They are much less often admitted after the first-stage exam and even if the second-stage exam is less competitive, it is the former negative effect that dominates overall. In particular, students who are around the first-stage threshold in the current system are more likely to be selected out at the first stage.

4.3.3 Changes in students' utilities and the impact on schools Table 4 presents summaries of changes in students' expected utility. We construct groups according to various quantiles of the distribution of the expected final grade. The closer to the top of

TABLE 4. Cutting seats: expected utility changes.

Expected Final Grade	ALL		$D = \text{Sobral}$		$D = \text{Fortaleza}$	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
0%–50%	–0.00029	0.00116	–0.00109	0.00185	–0.00011	0.00084
50%–60%	0.00001	0.00744	–0.00733	0.00563	0.00252	0.00622
60%–70%	0.00674	0.01655	–0.01125	0.00931	0.01206	0.01433
70%–80%	0.03122	0.02770	–0.00919	0.01272	0.03843	0.02306
80%–82%	0.04070	0.02813	–0.00096	0.00413	0.05493	0.01574
82%–84%	0.05491	0.02896	0.00540	0.00539	0.06567	0.01887
84%–86%	0.07304	0.03128	0.00107	0.00866	0.08482	0.01123
86%–88%	0.06124	0.03374	0.00257	0.00837	0.07762	0.01367
88%–90%	0.07932	0.03072	0.00813	0.00484	0.09027	0.01308
90%–92%	0.09239	0.03272	0.00617	0.00969	0.10224	0.01463
92%–94%	0.08806	0.04041	0.00991	0.00905	0.10696	0.01227
94%–96%	0.11009	0.03125	0.00834	0.00213	0.11839	0.01114
96%–98%	0.11178	0.03456	0.01104	0.00531	0.12249	0.01008
98%–100%	0.08939	0.04669	0.00760	0.00533	0.11185	0.01989
$E(\Delta U_i)$	0.01966		–0.00267		0.02487	
s.d. (ΔU_i)	0.03785		0.00817		0.04009	
Frequency ($\Delta U_i > 0$)	0.4363		0.2084		0.4894	

Note: 1. ALL contains all the students no matter what the original choices are. 2. $D = \text{Sobral}$ means the subpopulation of those who chose Sobral in the original system; and $D = \text{Fortaleza}$ means the subpopulation of those who chose Fortaleza in the original system. 3. $E(\Delta U_i)$ (resp., **s.d.**(ΔU_i)) is the sample average (resp., standard deviation) of the total utilitarian welfare change. 4. **Pr**($\Delta U_i > 0$) is the frequency of students whose expected utility changes are positive.

the distribution, the smaller the groups are (2% of the population only). As defined in Section 4.1, we set the unknown weight in utilities at $\mu = 0.8$.¹⁷

Consistently with changes in success probabilities, top students have significant utility improvements although this is also true for lower ranked students (above the 80% quantile). Nonetheless, focusing on means of expected utility hide very large dispersions in the 80–90% quantiles. This is best seen in the distribution of changes in utility (Supplementary Appendix S.1, Figure S.v) in which students in the 8th and 9th deciles are the ones whose changes in utility are the most dispersed. Furthermore, students just above the median tend to have lower expected utility in the counterfactual system and this is consistent with what we obtained for success probabilities. If we divide the sample by the original school choice, an indication of their preference, students who chose Fortaleza tend to benefit more than the ones who opted for Sobral. The influence of the second-stage exam seems to be much larger there than in Sobral. Overall, these results about this counterfactual experiment bring out a moderate total utilitarian welfare change. Yet, there are strong distributional effects and top students are better off and less able students are worse off.

The impact of cutting seats seems favorable for schools since the most able students now have a higher chance of admission since they are protected from the competition

¹⁷We also performed robustness checks by using weights μ varying from 0 to 1 (see Section 4.1). Results are shown in Table S.viii in the Supplementary Appendix. Differences are very small and our results are quantitatively robust to the value of μ .

of less able students at the second stage. This benefit comes in addition to cutting the costs of organizing and correcting the second-stage exam proofs.

4.4 *Enlarging the choice set*

In this experiment, students can submit an enlarged list of two schools if they wish. A choice list contains two elements d_1 and d_2 in which d_1 is the preferred school. Since our sample of interest only comprises students who positively value at least one of the schools, we have $d_1 \in \{S, F\}$. Yet, students can now apply to a second choice and $d_2 \in \{\emptyset, S, F\}/\{d_1\}$ in which $d_2 = \emptyset$ is the outside option chosen by students who do not give positive value to the second school. This mechanism belongs to the deferred-acceptance family with the additional twist that we keep the sequence of two exams as it is. The allocation of students after the first exam needs however to be adapted and this is the design that we now explain.

4.4.1 *Design of the experiment* To fix ideas, consider first a student who (1) has $V_S > 0$ and $V_F > 0$, and (2) chooses the list (S, F) . If after the first-exam, she is above the threshold for school S , her second choice does not matter.¹⁸ It is only if she would NOT be accepted to the second-stage exam in school S that she could compete for the second-stage exam in school F .¹⁹ She fails altogether when her grades are lower than both thresholds.

Consider first that at equilibrium $t_1^S > t_1^F$. After the first-stage exam, there are three possible outcomes for the student:

- $m_1 \geq t_1^S$: she takes the second exam of school S ,
- $m_1 < t_1^S$ and $m_1 \geq t_1^F$: she takes the second-stage exam of school F ,
- $m_1 < t_1^F$: she fails and takes the outside option.

While if $t_1^S < t_1^F$ (the probability of a tie being equal to zero),

- $m_1 \geq t_1^S$: she takes the second exam of school S ;
- $m_1 < t_1^S$: she fails and takes the outside option.

This sequence is easily adapted to students choosing the list (F, S) . Moreover, for students submitting a list (d_1, \emptyset) , the sequence of actions is the same as in the original mechanism. Students are selected into the second-stage exam for school d_1 if their grade is above its first-stage threshold.

Furthermore, given any choice among the four lists, $\{(S, F), (F, S), (S, \emptyset), (F, \emptyset)\}$ we can construct counterfactual success probabilities in each school P^S and P^F by adapting the algorithm we used before (see Supplementary Appendix S.3.2). For any value of success probabilities, we can then compute the optimal choice between $\{(S, F), (F, S), (S, \emptyset), (F, \emptyset)\}$. Details about how we get counterfactual thresholds and choices follow the lines of what was developed in Section 4.2.

¹⁸In particular, we discard the possibility of choosing a second ranked school after a success at the first-stage exam.

¹⁹See also the third experiment in which students choose according to the information they have on their performance at the first stage for a variation around these constraints.

4.4.2 Changes in thresholds The new thresholds for this counterfactual experiment are also shown in Table 3. For the first stage, the threshold of Sobral is now slightly larger than the original one while the threshold of Fortaleza remains roughly unchanged. This is an indication that Sobral is admitting better students while the effect on Fortaleza is negligible. Some of the students who were failing Fortaleza before can now compete for Sobral and get admitted after the first stage. Furthermore, some of the students who were choosing Sobral for strategic reasons in the original mechanism can now at no risk choose Fortaleza first and Sobral second. Deferred acceptance mechanisms lessen strategic motives and make choices more truthful (Abdulkadiroğlu and Sönmez (2003)) although the move is not necessarily Pareto-improving (Balinski and Sönmez (1999)). In the original system, students tended to choose Sobral as a “safety school” even when they truly preferred Fortaleza since success probabilities were higher at the former school. Giving students two choices attenuates the “safety school” effect although it does not eliminate it completely because of the two-stage nature of the exam. Yet, thresholds for the school in Fortaleza remains higher than for Sobral at both stages because it attracts more top-ability (m_0) students as is shown by preference estimates (see Table S.vii).

Large standard errors for counterfactual thresholds at the second-stage exam make differences with the current ones insignificant. Even if this counterfactual experiment moves some of the relatively good students after the first-stage exam from Fortaleza to Sobral, Sobral however still attracts less able students than Fortaleza in the second stage as in the first stage.

4.4.3 Changes in success probabilities Figure 3 reports changes in success probabilities for Sobral (see Figure S.vi for Fortaleza). Unlike the previous counterfactual experiment, the changes in Sobral and Fortaleza are now somewhat different. In Fortaleza, the change in success probabilities is negligible as thresholds are constant and the reallocation of choices from Sobral to Fortaleza not strong enough. In contrast, a fraction of students below the first admission threshold and above median has a lower success probability in Sobral in the counterfactual experiment. This is because better students who fail Fortaleza switch to Sobral to compete with them and lower ranked students are evicted since first-stage thresholds are now higher in Sobral. In other words, getting Sobral if failing Fortaleza is acting as an insurance device and students just above the first-stage threshold benefit from the existence of this insurance. Last, note that the change in success probabilities is small in this counterfactual compared with cutting seats since it affects students only through the allocation mechanism.

4.4.4 Changes in expected utilities and the impact on schools From the student perspective, this mechanism is also attractive since a majority of students—55%—will be (strictly) better off as shown in Table 5. Moreover, top students benefit more from the change than less able students because they are more likely to pass to the second-stage exam even if they happen to fail their preferred school. Deferred acceptance restricts less the possibilities of very top students since they can keep options open. In particular, students who preferred Sobral initially, benefit much more than those who preferred Fortaleza initially, seemingly because the pressure of competition at the top in Sobral is

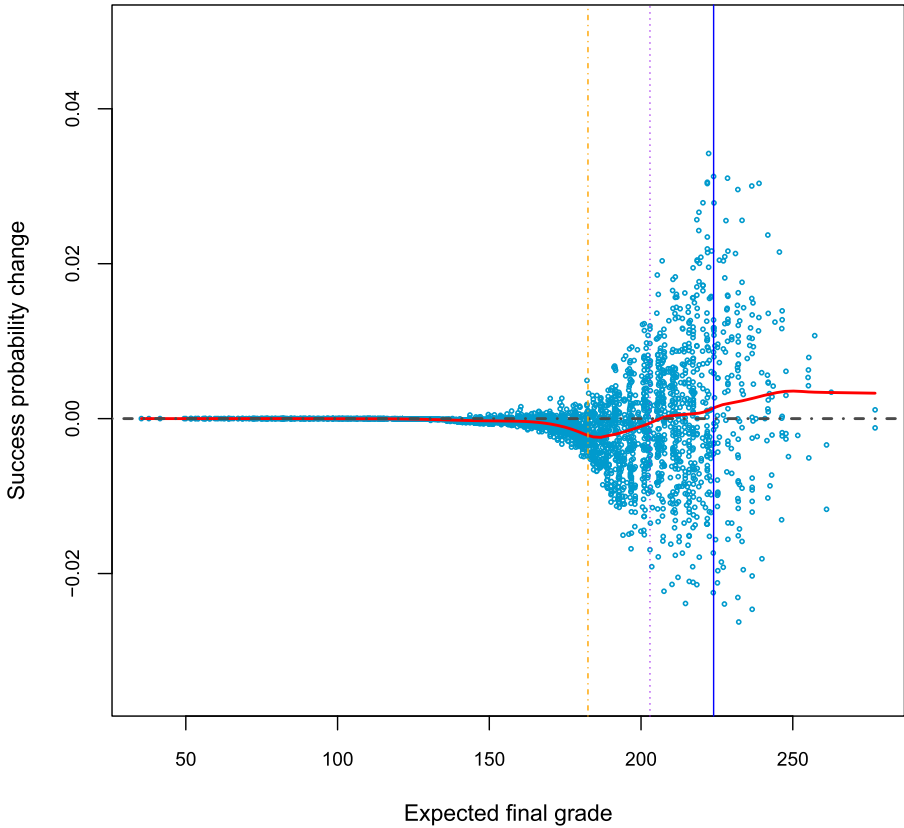


FIGURE 3. Two choices: success probability change in Sobral. Notes: See the notes of Figure 2.

lower since it loses its safety school status. In contrast, since Sobral has a lower threshold at the first-stage exam, students who prefer Sobral and are ranked around the first-stage threshold suffer from more competition from evicted students from Fortaleza. However, for those who preferred Fortaleza in the original system, expected utility mainly increases because of the second chance they get to compete for Sobral when they fail Fortaleza. The effect on expected utility is thus larger than the change in success probabilities.

In summary, enlarging the choice set improves the average ability of those who pass the first-stage exam in both schools. The majority of students are better off except students ranked around the first-stage threshold in the original system and who prefer the smallest school. From the perspective of the schools, Sobral should be more favorable to this mechanism since it can now attract higher ranked students. Fortaleza's thresholds remain the same although the composition of their recruitment might have changed since Sobral lost its safety school status. This seems however to moderately affect top students.

This confirms theoretical insights that the move to a deferred acceptance mechanism is likely to make both schools and more top students better off.

TABLE 5. Two choices: expected utility changes.

Expected Final Grade	ALL		<i>D</i> = Sobral		<i>D</i> = Fortaleza	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
0%–50%	0.00002	0.00076	0.00048	0.00144	−0.00009	0.00043
50%–60%	0.00176	0.00372	0.00537	0.00531	0.00052	0.00174
60%–70%	0.00727	0.01006	0.02106	0.01084	0.00320	0.00487
70%–80%	0.01706	0.01907	0.05619	0.01400	0.01008	0.00843
80%–82%	0.03719	0.03064	0.08629	0.00523	0.02042	0.01126
82%–84%	0.03140	0.03081	0.09163	0.00784	0.01831	0.01291
84%–86%	0.02817	0.03303	0.10573	0.00620	0.01548	0.01003
86%–88%	0.04673	0.03837	0.11457	0.00713	0.02780	0.01406
88%–90%	0.04323	0.03573	0.12548	0.00781	0.03058	0.01563
90%–92%	0.03728	0.03984	0.14298	0.01055	0.02520	0.01731
92%–94%	0.05830	0.04879	0.14871	0.00588	0.03643	0.02148
94%–96%	0.04137	0.04184	0.17341	0.00454	0.03059	0.01799
96%–98%	0.05055	0.04687	0.18008	0.00424	0.03677	0.02042
98%–100%	0.05964	0.06562	0.17849	0.01288	0.02702	0.02069
E (ΔU_i)	0.01143		0.03074		0.00693	
s.d. (ΔU_i)	0.02705		0.05073		0.01400	
Frequency ($\Delta U_i > 0$)	0.5431		0.7029		0.5058	

Note: 1. ALL contains all students no matter what the original choices are. 2. *D* = Sobral means the subpopulation of those who chose Sobral in the original system; and *D* = Fortaleza means the subpopulation of those who chose Fortaleza in the original system. 3. See the notes of Table 4.

4.5 Changing the timing

In the last counterfactual experiment, we try to evaluate the impact on students when they choose schools *after* learning their first-stage exam grade and no longer before this exam. Schools continue to rank students according to the same combination of *ENEM* and m_1 .

The new selection procedure is a serial dictatorship mechanism which is Pareto-optimal in the case of a single exam (for instance, [Abdulkadiroğlu and Sönmez \(1998\)](#)). It proceeds as follows. Starting from the first-ranked student and going down the ranking afterwards, each student chooses school *S* or *F* until the number of admitted students in one of the schools, say *j*, reaches four times the number of final seats in this school. This defines threshold t_1^j . The sequence continues going down the ranking although choice is now restricted to the other school $D \neq j$ or to opting out until the number of admitted students in that school reaches four times the number of final seats. The allocation of students to the second-stage exam is then complete. The game continues afterwards as in the current system.

As before, utilities V^S and V^F remain the same while this new mechanism affects the probabilities of success $P_{m_1}^S = \Pr\{m_2 > t_2^S | m_1\}$ and $P_{m_1}^F = \Pr\{m_2 > t_2^F | m_1\}$ which are now conditional on the first-stage grade m_1 . To define choices, suppose that $t_1^S > t_1^F$ which means in practice that Sobral seats are filled in faster than Fortaleza's. A student can face three cases:

- $m_1 > t_1^S$: the choice set is complete and consists in $\{S, F\}$. Schools are chosen by comparing $P_{m_1}^S V^S$ and $P_{m_1}^F V^F$ (since either $V^S > 0$ or $V^F > 0$).
- $m_1 < t_1^S$ and $m_1 \geq t_1^F$: the choice set is restricted to F and the student either opts for the second-stage exam in F if $V^F > 0$ or the outside option if not.
- $m_1 < t_1^F$: the only choice left is the outside option.

This algorithm is easily adapted to the case in which $t_1^S < t_1^F$ prevails. Additionally, we compute the same ex ante expected utilities by integrating out shocks in m_1 .

4.5.1 Changes in thresholds The new thresholds in this counterfactual experiment are shown in Table 3. Sobral has now a slightly lower threshold at the first stage and a slightly higher threshold at the second-stage exam while this is true but at the second stage for Fortaleza. The school in Fortaleza is overall more popular (see Table S.vii) and even more than the difference in offered seats. By making students choose in the order of first-stage grades, positions in Sobral at the second-stage exam are less likely to be filled earlier than Fortaleza's despite the one to four ratio (160/600). For instance, if more than 80% of the top 750 students prefer Fortaleza to Sobral, the 600 seats at Fortaleza would be filled in after those 750 students would reveal their choices while Sobral would still have 10 seats to fill in. Note that in simulations, such a solution can be very unstable with respect to the random draws of grade shocks and depend very much on revealed preferences and the first-stage randomness in selecting the set of students who can go to the second stage.

4.5.2 Changes in success probabilities Changes in success probabilities in Sobral are shown in Figure 4. Success probabilities, evaluated ex ante, now depend more on the first stage than before so that students performing well at the first stage increase their overall success probabilities while those performing worse have now lower success probabilities. There is also a large dispersion of these changes. Ex post dispersion increases with the final expected grade because it increases with the level of the initial success probabilities and this confirms the increasing importance of the first-stage grade. These conclusions are true for Fortaleza (see Section S.3) as well.

4.5.3 Changes in expected utilities and the impact on schools As this mechanism introduces an element of flexibility for students since they can condition their choices on their first-stage grades, their expected utility is on average mechanically larger than in the original system. Indeed, the frequency of an increase in expected utility is the largest in the three experiments. This mechanism is mainly attractive for the top students as shown in Table 6. In a nutshell, top students in the first stage are better protected from the competition of lower ranked students.

There are clear differences in utility changes among the top students conditional on their preferences for the schools. On average, students who were choosing Fortaleza in the original system would benefit more than those who preferred Sobral. This seems to be due to the difference in the sizes of the school because of the argument presented above when we were analyzing the impact on thresholds. Sobral seats are filled less quickly than Fortaleza's.

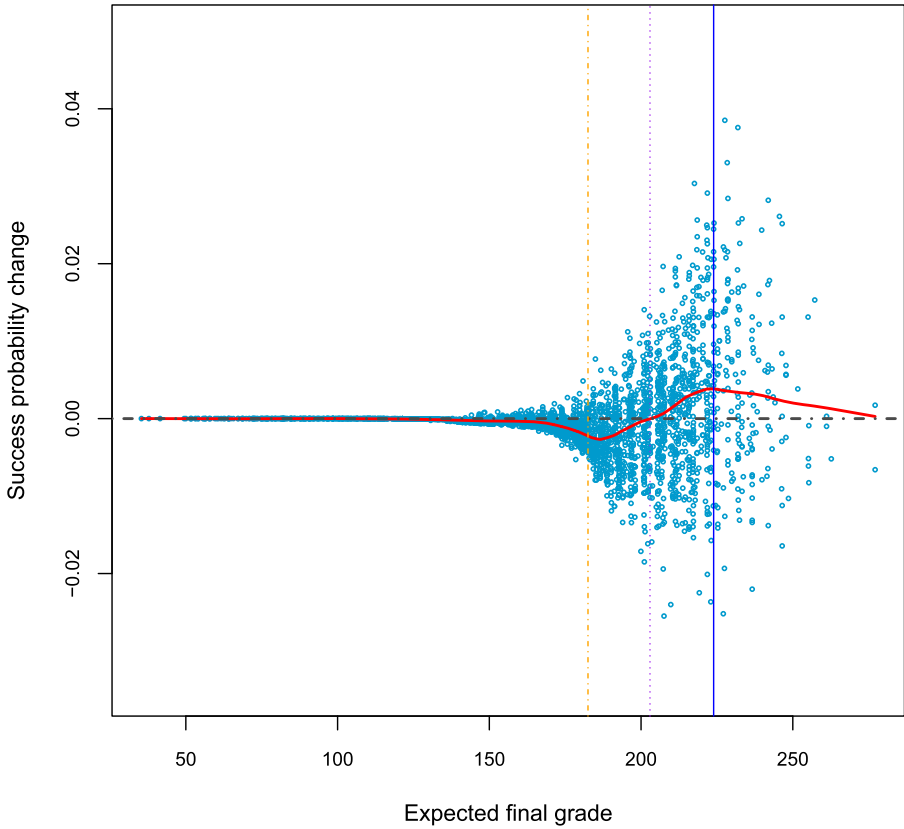


FIGURE 4. Timing change: success probability changes in Sobral. Notes: See the notes of Figure 2.

Overall, this counterfactual seems more friendly to top students. Nonetheless, such a system seems to select students with lower future academic success as shown by the analysis of [Wu and Zhong \(2014\)](#) using historical data on China provinces which have changed allocation mechanisms in this direction. Our data is too limited to explore this issue.

5. CONCLUSION

In this paper, we use data from entry exams and an allocation mechanism to colleges to provide an evaluation of changes in those mechanisms. We first use a model of school choices as well as performance to estimate parameters governing success probabilities and preferences. Expectations of sophisticated students are obtained by sampling into the Nash equilibrium conditions. Using those estimates, we can compute in a second step the impact of three counterfactual experiments on success probabilities and expected utility of students. This shows at what benefits and costs the current mechanism could be changed, not only in terms of aggregate utilitarian welfare but also in terms of potentially strong redistributive effects between schools and between students.

TABLE 6. Timing change: expected utility changes.

Expected Final Grade	ALL		$D = \text{Sobral}$		$D = \text{Fortaleza}$	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
0%–50%	0.00053	0.00158	–0.00019	0.00065	0.00070	0.00168
50%–60%	0.00731	0.00584	–0.00064	0.00182	0.01003	0.00394
60%–70%	0.01798	0.01034	0.00196	0.00291	0.02271	0.00612
70%–80%	0.03551	0.01394	0.00691	0.00491	0.04062	0.00722
80%–82%	0.04164	0.01984	0.00954	0.00386	0.05260	0.00654
82%–84%	0.04821	0.01877	0.01087	0.00588	0.05633	0.00683
84%–86%	0.05375	0.01971	0.00818	0.00574	0.06121	0.00674
86%–88%	0.05348	0.02207	0.01425	0.00571	0.06443	0.00745
88%–90%	0.06125	0.02169	0.01046	0.00434	0.06907	0.00865
90%–92%	0.06933	0.02109	0.01364	0.00633	0.07569	0.00932
92%–94%	0.06514	0.02591	0.01646	0.00808	0.07691	0.00989
94%–96%	0.07891	0.02258	0.01366	0.00598	0.08424	0.01289
96%–98%	0.07912	0.02262	0.01773	0.00570	0.08566	0.01054
98%–100%	0.06413	0.02878	0.01581	0.00784	0.07739	0.01454
E (ΔU_i)	0.01862		0.00288		0.02229	
s.d. (ΔU_i)	0.02703		0.00609		0.02865	
Frequency ($\Delta U_i > 0$)	0.6557		0.5166		0.6881	

Note: 1. ALL contains all students no matter what the original choices are. 2. $D = \text{Sobral}$ means the subpopulation of those who chose Sobral in the original system; and $D = \text{Fortaleza}$ means the subpopulation of those who chose Fortaleza in the original system. 3. See the notes of Table 4.

These cost-benefit analyses show that the choice of an allocation mechanism has sizeable consequences for both schools and students. The mechanism in place is neither fair nor strategic although it might be rationalized by the fact that some schools and/or groups of students would lose if it were changed. The political economy of such a choice of an allocation mechanism remains to be documented and analyzed and it would be interesting to develop the analysis of the ex ante game between schools and/or students that leads to the adoption of such or such mechanisms of selection and allocation. As a matter of fact, Federal universities in Brazil adopted in 2010, under pressure of the Federal government, a national allocation mechanism consisting of student submissions of a list of two preferred schools and a complicated learning mechanism. Some of us are in the process of collecting data to evaluate this new system.

Nonetheless, the previous mechanism allowed schools to tailor their selection procedures to the information they had about the prerequisites for their courses and any predictors of success or drop out of the students they selected. This fine tuning is lost in the new centralized procedure which abstracts away from the question of acquiring information that determines school preferences (Coles, Kushnir, and Niederle (2013)). Specifically, the new allocation mechanism used in Brazil (for instance, analyzed in Machado and Szerman (2017)) is based on a single grade given by an improved version of ENEM which nevertheless remains of poorer quality than the Vestibular analyzed in this paper since the additional information yielded by the two-stage exams is now lost. Universities were also reducing opportunistic behavior as shown by the last counterfactual since knowing results at the first-stage exam allows students to strategize better.

Our selection of two elite medical schools is admittedly specific and tailored to minimize departures from our simplifying assumptions. As preferences for these two schools are presumably closer than any other pair of schools, the impact of treatment on outcomes—that is, success probabilities and school choice—might be magnified by this selection. Whether this larger impact is translated into larger welfare effects is, however, ambiguous since differences between preferences are smaller.

On the modeling side, much remains to be done. Specifically, the modeling assumptions about expectations are strong and weakening them is high on the agenda. Identification however is bound to be weak since there is nothing in our data that might indicate whether agents are sophisticated, well or badly informed or even naive (He (2017), Agarwal and Somaini (2018)). The analysis shall thus proceed as an analysis of robustness that could lead to partial identification of the costs and benefits we have been describing above. It is also true that the question of why so many students are taking this exam although they have no chances to succeed remains pending. They could be overly optimistic and this relates to assumptions about expectations but they could also use the exam as a training device for the following year or for other exams of a similar type. This behavior seems to be easier to accommodate in the current framework.

APPENDIX A: EXISTENCE OF A NASH EQUILIBRIUM AND CONVERGENCE TO AN EQUILIBRIUM

When using the current mechanism or counterfactual experiments, the question of the existence of a Nash equilibrium is pending. This equilibrium is defined as the solution to the best response equations (1) and success probabilities that are mutually compatible and compatible with the equilibrium conditions (4). We rely on the theory of pseudo potential games as developed in Dubey, Haimanko, and Zapechelnyuk (2006)

In this discussion, we sketch the proof in a simpler game restricted to two schools $j \in \{S, F\}$ and a single stage exam and imposing some weak conditions. The extension to more schools or two exams complicates notation but does not affect the intuition. Conditions (4) become:

$$\sum_{i=1}^n [\mathbf{1}\{D_i = S\} \mathbf{1}\{m_1(Z_i, u_i, \beta) \geq t_1^S\}] = 4n^S,$$

$$\sum_{i=1}^n [\mathbf{1}\{D_i = F\} \mathbf{1}\{m_1(Z_i, u_i, \beta) \geq t_1^F\}] = 4n^F.$$

We will also assume that both schools are overdemanding by students who do not value positively both schools, that is,

$$\sum_{i=1}^n \mathbf{1}\{V_i^S > 0 \geq V_i^F\} > 4n^S,$$

$$\sum_{i=1}^n \mathbf{1}\{V_i^F > 0 \geq V_i^S\} > 4n^F,$$
(15)

so that thresholds in (4) are always defined by equalities.

Setting $\lambda_j(D_{(n)}) = \frac{4n^j}{\sum_{i=1}^n \mathbf{1}_{\{D_i=j\}}}$, we can write an explicit definition of the thresholds as the empirical $(1 - \lambda_j)$ -quantile of the distribution of grades in the sample of applicants to j :

$$T_1^j(Z_{(n)}^j, U_{(n)}^j) = F_{\{m_1(Z_i, u_i, \beta), D_i=j\}}^{-1}(1 - \lambda_j).$$

Note that the strategies of other students affect λ_j as well as the quantile so that expected success probabilities can be written as

$$P_0^j(D_{(n)}) = \mathbb{E}(\mathbf{1}\{m_1(Z_0, u_i, \beta) \geq T_1^j(Z_{(n)}^j, U_{(n)}^j)\}).$$

It is easy to formulate deep assumptions about the distribution function of grades that imply that the success probabilities strictly decrease when adding an additional competitor to the set of applicants to d . Indeed, let us order the strategy set $\{S, F\}$ as $S > F$. Extend the order to a partial order in strategies $D_{(n)}$ in the sample by positing that

$$D_{(n)} > D'_{(n)} \quad \text{iff} \quad D_i \geq D'_i \quad \text{and} \quad \text{for at least one } i \quad D_i > D'_i.$$

If the distribution of grade shocks is unbounded, adding competitors creates congestion and we have that

$$D_{(n)} > D'_{(n)} \quad \implies \quad P_0^S(D_{(n)}) < P_0^S(D'_{(n)}) \quad \text{and} \quad P_0^F(D_{(n)}) > P_0^F(D'_{(n)}).$$

It is now straightforward to prove that the game satisfies the *dual strong single crossing property*. Suppose indeed that $V_0^S > 0$ and that

$$P_0^S(D'_{(n)})V_0^S \leq P_0^F(D'_{(n)})V_0^F.$$

This implies

$$P_0^S(D_{(n)})V_0^S < P_0^S(D'_{(n)})V_0^S \leq P_0^F(D'_{(n)})V_0^F < P_0^F(D_{(n)})V_0^F.$$

This is also trivially satisfied when $V_0^S \leq 0$ and $V_0^F > 0$.

As this property of *dual strong single crossing* implies that this is a game of weak strategic substitutes with aggregation (Dubey, Haimanko, and Zapechelnyuk (2006)), it is a pseudo-potential game (Theorem 1, p. 81) and it has a Nash equilibrium (Proposition 1, p. 84). Furthermore, since the strategy set is finite, there are no best response cycles in the game. “If players start with an arbitrary strategic profile and each player (one at a time) unilaterally deviates to his unique best reply, then the process terminates in a Nash equilibrium after finitely many steps” (Remark 1, p. 85).

APPENDIX B: PROOFS IN SECTION 2

B.1 Proof of Proposition 1

Fix $\mathcal{J}_0 \subset \mathcal{J}$, a set of nonempty indices. The probability that the observed choice belongs to \mathcal{J}_0 , $\Pr(D \in \mathcal{J}_0 \mid Z, \max_{j \in \mathcal{J}} (V^j) > 0, X)$ is identified a.e. $P_{X,Z}$ and by equation (6) is equal to

$$\sum_{(j, \mathcal{J}_+); \mathcal{J}_+ \supset \{j\} \subset \mathcal{J}_0} Q(\mathcal{J}_+ \mid X) \Pr(\forall l \in \mathcal{J}_+, \Delta^{jl}(Z) > \log(V^l) - \log(V^j) \mid X, Z, \mathcal{J}_+, \mathcal{J}_+^c).$$

Because of Assumption CV, the support of any vector $\{\Delta^{jk}(Z)\}_{k \in \mathcal{J}/\{j\}}$ is $(-\infty, +\infty)^{\text{card}(\mathcal{J})-1}$. Consider the limit when, for all $j \in \mathcal{J}_0$ and all $k \in \mathcal{J}_0^c$, $\Delta^{jk}(Z)$ tends to $-\infty$,

$$\lim_{\forall (j,k) \in \mathcal{J}_0^c \times \mathcal{J}_0^c, \Delta^{jk}(Z) \rightarrow -\infty} \Pr(D \in \mathcal{J}_0 \mid Z, \max_{j \in \mathcal{J}} (V^j) > 0, X).$$

For all $k \in \mathcal{J}_0^c$, conditions $\Delta^{jk}(Z) > \log(V^k) - \log(V^j)$ are never satisfied and the limit above is thus equal to

$$\begin{aligned} & \sum_{(j, \mathcal{J}_+); \mathcal{J}_+ \supset \{j\} \subset \mathcal{J}_0} Q(\mathcal{J}_+ \mid X) \Pr(\forall l \in \mathcal{J}_+ \cap \mathcal{J}_0, \Delta^{jl}(Z) > \log(V^l) - \log(V^j) \mid X, Z, \mathcal{J}_+, \mathcal{J}_+^c) \\ &= \sum_{\mathcal{J}_+; \mathcal{J}_+ \subset \mathcal{J}_0} Q(\mathcal{J}_+ \mid X) \equiv Q^*(\mathcal{J}_0 \mid X) \end{aligned}$$

because the terms in the first line, $\Pr(\forall l \in \mathcal{J}_+ \cap \mathcal{J}_0, \dots)$ sum to one over $j \in \mathcal{J}_0$ for all $\mathcal{J}_+ \subset \mathcal{J}_0$. In consequence, $\forall \mathcal{J}_0 \subset \mathcal{J}$ and \mathcal{J}_0 nonempty, $Q^*(\mathcal{J}_0 \mid X)$ is identified a.e. P_X .

Consider now that $\mathcal{J}_0 = \{j\}$ is a singleton. Then $Q(\{j\} \mid X) = Q^*(\{j\} \mid X)$ is identified. By induction, suppose that for $K \geq 2$, $Q(\mathcal{J}_K \mid X)$ is identified for all \mathcal{J}_K such that $\text{card}(\mathcal{J}_K) = K$. Consider \mathcal{J}_{K+1} with $\text{card}(\mathcal{J}_{K+1}) = K + 1$ and

$$Q^*(\mathcal{J}_{K+1} \mid X) = \left[\sum_{\mathcal{J}_K; \mathcal{J}_K \subset \mathcal{J}_0, \text{card}(\mathcal{J}_K)=K} Q(\mathcal{J}_K \mid X) \right] + Q(\mathcal{J}_{K+1} \mid X)$$

which proves that $Q(\mathcal{J}_{K+1} \mid X)$ is identified. As this is true for $K = 1$, and if true for K , true for $K + 1$, $Q(\mathcal{J}_0 \mid X)$ is identified for all $\mathcal{J}_0 \subset \mathcal{J}$. \square

B.2 Proof of Proposition 2

We proceed by induction over the number of schools, J . The two-school case is proved in the text. To design the proof at the simplest level, we first derive the proof for $J = 3$ and $\mathcal{J} = \{1, 2, 3\}$. The general proof will follow the same lines but at a more abstract level and will show that if it is true for J , this is also true for $J + 1$.

Stage 1: From two schools to $J = 3$ Write the observed choice probabilities in equation (6) when $\Delta^{13}(Z) \rightarrow -\infty$, which is permitted by Assumption CV:

$$\begin{aligned} \Pr(D = 1 \mid Z, X) &= Q(\{1\} \mid X) \\ &+ Q(\{1, 2\} \mid X) \Pr(\Delta^{12}(Z) > \log(V^2) - \log(V^1) \mid X, Z, \mathcal{J}_+ = \{1, 2\}, \mathcal{J}_+^c = \{3\}), \end{aligned}$$

since alternative 1 is always dominated by alternative 3 when both V_1 and V_3 are positive. Given that $Q(\cdot)$ is identified and different from zero, this identifies

$$\Pr(\Delta^{12}(Z) > \log(V^2) - \log(V^1) \mid X, Z, \mathcal{J}_+ = \{1, 2\}, \mathcal{J}_+^c = \{3\}).$$

By generalizing this line of argument to any pair $\{j, k\}$ in $\{1, 2, 3\}$, this proves the identification of distributions in all quadrants of reduced dimension, $J = 2$.

We can return to equation (6)

$$\begin{aligned} & \Pr(D = j \mid Z, X) \\ &= \sum_{\mathcal{J}_+ \supset \{j\}} Q(\mathcal{J}_+ \mid X) \Pr(\forall k \in \mathcal{J}_+, \Delta^{jk}(Z) > \log(V^k) - \log(V^j) \mid X, Z, \mathcal{J}_+, \mathcal{J}_+^c) \end{aligned}$$

in which all terms are identified except when set $\mathcal{J}_+ = \mathcal{J}$. As $Q(\mathcal{J} \mid X)$ is positive by assumption, we derive from this equation an expression for $\Pr(\forall k \in \mathcal{J}, \Delta^{jk}(Z) > \log(V^k) - \log(V^j) \mid X, Z, \mathcal{J}_+ = \{1, 2, 3\}, \mathcal{J}_+^c = \emptyset)$ for all $j \in \mathcal{J}$ as a function of identified terms. By the complete variation assumption CV of $\Delta^{jk}(Z)$, this ensures the identification of the joint distribution $\Pr((\log(V^k) - \log(V^l))_{\forall k \in \mathcal{J}/\{l\}} \mid X, Z, \mathcal{J})$ if 1 is taken as the reference alternative. The property under induction is thus true for $J = 3$.

Stage 2: From J to $J + 1$ Assume now that the property is true for J . We now show that the property is true for $J + 1$. It follows the same steps as above:

(i) Assume that for all $j \in \mathcal{J}/\{l\}$, $\Delta^{jl}(Z) \rightarrow -\infty$ which identifies, through equation (6), for any $j \in \mathcal{J}/\{l\}$:

$$\Pr(\forall k \in \mathcal{J}/\{j, l\}, \Delta^{jk}(Z) > \log(V^j) - \log(V^k) \mid X, Z, \mathcal{J}_+ = \mathcal{J}/\{l\}, \mathcal{J}_+^c = \{l\}).$$

(ii) Return to equation (6)

$$\begin{aligned} & \Pr(D = j \mid Z, X) \\ &= \sum_{\mathcal{J}_+ \supset \{j\}} Q(\mathcal{J}_+ \mid X) \Pr(\forall k \in \mathcal{J}_+, \Delta^{jk}(Z) > \log(V^k) - \log(V^j) \mid X, Z, \mathcal{J}_+, \mathcal{J}_+^c) \end{aligned}$$

in which all terms are identified except the one corresponding to $\mathcal{J}_+ = \mathcal{J}$. As $Q(\mathcal{J} \mid X)$ is positive, we can derive an expression for $\Pr(\forall k \in \mathcal{J}/\{j\}, \Delta^{jk}(Z) > \log(V^k) - \log(V^j) \mid X, Z, \mathcal{J})$ for all $j \in \mathcal{J}$. By the complete variation assumption CV of $\Delta^{jk}(Z)$, this ensures the identification of the joint distribution $\Pr((\log(V^k) - \log(V^l))_{k \in \mathcal{J}/\{l\}} \mid X, Z, \mathcal{J})$ if 1 is taken as the reference alternative. Identification of differences between log values is thus true for $J + 1$. \square

REFERENCES

- Abdulkadiroğlu, A., N. Agarwal, and P. A. Pathak (2017), “The welfare effects of coordinated assignments: Evidence from the NYC high school match.” *American Economic Review*, 107 (12), 3635–3689. [1234]
- Abdulkadiroğlu, A., P. Pathak, A. E. Roth, and T. Sönmez (2006), “Changing the Boston school choice mechanism.” WP 11965, National Bureau of Economic Research. [1236]
- Abdulkadiroğlu, A., P. A. Pathak, and A. Roth (2009), “Strategy-proofness versus efficiency in matching with indifferences; redesigning the NYC high school match.” *American Economic Review*, 99 (5), 1954–1978. [1236]

- Abdulkadiroğlu, A. and T. Sönmez (1998), “Random serial dictatorship and the core from random endowments in house allocation problems.” *Econometrica*, 66 (3), 689–701. [1267]
- Abdulkadiroğlu, A. and T. Sönmez (2003), “School choice: A mechanism design approach.” *American Economic Review*, 93 (3), 729–747. [1236, 1265]
- Abizada, A. and S. Chen (2011), “The college admission problem with entrance criterion.” Unpublished manuscript. [1237]
- Agarwal, N. (2015), “An empirical model of the medical match.” *The American Economic Review*, 105 (7), 1939–1978. [1234, 1236]
- Agarwal, N. and P. Somaini (2018), “Demand analysis using strategic reports: An application to a school choice mechanism.” *Econometrica*, 86 (2), 391–444. [1234, 1235, 1238, 1239, 1244, 1246, 1248, 1271]
- Akyol, P. and K. Krishna (2017), “Preferences, selection, and value added: A structural approach.” *European Economic Review*, 91, 89–117. [1237]
- Alstadsæter, A. (2011), “Measuring the consumption value of higher education.” *CESifo Economic Studies*, 57 (3), 458–479. [1239]
- Arcidiacono, P. (2005), “Affirmative action in higher education: How do admission and financial aid rules affect future earnings?” *Econometrica*, 73 (5), 1477–1524. [1235]
- Avery, C., S. Lee, and A. E. Roth (2014), “College admissions as non-price competition: The case of South Korea.” WP 20774, National Bureau of Economic Research. [1237]
- Aygün, O. and I. Bo (2017), “College admission with multidimensional privileges: The Brazilian affirmative action case.” SSRN Archive. [1237]
- Azevedo, E. M. and J. D. Leshno (2016), “A supply and demand framework for two-sided matching markets.” *Journal of Political Economy*, 124 (5), 1235–1268. [1234, 1239, 1240, 1242, 1246]
- Balinski, M. and T. Sönmez (1999), “A tale of two mechanisms: Student placement.” *Journal of Economic Theory*, 84, 73–94. [1234, 1236, 1265]
- Blundell, R. and J. Powell (2003), “Endogeneity in non-parametric and semi-parametric regression models.” In *Advances in Economics and Econometrics: Theory and Applications*, Vol. II (M. Dewatripont, L. P. Hansen, and S. J. Turnovsky, eds.), 312–357, Cambridge University Press, Cambridge, U.K. [1251]
- Budish, E. and E. Cantillon (2012), “The multi-unit assignment problem: Theory and evidence from course allocation at Harvard.” *American Economic Review*, 102 (5), 2237–2271. [1234]
- Calsamiglia, C., C. Fu, and M. Güell (2018), “Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives.” WP 24588, National Bureau of Economic Research. [1234, 1236]

- Calsamiglia, C., G. Haeringer, and F. Klijn (2010), “Constrained school choice: An experimental study.” *The American Economic Review*, 1860–1874. [1234]
- Campbell, J. Y. (1987), “Does saving anticipate declining labor income?” *Econometrica*, 55, 1249–1273. [1249]
- Chade, H., G. Lewis, and L. Smith (2014), “Student portfolios and the college admissions problem.” *The Review of Economic Studies*, 81 (3), 971–1002. [1234, 1237]
- Che, Y. K. and Y. Koh (2016), “Decentralized college admissions.” *Journal of Political Economy*, 124, 1295–1338. [1234, 1237]
- Chen, Y. and O. Kesten (2017), “Chinese college admissions and school choice reforms: A theoretical analysis.” *Journal of Political Economy*, 125 (1), 99–139. [1234, 1237]
- Coles, P., A. Kushnir, and M. Niederle (2013), “Preference signaling in matching markets.” *American Economic Journal: Microeconomics*, 5 (2), 99–134. [1270]
- Dogan, M. K. and T. Yuret (2013), “Publication performance and student quality of Turkish economics departments/Türkiye’de İktisat Bölümlerinin Yayın Performansı ve Öğrenci kalitesi.” *Sosyoekonomi*, 1, 71. [1237]
- Dubey, P., O. Haimanko, and A. Zapechelnyuk (2006), “Strategic complements and substitutes and potential games.” *Games and Economic Behavior*, 54, 77–94. [1244, 1271, 1272]
- Epple, D., R. Romano, and H. Sieg (2006), “Admission, tuition, and financial aid policies in the market for higher education.” *Econometrica*, 74 (4), 885–928. [1235]
- Fack, G., J. Grenet, and Y. He (2017), “Estimating preferences in school choice mechanisms.” Unpublished manuscript. [1236]
- Fu, C. (2014), “Equilibrium tuition, applications, admissions, and enrollment in the college market.” *Journal of Political Economy*, 122 (2), 225–281. [1237]
- Hafalir, I. E., R. Hakimov, D. Kübler, and M. Kurino (2018), “College admissions with entrance exams: Centralized versus decentralized.” *Journal of Economic Theory*, 176, 886–934. [1234, 1235, 1237]
- Hastings, J., T. J. Kane, and D. O. Staiger (2009), “Heterogenous preferences and the efficacy of public school choice.” Working paper, Yale University. [1236]
- He, Y. (2017), “Gaming the school choice mechanism in Beijing.” WP 15–607, revised, Toulouse School of Economics. [1234, 1236, 1241, 1271]
- He, Y. and T. Magnac (2018), “A pigouvian approach to congestion in matching markets.” WP 17–870, revised, Toulouse School of Economics. [1236, 1257]
- Jacob, B., B. McCall, and K. Stange (2012), “The consumption value of education: Implications for the postsecondary market.” Working paper. [1239]
- Jensen, M. K. (2010), “Aggregative games and best-reply potentials.” *Economic Theory*, 43, 45–66. [1242]

- Lee, R. S. and M. Schwarz (2017), “Interviewing in two-sided matching markets.” *The RAND Journal of Economics*, 48 (3), 835–855. [1237]
- Machado, C. and C. Szerman (2017), “Centralized admission and the student-college match.” SSRN Archive. [1237, 1270]
- Manski, C. (1993), “Adolescent econometricians: How do youths infer the returns to schooling?” In *Studies of Supply and Demand in Higher Education* (C. T. Clotfelter and M. Rothschild, eds.). University of Chicago Press, Chicago. [1241]
- Manski, C. F. (1988), “Identification of binary response models.” *Journal of the American Statistical Association*, 83 (403), 729–738. [1246, 1249]
- Matzkin, R. L. (1992), “Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models.” *Econometrica*, 60 (2), 239–270. [1238]
- Matzkin, R. L. (1993), “Nonparametric identification and estimation of polychotomous choice models.” *Journal of Econometrics*, 58 (1), 137–168. [1235, 1238, 1246, 1247]
- Pathak, P. A. and T. Sönmez (2013), “Leveling the playing field: Sincere and sophisticated players in the Boston mechanism.” *American Economic Review*, 98 (4), 1636–1652. [1234]
- Roth, A. E. (2008), “Deferred acceptance algorithms: History, theory, practice, and open questions.” *International Journal of Game Theory*, 36, 537–569. [1234]
- Roth, A. E. (2018), “Marketplaces, markets and market design.” *American Economic Review*, 108 (7), 1609–1658. [1235]
- Roth, A. E. and M. A. O. Sotomayor (1992), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monographs, Vol. 18. Cambridge University Press, Cambridge. [1233]
- Selim, T. and S. Salem (2009), “Student placement in Egyptian colleges.” MPRA Paper No. 17596. [1237]
- Sönmez, T. and M. U. Ünver (2011), “Matching, allocation, and exchange of discrete resources.” In *Handbook of Social Economics*, Vol. 1, 781–852. North-Holland, Amsterdam. [1236]
- Wu, B. and X. Zhong (2014), “Matching mechanisms and matching quality: Evidence from a top university in China.” *Games and Economic Behavior*, 84, 196–215. [1269]
- Zhu, M. (2014), “College admissions in China: A mechanism design perspective.” *China Economic Review*, 30, 618–631. [1237]

Co-editor Peter Arcidiacono handled this manuscript.

Manuscript received 18 August, 2017; final version accepted 21 September, 2018; available online 21 November, 2018.