Equilibrium Tuition, Applications, Admissions and Enrollment in the College Market[*]

Chao Fu[†]

December 7, 2013

Abstract

I develop and estimate a structural equilibrium model of the college market. Students, having heterogeneous abilities and preferences, make college application decisions, subject to uncertainty and application costs. Colleges, observing only noisy measures of student ability, choose tuition and admissions policies to compete for more able students. Tuition, applications, admissions and enrollment are joint outcomes from a subgame perfect Nash equilibrium. I estimate the structural parameters of the model using data from the National Longitudinal Survey of Youth 1997, via a three-step procedure to deal with potential multiple equilibria. In counterfactual experiments, I use the model first to examine the extent to which college enrollment can be increased by expanding the supply of colleges, and then to assess the importance of various measures of student ability.

Keywords: College market, tuition, applications, admissions, enrollment, discrete choice, market equilibrium, multiple equilibria, estimation

[†]Department of Economics, University of Wisconsin, Madison, WI 53706, USA. Email: cfu@ssc.wisc.edu.

1   Introduction

Both the level of college enrollment and the composition of college students continue to be issues of widespread scholarly interest as well as the source of much public policy debate. In this paper, I develop and structurally estimate an equilibrium model of the college market. It provides insights into the determination of the population of college enrollees and permits quantitative evaluation of the effects of counterfactual changes in the features of the college market. The model interprets the allocation of students in the college market as an equilibrium outcome of a decentralized matching problem involving the entire population of colleges and potential applicants. As a result, counterfactuals that directly involve only a subset of the college or student population can produce equilibrium effects for all market participants. My paper thus provides a mechanism for assessing the market equilibrium consequences of changes in government policies on higher education.

While the idea of modeling college matching as a market equilibrium problem is not new, this paper makes advances relative to the current literature by simultaneously modeling three aspects of the college market that are plausibly regarded as empirically important and incorporating them into the empirical analysis. The three aspects are: 1) Application is costly to the student. Besides application fees, a student has to spend time and effort gathering and processing information and preparing application materials. Moreover, she also incurs nontrivial psychic costs such as the anxiety felt while waiting for admissions results. 2) Students differ in their abilities and preferences for colleges.[1] 3) While trying to attract and select more able students, colleges can only observe noisy measures of student ability, such as student test scores and essays. As a result, both sides of the market face uncertainties: for the student, admissions are uncertain, which, together with the cost of application, leads to a non-trivial portfolio problem: how many and which, if any, colleges to apply to? For the college, the yield of each admission and the quality of a potential enrollee are both uncertain. Colleges have to account for students' strategies in order to make inference about student quality. Colleges' policies are also interdependent because students' application portfolios and their enrollment depend on the policies of all colleges.

I model three stages of the market. First, colleges simultaneously announce their tuition. Second, students make application decisions and colleges simultaneously choose their admissions policies. Third, students make their enrollment decisions. My model incorporates tuition, applications, admissions and enrollment as joint outcomes from a subgame perfect Nash equilibrium (SPNE). SPNE in this model need not be unique. Multiplicity may arise from two sources: 1) multiple common self-fulfilling expectations held by the student about admissions policies, and 2) the strategic interplay among colleges.[2]

Building on Moro (2003), I estimate the model in three steps. The first two steps recover all the

---

[1] Throughout the paper, a student's ability refers to her readiness for college, not her innate ability.

[2] Models with multiple equilibria do not have a unique reduced form and this indeterminacy poses practical estimation problems. In direct maximum likelihood estimation of such models, one should maximize the likelihood not only with respect to the structural parameters but also with respect to the types of equilibria that may have generated the data. The latter is a very complicated task and can make the estimation infeasible.

structural parameters involved in the application-admission subgame without having to impose any equilibrium selection rule. In particular, each application-admission equilibrium can be uniquely summarized by the set of probabilities of admission to each college for different types of students. The first step, using simulated maximum likelihood, treats these probabilities as parameters and estimates them along with fundamental student-side parameters in the student decision model, thereby identifying the equilibrium that generated the data. The second step, based on a simulated minimum distance estimation procedure, recovers the college-side parameters by imposing each college's optimal admissions policy. Step three recovers the remaining parameters by matching colleges' optimal tuition with observed tuition levels.

To implement the empirical analysis, I use data from the National Longitudinal Survey of Youth 1997 (NLSY97), which provides detailed information on student applications, admissions, financial aid and enrollment. Some of my major findings are as follows: first, students not only attach different values to the same college, but also rank various colleges and the non-college option differently. That is, there is not a single best college for all, nor is attending college better than the non-college option for all. My first counterfactual experiment finds that increasing the supply of colleges has very limited effect on college attendance. In particular, when non-elite public colleges are expanded, at most 3.6% more students can be drawn into colleges, although the enlarged colleges adopt an open admissions policy and lower their tuition to almost zero. Therefore, neither tuition cost nor the number of available slots is a major obstacle to college access. A large group of students, mainly low-ability students, prefer the outside option over any of the college options.

Second, there are significant amounts of noise in various types of ability measures, including test scores and subjective measures such as student essays. My second counterfactual experiment assesses the importance of subjective measures by eliminating them from the admissions process. In response, elite colleges draw on higher tuition to help screen students. Non-elite colleges lower their tuition to compete for high-ability students, who apply to non-elite colleges as insurance in case they were mistakenly rejected by elite colleges. In equilibrium, enrollee ability drops in elite colleges and increases in non-elite colleges. Overall student welfare decreases; and the only winners are low-ability students, who become harder to distinguish from higher-ability students.

Although this paper is the first to estimate a market equilibrium model that incorporates tuition setting, applications, admissions and enrollment, it builds on various studies on similar topics. For example, Manski and Wise (1983) use a non-structural approach to study each stage of the college admissions problem in isolation. Most relevant to this paper, they find that applicants do not necessarily prefer the highest quality school.[3] Arcidiacono (2005) estimates a structural model to address the effects of college admissions and financial aid rules on future earnings. In a dynamic framework, he models student's application, enrollment and choice of college major and links education decisions to future earnings.

---

[3]Some examples of papers that focus on the role of race in college admissions include Bowen and Bok (1998), Kane (1998) and Light and Strayer (2002).

While an extensive empirical literature focuses on student decisions, little research has examined the college market in an equilibrium framework. One exception is Epple, Romano and Sieg (2006), ERS hereafter. In their paper, students differ in family income and ability (perfectly measured by SAT) and make a single enrollment decision.[4] Given its endowment and gross tuition level, each private college group chooses its financial aid and admissions policies to maximize the quality of education provided to its students.[5] Their model provides an equilibrium characterization of private colleges' financial aid and admissions strategies, where colleges with higher endowments enjoy greater market power and provide higher-quality education. With complete information, no uncertainty and no unobserved heterogeneity, their model predicts that students with the same SAT and family income would have the same admission, financial aid and enrollment outcomes. The authors assume measurement errors in SAT and family income, which are found to be large in order to accommodate data variations.[6]

This paper departs from ERS in several respects: 1) The college market is subject to information frictions and uncertainty: colleges can only observe noisy measures of student ability, and they do not observe student preferences. As a result, colleges are faced with complex inference problems in making their admissions decisions. Meanwhile, application becomes a non-trivial problem for the student, as is manifested by the popularity of various application guide programs. Both colleges and students will adjust their behavior according to how much information is available on the market. Consequently, evaluating the severity of information frictions is important for predicting the equilibrium effects of various counterfactual education policies. 2) Student application decisions differ substantially. For example, over 50% of high school graduates do not apply to any college. However, the college market includes not only college enrollees and/or those who do apply, but all potential college applicants. Alternative education policies will affect not only where applicants are enrolled, but also who will apply in the first place. Therefore, to evaluate the effects of these policies, it is necessary to understand the application decisions (including non application) made by all students and how these decisions interact with colleges' decisions. 3) Given the important role of public colleges, which accommodate the majority of college students, this paper models the strategic behavior of both public colleges and private colleges. 4) Students have different abilities and preferences for colleges, which are unobservable to researchers. Arguably, such heterogeneity may be the key force underlying data variations unexplained by observables. Hence it is important to incorporate them in the model. As the first two structural papers that study college market equilibrium, ERS and this paper complement one another. ERS provides a more comprehensive view on private colleges' financial aid strategy, which is especially important in explaining the

---

[4]In their paper, the application decision is not modeled. It is implicitly assumed that either application is not necessary for admission, or all students apply to all colleges or at least their best two equilibrium alternatives. Accordingly, their empirical analysis is based on a sample of college enrollees.

[5]Focusing on private colleges, they treat public colleges as an exogenous outside option for students.

[6]The authors note that "the model may not capture some important aspects of admission and pricing." (page 911)

allocation of elite students. This paper aims at understanding the allocation of typical students by endogenizing student application as part of the equilibrium in a frictional market, where both public colleges and private colleges act strategically.

Theoretically, I build on the work by Chade, Lewis and Smith (2011), CLS hereafter, who model the decentralized matching of students and two colleges. Students, with heterogeneous abilities, make application decisions subject to application costs and noisy evaluations. Colleges compete for better students by setting admissions standards for student signals.[7] As part of its contribution, my paper quantifies the significance of the two key elements of CLS: information frictions and application costs. Moreover, I extend CLS to account for some elements that are important, as acknowledged by the authors, to understand the real-world problem. On the student side, first, students are heterogeneous in their preferences for colleges as well as in their abilities, both of which are unknown to the colleges. Second, I allow for two noisy measures of student ability. One measure, as the signal in CLS, is subjective and its assessment is known only to the college. A typical example of this type of measure is the student essay. The other measure is the objective test score, which is known both to the student and to the colleges she applies to, and may be used strategically by the student in her applications.[8] On the college side, I model multiple colleges, which compete against each other via tuition as well as admissions policies.[9]

This paper is also related to studies on the estimation of models with multiple equilibria.[10] In discrete games studied in the IO literature, it is usually assumed that the researcher can observe data from different markets. In games with complete information, it is usually assumed that different markets are potentially in different equilibria. In Bresnahan and Reiss (1990), for a given value of the exogenous variables, the model predicts a unique number of entrants, which enables one to estimate and identify the parameters using maximum likelihood or method of moments. Tamer (2003) shows that if there are some values of a covariate for which the actions of all but one player are dictated by dominant strategies, then the problem boils down to discrete choice by this special single agent. In a more general setting (e.g., Ciliberto and Tamer (2009)), multiple equilibria exist with respect to the number of entrants and the support of the covariates is not rich enough, inference relies on partial identification and estimation is done via exploring the bounds on choice probabilities.

In discrete games with incomplete information, several studies (e.g., Aguirregabiria and Mira (2007) and Bajari, Benkard and Levin (2007) in dynamic games and Bajari, Hong, Krainer and Nekipelov (2010) in static games) use a two-step estimation procedure, assuming that the researcher observes multiple games/markets and that the same equilibrium is played across games. The first step estimates the conditional choice probabilities. The second step estimates the parameters that

---

[7]Nagypál (2004) analyzes a model in which colleges know student types, but students themselves can only learn their type through normally distributed signals.

[8]For example, a low-ability student with a high SAT score may apply to top colleges to which she would not otherwise apply; a high-ability student with a low SAT score may apply less aggressively than she would otherwise.

[9]As a price of these extensions, it is infeasible to obtain an analytical or graphical characterization of the equilibrium as in CLS.

[10]See de Paula (2013) for a comprehensive survey.

enter the payoff function by solving each player's decision problem given their equilibrium beliefs estimated from the first step. The assumption of a single equilibrium in the data is crucial for identification as it guarantees (joint with other restrictions) that the probability of one player choosing a specific action and his expected payoff from this choice using equilibrium beliefs are in a one-to-one relationship.

Moro (2003) develops an estimation strategy that applies to a different yet also very common framework: the data are from only one market, one side of the market consists of many small players and all observations derive from the same equilibrium. More importantly, each equilibrium can be uniquely summarized by an *unobserved* equilibrium object that can be treated as a *parameter*. He shows that under certain conditions one can consistently estimate *both* the fundamental parameters *and* the equilibrium that generated the data in two steps. As will be shown, my model setup falls into this framework.

The rest of the paper is organized as follows: Section 2 lays out the model. Section 3 explains the estimation strategy, followed by discussions about identification. Section 4 describes the data. Section 5 presents empirical results, including parameter estimates and model fit. Section 6 describes the counterfactual experiments. The last section concludes. The appendix contains some details and additional tables.

2 Model

2.1 Primitives

2.1.1 Players

There is a continuum of students, making college application and enrollment decisions. Students come from different family backgrounds (B), of which the student's home state (denoted l for location) is one element. They also differ in their abilities (one measure of which is SAT) and preferences for colleges.[11] There are J four-year colleges, indexed by $j = 1, 2, ...J$. Each college consists of a tuition office and an admissions office, and is endowed with a fixed capacity κ_j, where $\kappa_j > 0$ and $\sum_{j=1}^{J} \kappa_j < 1$, the total measure of students. There is also a two-year community college indexed by $j = J + 1$, which any student can attend without application. This paper focuses on the strategic behavior of four-year colleges; the community college will be treated as an exogenous option.

Assumptions Theoretically speaking, one can treat each college in real life as one player without much complication, however, it is infeasible to do so empirically for sample size and computation

---

[11]SAT can be low(1), medium(2) or high(3).

reasons.[12] I have made the following assumptions:

A1. There are 4 groups (g) of 4-yr colleges: (private, elite), (public, elite), (private, non-elite) and (public, non-elite). Colleges within a group are, for an average student, identical except for their locations. Denote g_j as the group College j belongs to.

A2. From a student's point of view, the location of a college matters only up to whether or not it is within her home state.

A3. All colleges face the same distribution of students. Given such symmetry, I focus on symmetric equilibrium, in which each college makes its own decision yet no college would benefit from deviating to a strategy that is different from the one used by others in its group.

With these assumptions, the model focuses on the main features of the college market. On the college side, it captures the fact that colleges with similar characteristics (within a group) are closer substitutes for one another than for those in other groups. As a result, the within-group competition is more fierce than that across groups. Moreover, it also captures the fact that the admissions policies and the tuition policies are similar among similar colleges. On the student side, it captures factors that are presumably the major ones considered by students: tuition cost, whether the college is private or public, elite or non-elite, in or out of one's home state.

Some other aspects, however, are abstracted. For example, although this paper can capture the most important aspect of students' geographical preferences, i.e., attachment to their home states, A2 treats all non-home states equally: there is no systematic reason that a student will prefer one over another. Similarly, although the model captures colleges' differential treatments of in-state versus out-of-state students, A3 abstracts from college strategies that depend on which specific states they are located in.[13]

2.1.2 Application Cost

Application is costly to the student. The cost of application is a non-decreasing function $C(\cdot)$ of the number of applications sent.

2.1.3 Financial Aid

A student may obtain financial aid that helps to fund her college education in general, and she may also obtain college-specific financial aid. The amounts of various financial aid depend on the student's family background and SAT, via financial aid functions $f_j(B, SAT)$, for $j = 0, ..., J + 1$, with 0 denoting the general aid and $j = 1, ..., J + 1$ college-specific aid.[14] In reality, although

---

[12]Epple, Romano and Sieg (2006) aggregate private colleges into 6 groups and treat each group as one college.

[13]Without A3, college strategies will vary with the distribution of students within their own states, even if colleges are identical. Theoretically, it is feasible to incorporate this aspect. Empirically, it is not because first, the number of students (applicants) observed per state is small (even smaller); deriving the distribution of students for each state from the sample may be problematic. Second, allowing college strategies to differ across states will significantly increase the dimension of the problem, making the computation and estimation infeasible.

[14]Ideally, a more complete model would endogenize tuition, applications, admissions, enrollment and financial aid. Unfortunately, this involves great complications that will make the empirical analysis intractable. As a compromise,

guidelines are available for students to calculate the expected financial aid they might obtain, the exact amounts remain uncertain. To capture this uncertainty, I allow the final realizations to be subject to post-application shocks $\eta \in \mathbb{R}^{J+2}$, distributed i.i.d. $N(0, \Omega_\eta)$. The realized financial aid for student i is given by

$$f_{ji} = \max\{f_j(B_i, SAT_i) + \eta_{ji}, 0\} \text{ for } j = 0, 1, .., J+1.$$

2.1.4 Student Endowment

By college age, each student is endowed with certain ability and preferences for colleges that are unobservable to the researcher. Abilities and preferences are potentially correlated. They are modeled as follows: students are of different types (K); and those within a type may share similar preferences more than they do with students of other types. These unobservable types are correlated with SAT and family background, and are distributed according to $P(K|SAT, B)$.

A student type K has two dimensions with $K \equiv (A, z)$. The first dimension (A) represents the quality of a student that colleges care about, i.e., student ability, which can be low (1), medium (2) or high (3). The second dimension $z \in \{1, 2\}$ allows for systematic heterogeneity in preferences among students of the same ability. For example, some students may prefer big (public) universities that offer greater diversity and a wider range of student activities; while some may prefer small (private) colleges where they can get more personal attention from professors.

In addition, each student may have her own idiosyncratic tastes for colleges that are not representative of her type. For example, a student may prefer a particular college because her parents attended that college. To capture such heterogeneity, a type-K student i's preferences for colleges are modeled as a random vector $u_i \equiv \{u_{ji}\}_{j=1}^{J+1}$, with

$$u_{ji} = \overline{u}_{g_j K} + \epsilon_{1 g_j i} + \epsilon_{2ji},$$

where g_j represents the group College j belongs to. $\overline{u}_{g_j K}$ is the preference for college group g_j for an average type-K student.[15] $\epsilon_{1 g_j i} \sim N(0, \sigma_{\epsilon_{1 g_j}}^2)$ is student i's idiosyncratic taste for group g_j. $\epsilon_{2ji} \sim N(0, \sigma_{\epsilon_2}^2)$ captures her personal taste for college j regardless of its group.[16] Hence, a student's tastes for colleges are correlated within a college group.

Finally, students also differ in their (dis)tastes for studying out of their home states, modeled as $\xi_i \sim N\left(\overline{\xi}_K, \sigma_\xi^2\right)$. Given tuition profile $t \equiv \{\{t_{jl}\}_l\}_j$, the ex-post value of attending college j for

---

Epple, Romano and Sieg (2006) abstract from application decisions and hence the effects of college policies on the pool of applicants, so that they can better focus on college's financial aid strategies. I carry out my analysis in a way that complements their work: I endogenize application decisions and allow colleges to choose gross tuition while leaving financial aid exogenous.

[15] Treating each $\overline{u}_{g_j K}$ as one parameter, the model allows student type-specific preferences to be correlated across various college groups in a non-parametric fashion. Students' observable characteristics (SAT, B) are correlated indirectly with their preferences via their correlation with student type K.

[16] All community colleges are treated as one single option and $\epsilon_{2ji} = 0$ for $j = J+1$.

student i is

$$U_{ji}(t) = (-t_{jl_i} + f_{0i} + f_{ji}) + u_{ji} - I(l_j \neq l_i)\xi_i, \tag{1}$$

where t_{jl_i} is college j's tuition for a student from state l_i, hence the first parenthesis of (1) summarizes student i's net monetary cost to attend college j. The last term specifies that if the student's home location differs from college j's location l_j, (dis)utility ξ_i applies.[17]

In addition, an outside non-college option is always available to the student and its value is normalized to zero. Thus, students' preferences for colleges are relative to their individual preferences for the non-college option, all of which are endowed on them by college age.[18]

2.1.5 College Payoff

Colleges care about the ability of their enrollees and their net tuition revenues. For a private college j, its payoff (W_j) is

$$W_j = \int (\omega_{a_i} + m_{1j}\pi_{ji}) \, dF_j^*(i) + m_{2j}\frac{\Pi_j^2}{N_j} \text{ if } j \text{ is private.} \tag{2}$$

ω_a is the value of ability $A = a$, with $\omega_{a+1} > \omega_a > 0$, $\pi_{ji} \equiv t_j - f_{ji}$ is the net tuition revenue from student i, and m_{1j} measures college j's valuation of net tuition relative to student ability.[19] Each student's contribution is aggregated over $F_j^*(i)$, the endogenous distribution of college j's enrollees. The second term in (2) captures college j's potentially nonlinear preference for revenue, where Π_j is j's total net tuition revenue and N_j is its total enrollment. Π_j^2 is adjusted by N_j to keep the second term at the same magnitude as the first term.

A public college may treat in-state students differently from out-of-state students, with an objective function

$$W_j = \sum_{\iota=0}^{1} \left(\int (\omega_{a_i} + m_{1j\iota}\pi_{ji}) \, dF_{j\iota}^*(i) + m_{2j\iota}\frac{\Pi_{j\iota}^2}{N_{j\iota}} \right) \text{ if } j \text{ is public,} \tag{3}$$

$$\iota \equiv I(l_i = l_j).$$

$F_{j0}^*(i)$ $(F_{j1}^*(i))$ is the endogenous distribution of out (in) state enrollees in college j; Π_{j0} (Π_{j1}) is j's total net tuition revenue from out (in) state enrollees, and N_{j0} (N_{j1}) is the total number of out (in) state enrollees. For example, since public colleges are partly state-funded, they may be much more constrained from collecting high tuition from in-state students than from out-of-state students; it is possible that m will differ across ι's.

---

[17] Community colleges are always in state.

[18] In this paper, students' ability and preferences are taken as initial conditions. For research on early childhood human capital formation, see, for example, Cunha, Heckman and Schennach (2010).

[19] Given symmetry, the tuition weights m's are restricted to be the same within a college group.

Discussions It is common in the literature to assume that two factors are key to colleges' objectives: first, the quality of enrollees; and second, monetary inputs that fund faculty and facilities. I assume that a college's payoff depends on these two factors, which is in line with previous studies on college behavior. For example, although specific forms differ across studies, both Rothschild and White (1995) and Epple, Romano and Sieg (2006) assume that education production depends on student ability and monetary inputs, the value of the latter being equal to net tuition in equilibrium.[20] Another accepted fact in the literature is that the vast majority of colleges do not aim at profit maximization.[21] In other words, their preferences for revenue may be bounded.[22] Without a deeper study on why such preferences may exist, which is beyond the scope of this paper, I allow for, without imposing, such preferences. I assume a quadratic specification because it is parsimonious while flexible enough to entertain preferences that are concave, convex or linear, to be determined by the data.

The model captures critical aspects of the college market that distinguish it from oligopoly markets studied in the IO literature, e.g., Berry, Levinsohn and Pakes (1995, 2004). Most importantly, unlike a typical firm, which does not care about the identity of its customers, a college values but does not observe the ability of its applicant. Faced with information frictions and a capacity constraint, a college has to solve a non-trivial inference problem as it tries to fill its capacity with higher-ability students. Given these complications, I have assumed away college-side unobservables that affect college payoffs in order to keep the exercise feasible. Yet, the model allows students to differ in their preferences for any given college, and colleges to differ in their tradeoffs between student ability and revenue.[23] Adding college-level unobservables will bring the model closer to reality. However, it involves nontrivial technical problems and is left for future research.[24]

2.1.6 Timing

Stage 1: Colleges simultaneously announce tuition levels, to which they commit.

Stage 2: Students make application decisions; colleges simultaneously choose admissions policies.

Stage 3: Students learn about admission and financial aid results, and make enrollment decisions.[25]

---

[20]In Epple, Romano and Sieg (2006), there is a third factor: the average family income among enrollees, which negatively affects a college's objective.

[21]For example, Winston (1999) emphasizes that higher education is a nonprofit enterprise. Chade, Lewis and Smith (2011) assume colleges maximize total enrollee ability. Howell (2010) assumes that colleges maximize their reputations that depend on enrollee characteristics. There are also studies that treat colleges as profit maximizers, for example, Rothschild and White (1995), who show that equilibrium prices in a competitive market with perfect information achieve efficient allocation of students.

[22]Epple, Romano and Sieg (2006) assume that the maximum prices colleges can charge, i.e., their tuition levels, are exogenously given.

[23]On the student side, besides individual idiosyncratic tastes, preference parameters are (student type, college group)-specific. On the college side, college objectives and constrains differ across college groups.

[24]Besides the increase in computational burden, with college-side unobservables, one will have to solve the applications-admissions game and deal with the multiple equilibria problem during the estimation. Details are available upon request.

[25]This paper excludes early admissions, which is a very interesting and important game among top colleges. See, for example, Avery, Fairbanks and Zeckhauser (2003), and Avery and Levin (2010). For college applications in general,

2.1.7 Information Structure

Upon student i's application, each college she applies to receives a signal $s \in \{1, 2, 3\}$ (low, medium, high) drawn from the distribution $P(s|A_i)$, the realization of which is known only to the college. For $A < A'$, $P(s|A')$ first order stochastically dominates $P(s|A)$.[26] Unconditionally, a student's signals to various colleges are correlated because they all measure the student's ability. Conditional on the student's ability, the residuals embedded in these signals are assumed to be i.i.d. random. Such randomness is meant to capture the idiosyncratic interpretations of the student's application materials by different admission officers across colleges.

$P(s|A)$, the distributions of characteristics, preferences, payoff functions and financial aid functions are public information. An individual student's SAT_i score is known both to her and to the colleges she applies to. A student has private information about her type K_i, taste ϵ_i and family background B_i ($l_i \in B_i$). To ease notation, let $X_i \equiv (K_i, B_i, \epsilon_i)$. After application, the student observes her financial aid shocks. The following table summarizes, in addition to the public information, information available to the student and admissions office j when they make decisions. The last column also shows student characteristics that are observable to the researcher.

Information Sets

	Student	Admissions Office j	Researcher
Application-Admission	SAT_i, X_i	$SAT_i, s_{ji}, (l_i)$	SAT_i, B_i
Enrollment	SAT_i, X_i, η_i	–	SAT_i, B_i

For any individual applicant, admissions office j observes her SAT_i and the signal (s_{ji}) she sends to j. If the admissions office can discriminate based on students' origins, it also observes l_i. For the student, the admission probability depends on her SAT and ability (instead of signal because she cannot observe her signal but her ability governs her signal distribution), and in the case of origin-based discrimination, it also depends on her home location. I do not make assumptions about whether colleges practice origin-based discrimination in their admissions; instead, the estimation procedure outlined later allows me to infer it from the data. To ease notation, I will present the model without such discrimination in admissions.[27]

2.2 Applications, Admissions and Enrollment

In this subsection, I solve the student's problem backwards and the college's admissions problem, taking as given the tuition levels announced in Stage 1 of the game.

---

however, early admissions account for only a small fraction of the total applications. For example, in 2003, 17.7% of all four-year colleges offered early decision. In these colleges, the mean percentage of all applications received through early decision was 7.6%. *Admission Trends Survey* (2004), National Association for College Admission Counseling. For similar reasons, this paper abstracts from post-admission negotiations that may be important in top private colleges.

[26]That is, if $A < A'$, then for any $s \in \{1, 2, 3\}$, $\Pr(s' \leq s|A) \geq \Pr(s' \leq s|A')$.

[27]All derivation goes through for the model with such discrimination: one only needs to add l_i into the arguments of admissions probability faced by students and admissions policies set by colleges.

2.2.1 Enrollment Decision

Given her admission and financial aid results, student i chooses the best among her outside option and admissions on hand, i.e., $\max\{U_{0i}, \{U_{ji}(t)\}_{j\in O_i}\}$, where O_i denotes the set of colleges that have admitted student i, which always includes the community college. Let

$$v(O_i, X_i, \eta_i|t) \equiv \max\{U_{0i}, \{U_{ji}(t)\}_{j\in O_i}\} \tag{4}$$

be the optimal ex-post value for student i, given admission set O_i; and denote the associated optimal enrollment strategy as $d(O_i, X_i, \eta_i|t)$.

2.2.2 Application Decision

Given her admissions probability $p_j(A_i, SAT_i|t)$ to each college j, the value of application portfolio Y for student i is

$$V(Y, X_i, SAT_i|t) \equiv \sum_{O\subseteq\{Y,J+1\}} \Pr(O|A_i, SAT_i, t)E\left[v(O, X_i, \eta_i|t)\right] - C(|Y|), \tag{5}$$

where the expectation is over financial aid shocks, $|Y|$ is the size of portfolio Y, and

$$\Pr(O|A_i, SAT_i, t) = \prod_{j\in O} p_j(A_i, SAT_i|t) \prod_{j'\in Y\backslash O} (1 - p_{j'}(A_i, SAT_i|t))$$

is the probability that the set O of colleges admit student i. The student's application problem is

$$\max_{Y\subseteq\{1,\ldots,J\}} \{V(Y, X_i, SAT_i|t)\}. \tag{6}$$

Let the optimal application strategy be $Y(X_i, SAT_i|t)$.

2.2.3 Admissions Policy

Given tuition announced by all colleges, admissions office j chooses its policy subject to its capacity constraint. Observing only (s, SAT) of its applicants, the office treats everyone with the same (s, SAT) equally with policy $e_j(s, SAT|t)$. Its optimal admissions policy must be a best response to other colleges' admissions policies while accounting for students' strategic behavior. In particular, from (s, SAT), the college has to infer, first, the probability that a certain applicant will accept its admission, and second, the expected ability of this applicant *conditional on her acceptance of the admission*, both of which depend on the strategies of all other players.[28] For example, whether or not a student will accept college j's admission depends on whether she also applies to other colleges

---

[28]Conditioning on acceptance is necessary to make a correct inference about the student's ability because of the potential "winner's curse": the student might accept college j's admission because she is of low ability and is rejected by other colleges.

(which is unknown to college j), and if so, whether or not she will be accepted by each of those colleges. In addition, college j needs to integrate out all financial aid shocks that may occur to the student. In the appendix, I provide the formal theoretical derivation and the implementation of $\{e_j(s, SAT|t)\}$.

2.2.4 Probability of Admissions

The probability of admissions for different (A, SAT) groups of students, $\{p_j(A, SAT|t)\}$, summarizes the link among various players. Knowledge of p makes the information about admissions policies $\{e_j(s, SAT|t)\}$ redundant. Students' application decisions are based on p. Likewise, based on p_{-j}, college j can make inferences about its applicants and therefore choose its admissions policy. The relationship between p and e is given by:[29]

$$p_j(A, SAT|t) = \sum_s P(s|A)e_j(s, SAT|t). \tag{7}$$

2.2.5 Application-Admission Equilibrium

Definition 1 *Given tuition profile t, a symmetric application-admission equilibrium, denoted as $AE(t)$, is $(d(\cdot|t), Y(\cdot|t), e(\cdot|t), p(\cdot|t))$, such that*
(a) $d(O, X, \eta|t)$ is an optimal enrollment decision for every (O, X, η);
(b) Given $p(\cdot|t)$, $Y(X, SAT|t)$ is an optimal college application portfolio for every (X, SAT), i.e., solves problem (6);
(c) For every j, given $(d(\cdot|t), Y(\cdot|t), p_{-j}(\cdot|t))$, $e_j(\cdot|t)$ is an optimal admissions policy, and $e_j(\cdot|t) = e_{j'}(\cdot|t)$ if $g_j = g_{j'}$;
(d) p_j and e_j satisfy (7) (consistency).

2.3 Tuition Policy

Before the application season begins, college tuition offices simultaneously announce their tuition policies, understanding that their announcements are binding and will affect the application-admission subgame.[30] Let $E(W_j|AE(t))$ be college j's expected payoff under $AE(t)$. Given t_{-j} and the equilibrium profiles $AE(\cdot)$ in the following subgame, college j's problem is

$$\max_{\tilde{t}_{jl} \geq 0} \{E(W_j|AE(\tilde{t}_j, t_{-j}))\} \tag{8}$$

$$s.t.\ \tilde{t}_{jl} = \tilde{t}_{jl'} \text{ for all } l \text{ and } l' \text{ if } j \text{ is private,}$$

$$\tilde{t}_{jl} = \tilde{t}_{jl'} \text{ for all } l, l' \neq l_j \text{ if } j \text{ is public.}$$

---

[29]The role of p as the link among players and the mapping (7) are of great importance in the estimation strategy to be specified later.

[30]Although from the researcher's point of view the subsequent game could admit multiple equilibria, I assume that the players agree on the equilibrium selection rule.

The constraints specify that tuition must be the same for all attendees in a private college.[31] Public colleges may charge different tuition for in-state than for out-of-state students, but all out-of-state students face the same tuition.[32]

Independent of its preference for revenue, each college considers the strategic role of its tuition in the subsequent $AE(\tilde{t}_j, t_{-j})$. On the one hand, low tuition makes the college more attractive to students and more competitive in the market. On the other hand, high tuition serves as a screening tool and leads to a better pool of applicants if high-ability students are less sensitive to tuition than low-ability students.[33] Together with its preference for revenue, such trade-offs determine the college's optimal tuition level.

2.4 Subgame Perfect Nash Equilibrium

Definition 2 *A symmetric subgame perfect Nash equilibrium for the college market is*
$(t^*, d(\cdot|\cdot), Y(\cdot|\cdot), e(\cdot|\cdot), p(\cdot|\cdot))$ *such that:*
(a) For every t, $(d(\cdot|t), Y(\cdot|t), e(\cdot|t), p(\cdot|t))$ constitutes an $AE(t)$, according to Definition 1;
(b) For every j, given t^*_{-j}, t^*_j is optimal for college j, i.e., solves problem (8), and $t^*_j = t^*_{j'}$ if $g_j = g_{j'}$.

In the appendix, I prove the existence of equilibrium for a simplified version of the model. Numerically, I have found equilibrium in the full model throughout my empirical analyses.

3 Estimation Strategy and Identification

3.1 Estimating the Application-Admission Subgame

The estimation is complicated by potential multiple equilibria in the subgame and the fact that researchers do not observe the equilibrium selection rule.[34] One way to deal with this complication is to impose some equilibrium selection rule assumed to have been used by the players and to consider only the selected equilibrium. However, for models like the one in this paper, there is not a single compelling selection rule (from the researcher's point of view).[35] Building on Moro (2003), I use a two-step strategy to estimate the application-admission subgame without having to impose any equilibrium selection rule.

---

[31] Given students' home bias, private colleges may want to charge higher tuition for in-state students. Without a deeper investigation into why this is not the case, I impose this restriction to reconcile with the data.

[32] For sample size and computational concerns, I abstract from interstate tuition reciprocity practiced in some states.

[33] This is a possible scenario. However, in the estimation, I do not impose any restriction on the relationship between student ability and their sensitivity to prices.

[34] The problem of possible multiple equilibria is a difficult, yet frequent problem in structual equilibrium models. For example, the model by Epple, Romano and Sieg (2006) also admits multiple equilibria, and the authors assume unique equilibrium in their estimation and other empirical analyses.

[35] See, for example, Mailath, Okuno-Fujiwara and Postlewaite (1993), who question the logical foundations and performances of many popular equilibrium selection rules.

Each application-admission equilibrium is uniquely summarized in the admissions probabilities $\{p_j(A, SAT, l)\}$ or $\{p_j(A, SAT)\}$, depending on whether origin-based discrimination is allowed. The vector p provides sufficient information for players to make their unique optimal decisions. In the student decision model, the unobservable tastes of an individual student do not affect the equilibrium; and p is taken as given just like all the other parameters are. Step One treats p as parameters and estimates them along with structural student-side parameters. As shown in the identification section, the student-side model is identified, so is the equilibrium that generated the data.[36] In the second step, one only needs to solve each college's decision problem instead of the game between colleges. The reason is the following: the p of other colleges is exactly what a college was reacting to; and p is a known fixed *parameter* from the first stage estimation. Given model parameters and the p from the first step, the researcher can solve for a college's unique admissions policies $e_j(s, SAT|\cdot)$, which yield a new set of admissions probabilities.[37] Step two uses this logic to search for the college-side parameters that bring these probabilities to match the equilibrium admissions probabilities estimated in Step One.

3.1.1 Step One: Student-Side Parameters and Equilibrium Admissions Probabilities

I implement the first step via simulated maximum likelihood estimation (SMLE): together with estimates of the fundamental student-side parameters $\left(\widehat{\Theta}_0\right)$, the estimated equilibrium admissions probabilities $\widehat{p}$ should maximize the probability of the observed outcomes of applications, admissions, financial aid and enrollment, conditional on observable student characteristics, i.e., $\{(Y_i, O_i, f_i, d_i | SAT_i, B_i)\}_i$. Θ_0 is composed of 1) preference parameters Θ_{0u}, 2) application cost parameters Θ_{0C}, 3) financial aid parameters Θ_{0f}, and 4) the parameters involved in the distribution of types Θ_{0K}.

Suppose student i is of type K. Her contribution to the likelihood, $L_{iK}(\Theta_{0u}, \Theta_{0C}, \Theta_{0f}, p)$, is composed of the following parts:

$L_{iK}^Y(\Theta_{0u}, \Theta_{0C}, \Theta_{0f}, p)-$ the contribution of applications Y_i,

$L_{iK}^O(p)-$ the contribution of admissions $O_i | Y_i$,

$L_{iK}^f(\Theta_{0f})-$ the contribution of financial aid $f_i | O_i$, and

---

[36] Given admissions probabilities, students' application strategies are independent, which yields a unique equilibrium in the student-side problem. This may not hold if students directly value the quality of their peers. With peer effects, multiple equilibria may coexist in both the student-side and the college-side problem, inducing substantial complications into the model. The existence of peer effects has been controversial in the higher-education literature. (See, for example, Sacerdote (2001), Zimmerman (2003), Arcidiacono and Nicholson (2005) and Dale and Krueger (1998)). In this paper, I focus on the interactions between colleges and students and the competition among colleges, leaving the inclusion of interactions among students for future research.

[37] Notice that a college also observes individual students' signals, while the researcher does not. Therefore, the researcher cannot predict the admissions result for each individual student. However, given parameter values, the researcher can predict the distribution of applicants and their signals, which is sufficient to solve for the admissions policies $e_j(s, SAT)$.

$L_{iK}^d(\Theta_{0u}, \Theta_{0f})-$ the contribution of enrollment $d_i|(O_i, f_i)$, such that

$$L_{iK}(\cdot) = L_{iK}^Y(\cdot)L_{iK}^O(\cdot)L_{iK}^f(\cdot)L_{iK}^d(\cdot).$$

Now, I will specify each part in detail. Conditional on (K, SAT_i, B_i), there are no unobservables involved in the probabilities of $O_i|Y_i$ and $f_i|O_i$. The probability of $O_i|Y_i$ depends only on ability, SAT (and l), and is given by

$$L_{iK}^O(p) \equiv \Pr(O_i|Y_i, A, SAT_i, l_i) = \prod_{j \in O_i} p_j(A, SAT_i, l_i) \prod_{j' \in Y_i \setminus O_i} [1 - p_{j'}(A, SAT_i, l_i)].$$

The probability of the observed financial aid $L_{iK}^f(\Theta_{0f})$ depends only on SAT and family background via the financial aid functions.[38]

The choices of Y_i and $d_i|(O_i, f_i)$ both depend on the unobserved idiosyncratic tastes ϵ. Let $J_i^f \subseteq \{0, O_i\}$ be the sources of observed financial aid for student i, where 0 denotes general aid. Let $G(\epsilon, \{\eta_j\}_{j \in \{0, O_i\} \setminus J_i^f})$ be the joint distribution of idiosyncratic taste and shocks to unobserved financial aid,
$L_{iK}^Y(\Theta_{0u}, \Theta_{0C}, \Theta_{0f}, p)L_{iK}^d(\Theta_{0u}, \Theta_{0f}) \equiv$

$$\int I(Y_i|K, SAT_i, B_i, \epsilon)I(d_i|O_i, K, B_i, \epsilon, \{\eta_j\}_{j \in \{0, O_i\} \setminus J_i^f}, \{f_{ji}\}_{j \in J_i^f})dG(\epsilon, \{\eta_j\}_{j \in \{0, O_i\} \setminus J_i^f}).$$

The multi-dimensional integration has no closed-form solution and is approximated by a kernel smoothed frequency simulator (McFadden (1989)).[39]

To obtain the likelihood contribution of student i, I integrate over the unobserved type:

$$L_i(\Theta_0, p) = \sum_K P(K|SAT_i, B_i; \Theta_{0K})L_{iK}(\Theta_{0u}, \Theta_{0C}, \Theta_{0f}, p). \tag{9}$$

Finally, the log likelihood for the entire random sample is

$$\pounds(\Theta_0, p) = \sum_i \ln(L_i(\Theta_0, p)). \tag{10}$$

3.1.2 Test the Existence of Origin-Based Admissions

In Step One, two versions of the student decision model are estimated. In the first version, $p_j(A, SAT, l)$ is allowed to depend on whether or not the student is in-state $I(l_i = l_j)$.[40] In the second version, it is restricted that $p_j(A, SAT, l) = p_j(A, SAT, l')$ for all l, l'. Since the first version

---

[38] I also allow for measurement errors in financial aid.

[39] See the appendix for details.

[40] Version 1 includes sub-versions where $p_j(\cdot)$ is allowed to depend on $I(l_i = l_j)$ for different subsets (including all) of the college groups.

nests the second, one can test whether or not admissions depend on a student's origin via a likelihood ratio test. In my estimation, the likelihood ratio test fails to reject the hypothesis that admissions are origin-independent, which has major implications for the specification and estimation of the college side of the model as follows.[41]

Admissions Office's Information Set The test result is consistent with a specification where a student's origin (l) is not in the admissions office's information set.[42] An observationally equivalent alternative is that l is observed, but the admissions office is constrained to admit comparable students from different states equally. In this paper, I assume the former.

Admissions Office's Objective Consistent with the test result, only ability measures matter for admissions. This can be rationalized by an admissions process that is purely merit based and aimed at maximizing total enrollee ability subject to capacity constraints. Alternatively, net tuition revenue may also be taken into account by the admissions office, although admissions policies do not depend on students' origins. Between these two observationally equivalent specifications, I choose the former because first of all, it is consistent with the need-blind admissions practiced by a lot of colleges, especially the elite ones. Second, it significantly facilitates the estimation. Given that the goal of admissions is the maximization of total enrollee ability, to solve the admissions problem, knowledge about a college's preference for revenue is unnecessary. Thus, to estimate parameters that govern the admissions process, there is no need to jointly estimate colleges' revenue preference parameters: one can estimate the former via solving individual college's admissions decision problem in Step Two, and recover the latter in Step Three.[43]

3.1.3   Step Two: Estimate Admission-Related College-Side Parameters

In Step Two, I use simulated minimum distance estimation (SMDE) to recover college-side parameters Θ_2, including signal distribution $P(s|A)$, capacity constraints κ and values of abilities ω. Based on $\widehat{\Theta}_0$, I simulate a population of students and obtain their optimal application and enrollment strategies under $\widehat{p}$. The resulting equilibrium enrollment in each college group should equal its expected capacity. These equilibrium enrollments, together with $\widehat{p}$, serve as targets to be matched in the second-step estimation.

   The estimation explores each college's optimal admissions policy given the proper information set as tested in Step one. Taking student strategies and $\widehat{p}_{-j}$ as given, college j chooses its admissions policy e_j, which is generically unique and leads to the admissions probability to college j from

---

[41]There are some differences between the observed admissions rates for in-state and out-of-state students with the same SAT, which can be explained by origin-based discrimination in admissions and/or student self selection. The likelihood ratio test fails to reject the hypothesis that student self selection is sufficient to explain such differences.

[42]This includes the case where l is observed but ignored.

[43]Otherwise, one has to solve the college's tuition problem hence the application-admissions equilibrium in order to estimate admission-related parameters.

students' perspectives, according to equation (7). Ideally, the admissions probabilities derived from Step Two should match $\widehat{p}$ from Step One, and the capacity parameters in Step Two should match equilibrium enrollments. The estimates of the college-side parameters minimize the weighted sum of the discrepancies, which arise from the first-step estimation errors. Let $\widehat{\Theta}_1 = [\widehat{\Theta}_0', \widehat{p}']'$; the objective function in Step Two is

$$\min_{\Theta_2}\{q(\widehat{\Theta}_1, \Theta_2)'\widehat{W}q(\widehat{\Theta}_1, \Theta_2)\}, \tag{11}$$

where $q(\cdot)$ is the vector of the discrepancies mentioned above, and $\widehat{W}$ is an estimate of the optimal weighting matrix.[44] The choice of W takes into account that $q(\cdot)$ is a function of $\widehat{\Theta}_1$, which are point estimates with variances and covariances.[45]

3.2 Step Three: Tuition Preference

Given other colleges' equilibrium (data) tuition t_{-j}^*, I solve college j's tuition problem (8).[46] Under the true tuition preference parameters m, the optimal solution should match the tuition data.[47] The objective in Step Three is

$$\min_{m}\{(t^* - t(\widehat{\Theta}, m))'(t^* - t(\widehat{\Theta}, m))\},$$

where t^* is the data tuition profile, $t(\cdot)$ consists of each college's optimal tuition, and $\widehat{\Theta} \equiv [\widehat{\Theta}_0, \widehat{\Theta}_2]$ is the vector of fundamental parameter estimates from the previous two steps. I obtain the variance-covariance of $\widehat{m}$ using the Delta method, which exploits the variance-covariance structure of $\widehat{\Theta}$.

3.3 Identification

This subsection gives an overview of the identification. Discussion about the identification of specific parameters will be provided along with the estimation results. The identification relies on the following assumptions.

IA1: the number of student types is finite; idiosyncratic tastes are separable and independent from type-specific mean preferences; tastes are drawn from an i.i.d. single-mode distribution, with mean normalized to zero, and tastes are independent of (SAT, B, K).

IA2: at least one variable in the financial aid functions is excluded from the type distribution

---

[44]See the appendix for details.

[45]The standard errors of the parameter estimates in the second step and the third step account for the estimation errors in the previous step(s).

[46]Details are in the appendix.

[47]Given that there is only a single college market, there are at most 2 tuition levels observed per college group, the basis for the estimation of the colleges' objective functions. Therefore, pursing a conventional estimation approach is not sensible. Instead, I treat the nonlinear tuition best response functions as exact, which implies that the researcher observes all factors involved in a college's tuition decision, and saturate the model. This approach also enables me to recover the tuition preference parameters without solving the full tuition game. As is shown below, the fit to the tuition data is quite good, although there is no statistical criterion that can be applied.

function; conditional on (SAT, y), this variable is independent of K.

The intuition of identification is as follows. In the data, different application portfolios are chosen at different frequencies; the model predicts that students within the same type tend to choose similar application portfolios. Given IA1, the modes of these choices informs one of the number of types and the fraction of each type. The distributions of student type-related characteristics (assumed to be SAT and family income y) will differ around various modes, which informs one of the correlation between type K and (SAT, y).

Given IA2, students with the same (SAT, y) may differ in other family background variables that affect their expected financial aid. Such differences will lead to different application behaviors, for example, application versus non-application, within the same type. The sensitivity of students application choices to their expected financial aid conditional on (SAT, y) identifies type-specific expected utility from applying, which is a composite of application cost, type-specific admissions probabilities and preferences for colleges. For example, for a type whose expected utility from applying is marginal, their application behavior will differ a lot with the amount of financial aid they expect to obtain. Given the identification of type distribution, type-specific admissions probabilities are identified from the correlation between family income and admissions probabilities within an SAT group, because family income is assumed to affect admissions probabilities only via student type. Finally, type-specific preferences for colleges can be separated from application cost because application costs are common across all students, however, students of the same type but different SAT scores will face different admissions probabilities, hence different expected benefits from applying.

The arguments above do not depend on specific parametric assumptions. For example, Lewbel (2000) shows the identification of similar semiparametric models when an IA2-like excluded variable with a large support exists. However, to make the exercise feasible, I have assumed specific functional forms. Assuming student tastes are multinomial normal, the appendix shows a formal proof of identification.

Observing the same student multiple times via her applications, admissions and enrollment strengthens identification. For example, someone with a strong preference to attend college but low ability will diversify her risks by sending out more applications, but may be rejected by most of the college groups she applies to. Besides their sizes, the contents of application portfolios are also informative. In the model, a student's preferences for different colleges are correlated via her type-specific preference parameters. Consider students with the same SAT and family background, hence the same expected net tuition and ability. Without heterogeneity along the z dimension of student type, i.e., the dimension that captures students' preferences for public relative to private colleges, these students differ only in their i.i.d. idiosyncratic tastes. As a result, there should not be any systematic difference between their application portfolios. However, in the data when these students send out multiple applications, some concentrate on public colleges and some on private

colleges.[48] The patterns of such concentration, therefore, inform one about the distribution of z and its effects on students' preferences.

4 Data

4.1 NLSY Data and Sample Selection

In NLSY97, a college choice series was administered in years 2003-2005 to respondents from the 1983 and 1984 birth cohorts who had completed either the 12th grade or a GED at the time of interview. Respondents provided information about each college to which they applied, including name and location; any general financial aid they may have received; whether each college to which they applied had accepted them for admission, along with financial aid offered. Information was asked about each application cycle.[49] In every survey year, the respondents also reported on the college(s), if any, they attended during the previous year. Other available information relevant to this paper includes SAT/ACT score and financial-aid-relevant family information (home state, family income, family assets, race and number of siblings in college at the time of application).

The sample I use is from the $2,303$ students within the representative random sample who were eligible for the college choice survey in at least one of the years 2003-2005. To focus on first-time college application behavior, I define applicants as students whose first-time college application occurred no later than 12 months after they became eligible. Under this definition, $1,756$ students are either applicants or non-applicants.[50] I exclude applications for early admission. I also drop observations where some critical information, such as the identity of the college applied to, is missing. The final sample size is $1,646$.

4.2 College Groups and Choice Set

The elite/non-elite division of colleges is based on U.S. News and World Report 2001-2005.[51] The top 30 private universities and top 20 liberal arts colleges are considered as (private, elite). The (public, elite) group includes the top 30 public universities; and if no college in a state appears on that list, the best public (flagship) university within that state is included. Consistent with the cases of almost all states, I assume there is one elite public college per state and at most one application can be sent to the (public, elite) group in one's home state. Arguably, from a student's point of view, the flagship university in one's own state can be considered as (public, elite) even if it

---

[48]For example, among applicants with SAT above 1200 and family income above the 75th percentile, 46% applied only to public colleges and 21% applied only to private colleges.

[49]An application cycle includes applications submitted for the same start date, such as fall 2002.

[50]I exclude students who were already in college before their first reported applications. If a student is observed in more than one cycle, I use only her/his first-time application/non-application information.

[51]The report years I use correspond to the years when most of the students in my sample applied to colleges, and the rankings had been very stable during that period.

is not ranked at the top nation wide. However, it is far less realistic to assume that every state has a private elite college. Meanwhile, the data suggest that whether or not a college is in one's home state may not be a significant factor differentiating colleges within the (private, elite) group.[52] For these reasons and concerns about the sample size, I assume private elite colleges as national and abstract location from their characteristics.[53]

Table 1 Four-Year College Groups

	(pri,elite)	(pub,elite)	(pri,non)	(pub,non)
Num. of colleges (Potential[a])	51	56	1921	595
Num. of colleges (Applied[b])	37	56	312	268
Capacity[c] (%)	1.0	7.7	11.5	21.9

a. Total number of colleges in each group (IPEDS).

b. Number of colleges applied to by some students in the sample.

c. Capacity = Num. of students in the sample enrolled in each group/sample size.

To keep the estimation tractable, I assume that within each of the four groups of 4-yr colleges, a student can send out at most two applications.[54] This assumption is not as restrictive as it seems. First of all, as long as the student can apply for more than one college within a group, the model will be able to capture the competition between colleges within a group. This is true because the "threat" to a college is the *one* best competing alternative a student has. Moreover, the assumption is in line with the majority of students' behavior: 83% of applicants applied to no more than 2 colleges within each of the four college groups. It does, however, abstracts from some very interesting but non-typical aspects of the data, such as the behavior of some "elite" students who apply to many elite colleges. The empirical definitions of application, admission and enrollment, as well as the interpretation of the number of colleges are adjusted to accommodate the aggregation of colleges, as specified in Appendix B4-B5.

4.3 Summary Statistics

Table 2 summarizes characteristics among students who did (not) apply to 4-yr colleges, and those who attended a 2-yr (4-yr) college. Clear differences emerge between non-applicants and applicants: the latter are much more likely to be female, white, with higher SAT scores and with higher family

---

[52]Among students who applied to private elite colleges, about 80% applied to such colleges out of home state.

[53]With the small number of students who applied to any private elite college, dividing this group by location will generate a lot of "empty cells," i.e., choices not chosen by any student in the sample. This will cause problems to the estimation and make the parameter estimates highly imprecise.

[54]That is, the maximum number of applications is set at 8. Allowing for more applications will considerably increase the computation burden since the number of possible application portfolios grows exponentially with the number of applications. Because almost all states have only one public elite college, it is further resticted that at most one application can be sent to this group in state.

income.[55] About 23% of students in the sample attended a 2-yr college, while 42% attended a 4-yr college. Compared to the former, the latter are more likely to be female, white, with higher SAT scores, but the average family incomes are similar across these two groups.

Table 3 summarizes the distribution of the sizes of application portfolios. Among all students, over 54% did not apply to any 4-yr college. This is mainly driven by the predominant non-application decisions among low SAT students; most students with higher SAT scores applied. Among applicants, most applied to only one college.

Table 4 shows group-specific application rates and admissions rates. The application rate increases as one goes from (pri,elite) to (pub,non).[56] However relative to their capacities (shown in Table 1), elite colleges receive disproportionately higher fractions of applications than non-elite colleges. For example, (pub,non) is almost 22 times as large as (pri,elite), but the application rate for (pub,non) is only 7 times as high as that for (pri,elite). Consistently, the admissions rate increases monotonically from 51% in (pri,elite) to 94% in (pub,non). Colleges' selectivity can also be seen from the composition of their enrollees. In (pri,elite) group, 94% of enrollees have high SAT, while only 18% of enrollees have high SAT in (pub,non) group.

Table 2 Student Characteristics

	Non-Applicants	Applicants	2-Yr Attendees	4-Yr Attendees
Female	43.0%	53.1%	47.1%	53.5%
Black	17.6%	13.4%	15.2%	12.3%
Family Income^a	39,822 (32,428)	68,231 (51,208)	70,605 (51,279)	70,179 (50,995)
$SAT^b = 1$	80.2%	16.6%	58.0%	14.0%
$SAT = 2$	16.7%	59.7%	35.8%	60.3%
$SAT = 3$	3.1%	23.7%	6.2%	25.7%
Observations	892	754	374	693

a. in 2003 dollars, standard deviations are in parentheses.

b. SAT=1 if SAT or ACT equivalent is lower than 800 (Obs: 840).[57]

SAT=2 if SAT or ACT equivalent is between 800 and 1200 (Obs: 599).

SAT=3 if SAT or ACT equivalent is above 1200 (Obs: 207).

---

[55]Similar patterns have been found in other studies using different data. For example, Arcidiacono (2005), using data from the National Longitudinal Study of the Class of 1972, and Howell (2010), using data from National Education Longitudinal Study of 1988 report similar patterns.

[56]Application rates across groups will not necessarily add up to 100%, since some students applied to multiple college groups.

[57]Students who did not take the SAT or ACT test are categorized into SAT=1 group, since the other observable characteristics of these students and the outcomes of their applications, admissions and enrollment are very similar to those with low SAT/ACT scores.

Table 3 Number of Applications (%)

	$n = 0$	$n = 1$	$n \geq 2$
All Students	54.2	28.0	17.8
$SAT = 1$	85.1	12.1	2.8
$SAT = 2$	24.9	45.2	29.9
$SAT = 3$	13.0	43.0	44.0

Table 4 Application & Admission: All Applicants

(%)	(pri,elite)	(pub,elite)	(pri,non)	(pub,non)
Application Rate	9.7	31.8	44.6	71.5
Admission Rate	53.4	83.0	91.4	94.0
SAT=3 Enrollees	93.8	36.2	27.9	17.8

Num of all applicants: 754

Application rate=num. of group-specific applications/num. of all applications

Admission rate=num. of group-specific admissions/num. of group-specific applications

One pattern not shown in the tables is students' home bias: 66% of all 4-yr applicants applied to in-state colleges only, and 76% of all 4-yr attendees went to in-state colleges. This can be partly explained by the tuition differences shown in Table 5, where the within-group average tuition is based on information from the Integrated Postsecondary Education Data System (IPEDS). Public colleges price-discriminate against out-of-state students by charging them 3 times as much as they do in-state students, although still lower than tuition charged by private colleges. The last two rows summarize financial aid data. Relative to students admitted to elite colleges, a higher fraction of students admitted to non-elite colleges receive college financial aid. In addition, 40% of admitted students receive some outside financial aid that helps to fund college attendance in general.

Table 5 Tuition and Financial Aid

		(pri,elite)	(pub,elite)	(pri,non)	(pub,non)	2-yr College	General
Tuition[a]	(In-State)	27,033	5,000	17,296	3,969	2,744	–
	(out-of-state)		14,435		10,215	–	
Aid Recipients[b]		25%	24.1%	49.5%	27.2%	–	39.9%
Average Aid Offered		12,440	6,962	11,389	5,208	3,095	4,326

a. Tuition and aid are measured in 2003 dollars.

b. Num. of aid offers/num. of admissions in the sample. N/A for 2-yr colleges due to open admissions.

5 Empirical Results

Based on the likelihood ratio test, I report the results for the model where in-state and out-of-state students with the same (SAT, A) face the same admissions probabilities.

5.1 Parameter Estimates

5.1.1 Student Preferences for Colleges

Table 6 Preferences for Colleges

($1,000)	(pri,elite)	(pub,elite)	(pri,non)	(pub,non)	2-yr
$\overline{u}_g$(A=1,z=1)[a]	−187.7 (188.0)	−183.2 (5.1)	−123.5 (3.8)	−188.6 (4.4)	−38.1 (1.7)
$\overline{u}_g$(A=2,z=1)	−42.2 (66.5)	−37.2 (4.6)	31.0 (1.4)	56.8 (2.1)	36.1 (1.4)
$\overline{u}_g$(A=3,z=1)	−52.8 (21.4)	127.3 (0.4)	8.2 (7.6)	73.2 (3.9)	9.8 (4.5)
$\overline{u}_g$(A=2,z=2)	−74.4 (29.4)	−115.7 (34.9)	96.6 (4.6)	19.4 (3.19)	−13.3 (5.6)
$\overline{u}_g$(A=3,z=2)	139.9 (14.3)	30.4 (14.5)	35.6 (19.5)	−66.2 (16.4)	−12.7 (33.2)
$\sigma^2_{\epsilon_{1g}}$ (college group)	49.9 (8.4)	24.9 (3.0)	42.3 (1.0)	57.4 (1.8)	61.4 (1.2)
$\sigma^2_{\epsilon_2}$ (specific college)			61.5 (1.2)		

[a] The restriction $\overline{u}_g$(A=1,z=2) = $\overline{u}_g$(A=1,z=1) holds at 10% significance level.

There is significant heterogeneity in students preferences for colleges, both across student types and within each type. Rows 1 to 5 of Table 6 show the values of college groups for an average student of a given type, relative to the non-college option. For an average low-ability (A=1) student, the non-college option is better than any college option. This explains why the majority of (low family income, low SAT) students, who are most likely to be of low ability, do not apply to or attend any college in the data. Due to their low family income, these students would obtain very generous financial aid if they were admitted to any college. Moreover, from an individual student's point of view, there is a nontrivial probability that she would be admitted to some college. Given the apparent "unclaimed" benefits for these students, their predominant choices of the non-college option indicate that the values of colleges must be low for most of them, a finding consistent with previous literature such as Cunha, Heckman and Navarro (2005).[58,59]

In most cases, middle-A students rank non-elite colleges over elite colleges, while the opposite is true for high-A students. Such patterns are not completely surprising. For example, it is reasonable to believe that the effort costs required in elite colleges are higher than those required in non-elite colleges, and that these costs decrease with student ability. Considering the effort costs and the probabilities of success in different colleges, a mediocre student might be better off attending a

---

[58] Cunha, Heckman and Navarro (2005) find very high psychic costs of attending college (median around $500,000), which stand in for expectational errors and attitudes towards risk that explain why agents who face high gross returns do not go to college.

[59] Another potential but perhaps minor explanation is borrowing constraint. For example, Cameron and Heckman (1998) and Keane and Wolpin (2001), find that borrowing constraints have a negligible impact on college attendance, based on which I assume no borrowing constraint. Lochner and Monge-Naranjo (2011) find that conditional on AFQT, the correlation between family income and college attendance is weaker for the NLSY79 cohort than for the NLSY97 cohort, suggesting the later cohort may be more constained. Alternatively, their finding can be explained by the stronger correlation between family income and students' college ability, even after controlling for AFQT. That is, there are more students from the later cohorts that are constrained in their childhood when their pre-college human capital is formed.

non-elite college.

Holding ability constant, z-1 type in general value public and 2-yr colleges over private colleges, while the opposite holds for z-2 type. Private colleges and public colleges have different features that may fit some students better than others. For example, private colleges are usually smaller than public colleges, which may be an advantage for some students but a disadvantage for others.

By introducing types, the model explains the systematic differences in students' choices. The residual non-systematic differences in student choices are accounted for by their idiosyncratic preferences, which feature significant dispersions both for college groups ($\sigma_{\epsilon_{1g}}$) and for specific colleges (σ_{ϵ_2}). In sum, not only do students attach different values to the same college, but they also rank colleges differently. For example, attending an elite college is not optimal for all students.[60] Instead, each option (including the outside option) offered in the college market best caters to some groups of students.

5.1.2 Home Bias

Table 7 shows the disutility of attending colleges out of one's home state, which includes both extra monetary costs such as costs for transportation and residence, as well as psychic cost.[61] Such costs are found to be lower for high-A students, who are presumably better at adapting to new environment. Students who prefer public colleges over private colleges ($z = 1$) exhibit greater unwillingness to study far away from home. The identification of type-specific home biases comes from the correlation of student choices and their characteristics. For example, the fraction of applicants who applied only within home states is 50% among high-SAT applicants, as compared to 70% among other applicants. Similarly, controlling for the number of applications, for example at 2, the fraction of applicants who applied only within home states is 50% among students who applied to at least one private college, as compared to 64% among those who only applied for public colleges. Finally, the dispersion of student decisions to apply out of state among similar students identifies σ_ξ.

Table 7 Out-of-State Utility Cost

$1,000	(A=3,z=2)	(A<3,z=2)	(A=3,z=1)	(A<3,z=1)
Mean ($\overline{\xi}_K$)	22.5 (1.1)	26.1 (0.6)	37.1	40.7
$\overline{\xi}$(A,z=1)$-\overline{\xi}$(A,z=2)		14.6 (0.5)		
Dispersion (σ_ξ)		35.1 (0.5)		

[a] The restriction $\overline{\xi}$(A=1,z)=$\overline{\xi}$(A=2,z) holds at 10% significance level.

Remark   All three taste dispersions, across college groups ($\sigma_{\epsilon_{1g}}$), specific colleges (σ_{ϵ_2}) and home bias (σ_ξ), are necessary to explain the data. For example, suppose $\sigma_{\epsilon_2} = 0$, the following application

---

[60]This is consistent with findings from some other studies, for example, Dale and Krueger (2002).

[61]Studies on migration decisions often find that both the mean and the dispersion of moving costs are substantial. For example, Kennan and Walker (2011).

profile would not happen, where a student applied to two colleges, one out-of-state public and the other in-state private. The fact that she applied to an out-of-state public college, willing to pay out-of-state tuition, but not an in-state counterpart, reveals her taste for studying out of home state. However, if that is the case, she should not have chosen an in-state private college over an out-of-state counterpart. It is cases like this that identify σ_{ϵ_2}.

5.1.3 Application Costs

Table 8 Application Costs

$1,000	$n = 1$	$n = 2$	$n = 3$	$n \geq 4$
$C(n) - C(n-1)$	1.90 (0.3)	0.90 (0.05)	0.33 (0.02)	0.27 (0.03)

The cost for the first application is about $1,900$, but as the number of applications increases, the marginal cost rapidly decreases, suggesting the existence of some economies of scale. In interpreting application costs, on the one hand, one must remember that they incorporate all factors that make application costly, i.e., all student-side barriers to applying for colleges other than their ability and preferences. For example, the cost to collect information and prepare application materials, the stress to meet the application deadlines, and the anxiety felt while waiting for admissions results. On the other hand, student ability and preferences are far more important in explaining the application patterns found in the data. As an example, if one fixes all the other parameters and reduces application costs by half and simulate the student decision model, the fraction of non-applicants remains at 51%, as compared to 54% in the baseline model. For many students, application costs are irrelevant to their decisions. For example, average low-ability students, who derive negative utilities from colleges, will not apply even if application is costless.

The same cannot be said for students that are at the margin of applying and not applying, applying more and applying less. For example, since SAT is only a noisy measure of student ability, students of the same type may have different SAT scores, hence different admissions probabilities. Were application costs negligible, there should not be noticeable adjustment in their application choices to the admissions probabilities, since they share the same preferences on average. The extent to which such adjustment exists informs one of the importance of application costs.[62]

5.1.4 Ability Measures

One important feature of this model is information friction: students can only convey their abilities to colleges via noisy ability measures. The following two tables show the severity of such information friction. Based on the ability distribution parameter estimates, each row of Table 9.1 shows the distribution of SAT scores given ability, where each row adds up to 100%. Over 90% of the cases, Ability-1 students will score low in SAT, which makes it relatively easy to distinguish them from

---

[62]See the appendix for details.

the others based on SAT. However, SAT is less useful in distinguishing between medium-ability and high-ability types.

Table 9.1 SAT and Ability: Simulation

%	$P(SAT=1|A)$	$P(SAT=2|A)$	$P(SAT=3|A)$
$A = 1$	91.0	8.3	0.7
$A = 2$	18.3	65.8	15.9
$A = 3$	1.5	52.4	46.1

Table 9.2 reports the distribution of signals conditional on ability. Ability-2 students distinguish themselves from Ability-1 students primarily by their low probability of sending out low signals. Ability-3 students are much more likely than others to send high signals, and almost never send out low signals.[63]

Table 9.2 Signal Distribution

%	$P(s = 1|A)$	$P(s = 2|A)$	$P(s = 3|A)$
$A = 1$	94.1 (0.4)	1.4	4.5 (5.1)
$A = 2$	7.7 (0.5)	87.8	4.6 (2.4)
$A = 3$	0.04 (0.7)	46.7	53.3 (8.8)

Previous studies, for example, Cameron and Heckman (2001), have noted that family income has substantial influence on forming students' ability. Table 9.3 reinforces this finding by showing ability distribution among different family income groups, where each column adds up to 100%. Students from low income families are most likely to be of low ability, and hence very low preferences for colleges. The effect of such preferences on college attendance will be illustrated in the first counterfactual experiment.

Table 9.3 Family Income and Ability

%	$P(A|\text{Low Inc})$	$P(A|\text{Middle Inc})$	$P(A|\text{High Inc})$
$A = 1$	71.8	49.8	20.8
$A = 2$	25.1	40.2	49.6
$A = 3$	3.1	10.0	26.7

Low Inc: if family income is below 25th percentile (group mean $10,017)

Middle Inc: if family income is in 25-75th percentile (group mean $45,611)

High Inc: if family income is above 75th percentile (group mean $110,068)

5.2 Model Fit

Given the parameter estimates, I simulate the subgame perfect Nash equilibrium model (SPNE) and compare model predictions with the data.[64] Tables 10 shows the number of applications.

---

[63]Table G1.4 in the appendix shows the distribution of ability, SAT and signals among applicants.

[64]Model fits by SAT and by family income are in the appendix.

Table 11 shows the allocation of these applications and the admissions rates by college groups. Table 12 displays the fits of student final allocation in terms of college groups: the model slightly under-predict the fraction of students attending non-elite colleges. As Table 13 shows, the model replicates the pattern that most applications and attendance occur within home states. Finally, Table 14 contrasts model predicted tuition levels with the data.

Table 10 Model v.s. Data

Num of Applications (%)

Size	Data	Model
0	54.2	54.5
1	28.0	27.8
2 or more	17.8	17.7
χ^2 Stat	0.06	

$\chi^2_{2,0.05} = 5.99$

Table 11 Model v.s. Data

Application & Admission: Applicants (%)

Application Rate	Data	Model
(pri,elite)	9.7	9.4
(pub,elite)	31.8	29.0
(pri,non)	44.6	44.4
(pub,non)	71.5	67.6
Admission Rate		
(pri,elite)	53.4	58.5
(pub,elite)	83.0	90.1
(pri,non)	91.4	91.5
(pub,non)	94.0	95.9

* All Pass $\chi^2_{1,0.05}$ test.

Table 12 Model v.s. Data

Final Allocation of Students (%)

	Data	Model
(pri,elite)	1.0	1.5
(pub,elite)	7.7	8.0
(pri,non)	11.5	10.9
(pub,non)	21.9	20.2
2-yr college	22.7	22.9
Non-college	35.2	36.5
χ^2 Stat.		6.98

$\chi^2_{5,0.05}=11.07$

Table 13 Model v.s. Data: Home Bias

%	Data	Model
Home-Only Applicants[a]	65.6	67.5
Home-State Attendees[b]	76.2	78.0

[*] Both pass $\chi^2_{1,0.05}$ test.

[a] % students who apply only within home states among all 4-yr applicants.

[b] % students who attend home-state 4-yr colleges among all 4-yr attendees.

Table 14 Model v.s. Data: Tuition

$	(pri,elite)	(pub,elite)		(pri,non)	(pub,non)	
		In-State	Out-of-State		In-State	Out-of-State
Data	27,033	5,000	14,435	17,296	3,969	10,215
Model	27,530	5,090	13,892	16,891	3,451	10,540

6 Counterfactual Experiments

With the estimated model, I conduct two counterfactual experiments. Comparisons are made between the baseline SPNE and the new SPNE.[65]

6.1 Creating More Opportunities

To what extent can the government expand college access by increasing college capacities? To answer this question, I consider two counterfactual scenarios. In the first, community college tuition is maintained at its current level ($2,744), which also serves as the lower bound for 4-yr college

---

[65]In simulating the baseline model and the counterfactual experiments, I tried a wide range of initial guesses in my search for equilibrium. For each model, I find only one equilibrium.
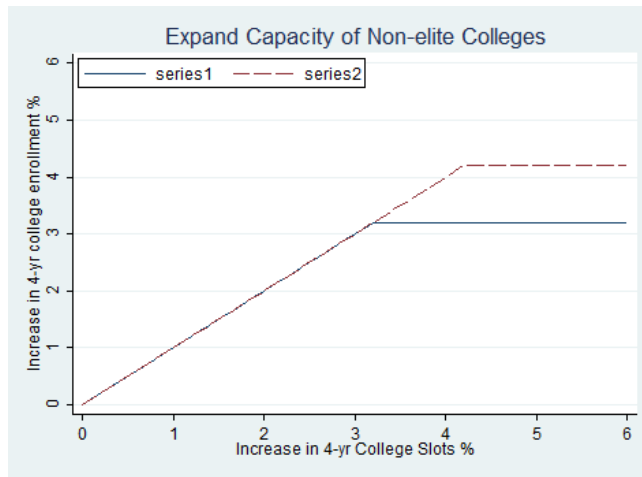
Figure 1: Expansion of Non-Elite Colleges

tuition. In a second, more aggressive scenario, community colleges become free and the lower bound on 4-yr college tuition is set to zero. Under each scenario, I conduct a series of expansion experiments and increase the capacities of non-elite public colleges by growing magnitudes while keeping the capacities of other colleges fixed.[66] For each capacity configuration, I solve the SPNE and examine the responses of colleges and students.[67]

The response of college enrollment to the increase in supply is shown in Figure 1. In each series, at the beginning of the expansion, there is a one-to-one response of 4-yr college enrollment to the increase in supply. Then, enrollment reaches a satiation point where there is neither excess demand nor excess supply of slots in non-elite public colleges and the equilibrium outcomes remain the same thereafter. The following tables report the two cases when the supply of non-elite public colleges is at its satiation point, labeled "New 1" ("New 2") for the first (second) scenario.

Table 15.1 shows changes in tuition. Under both counterfactual scenarios, non-elite public colleges cut their tuition levels for both in-state and out-of-state students to the lower bound, in order to attract enough students.[68] In response to the drastic action in non-elite public colleges, both non-elite private colleges and elite public colleges lower their tuition. Elite private colleges, in contrast, increase their tuition. The reason is as follows: relative to the number of students with high ability and strong preferences for elite private colleges, the total slots in these colleges are still scarce. When other colleges lower their tuition, an elite private college need not lower its tuition to attract enough students. Rather, increasing its tuition helps to screen out lower-ability students, who have less to gain from attending elite colleges and hence are more price-sensitive than high-

---

[66]Similar results hold in analogous experiments with non-elite private colleges' capacity. I increase the supply of non-elite colleges because they accommodate most college attendees and are most relevant to the overall access to college education.

[67]Appendix F studies the effect of tuition reduction on college enrollment, based on the student decision model. The effect is found to be small, which echoes results presented in this subsection.

[68]Colleges do not have to fill their capacities, and they can charge high tuition and leave some slots vacant. However, under the current situation and the estimated parameter values, it is not optimal for them to do so.

ability students. Therefore, even though an elite private college compete with colleges both in other groups and within its own group, increasing tuition is a good strategy. None of the other colleges, including the elite public, enjoy such market power: competition forces them to lower tuition.

Table 15.1 Increasing Supply: Tuition

$	(pri,elite)	(pub,elite)		(pri,non)	(pub,non)	
		In-State	Out-of-State		In-State	Out-of-State
Baseline	27,530	5,090	13,892	16,891	3,451	10,540
New 1	29,152	4,177	11,568	15,291	2,744	2,744
New 2	29,952	2,917	7,308	13,631	0	18

In both counterfactual cases, the expanded non-elite public colleges admit all of their applicants. Under Scenario 1, admissions rates also increase in all the other colleges. The major driving forces for the increased admissions rates are likely to differ across college groups. For colleges other than the elite private, higher admissions rates and lower tuition reflect their efforts to enroll enough students. The elite private colleges increase their admissions rates mainly because they are faced with a better self-selected applicant pool, as a result of the enlarged tuition gap. Under Scenario 2, elite private colleges continue to increase both their tuition and admissions rates. The admissions rates in elite public colleges and non-elite private colleges are slightly lower than their baseline levels due to their dramatic tuition reduction.

Table 15.2 Increasing Supply: Admissions

%	(pri,elite)	(pub,elite)	(pri,non)	(pub,non)
Baseline	58.5	90.1	91.5	95.9
New 1	63.6	90.8	92.1	100.0
New 2	71.8	89.8	91.3	100.0

Table 15.3 displays the attendance rates among all students. In Scenario 1, 4-yr college attendance rate increases by 2.6%, while 2-yr colleges lose 1% of their enrollees. In Scenario 2, 3.6% more students are drawn into colleges. Since the supply of non-elite public colleges exceeds demand if they are enlarged further, these increases represent the upper limits to which the government can increase college attendance by such expansions.

Table 15.3 Increasing Supply: Attendance

%	Baseline	New 1	New 2	All Open&Free
4-Yr	40.6	43.2	44.2	55.6
2-Yr	22.9	21.9	22.9	18.0

I also conduct a partial equilibrium experiment where all colleges are open and free. This is an extreme situation with unlimited supply of all colleges. The result is reported in the last column of Table 15.3. Four-year college attendance rate increases by 15%. Two-year colleges lose 5% of their

enrollees, however, most of their enrollees choose to stay instead of attending 4-yr colleges for free, highlighting the importance of 2-yr colleges. Considering total college enrollment, some students (10%) are indeed constrained by tuition and/or available slots. However, the vast majority of students who do not attend college under the base SPNE prefer the outside option over any college option.

To explain why expansion has such limited effects on enrollment, Table 15.4 shows the attendance rates by student ability. Under the baseline, only 28% of low-ability students attend any college and almost none attend 4-yr colleges. When all colleges become free and open, 18% more of them will be attracted to colleges, while the majority still choose the non-college option. In contrast, almost all students of higher ability attend colleges, mostly 4-yr ones. Therefore, the major barrier to college access is student ability and associated preferences, not college capacity or tuition.[69]

Table 15.4 Increasing Supply: Attendance by Ability

%	Baseline	New 1	New 2	All Open&Free
$A = 1$				
4-Yr	1.0	3.5	4.3	18.9
2-Yr	27.0	26.5	29.2	26.5
$A = 2$				
4-Yr	72.3	75.1	76.7	86.9
2-Yr	24.0	21.9	21.1	12.7
$A = 3$				
4-Yr	93.3	94.1	94.4	97.8
2-Yr	5.8	5.3	5.1	2.2

6.2   Ignoring Signals

In some countries, such as China, college admissions are based almost entirely on scores in a nation-wide test. Although such a system may save resources invested in the admissions process, such as the human resource employed in reading thousands of student essays, it ignores a valuable source of information about student ability. In the second counterfactual experiment, I assess the consequences of ignoring signals in the admissions process.[70]

Table 16.1 shows the changes in tuition under the new SPNE. Elite colleges draw on higher tuition to screen students when the information on ability provided by signals becomes unavailable, because their target enrollees, i.e., high-ability students, are less price-sensitive. However, the screening effect of tuition is not strong enough to make up for the loss of information embedded

---

[69]This finding is in line with earlier research. See, for example, Cameron and Heckman (1998, 2001) and Keane and Wolpin (2001).

[70]In a different counterfactual experiment, I examine the opposite case, where admission offices can only use student signals to distinguish between applicants. The results are similar. For example, elite (non-elite) colleges increase (decrease) their tuition. Details are available upon request.

in signals, and more high-ability applicants will be mistakenly rejected by elite colleges. Knowing this fact, non-elite colleges lower their tuition to compete for these students, who apply to them as insurance.

Table 16.1 Ignore Signals: Tuition

$	(pri,elite)	(pub,elite)		(pri,non)	(pub,non)	
		In-State	Out-of-State		In-State	Out-of-State
Baseline	27,530	5,090	13,892	16,891	3,451	10,540
New	30,028	5,131	14,079	14,800	3,083	9,426

In response to these tuition reductions, more students apply to colleges (Table 16.2). However, applicants apply less. As admissions only depend on SAT and students know their SAT scores, there is less uncertainty hence less need for portfolio diversification. The reduction in uncertainty is especially true for high-SAT applicants, who are now admitted in all colleges (Table 16.3). The overall admissions rates, however, decrease in all colleges, due to the low admissions rates among low-SAT applicants.

Table 16.2 Ignore Signals: Num of Applications

%	Num $= 0$	Num $= 1$	Num ≥ 2
Baseline	54.5	27.8	17.7
New	52.7	34.2	13.1

Table 16.3 Ignore Signals: Admission Rates

%	All		$SAT= 1$		$SAT= 2$		$SAT= 3$	
	Base	New	Base	New	Base	New	Base	New
(pri,elite)	58.5	66.7	22.1	N/A	41.3	32.9	76.2	100.0
(pub,elite)	90.1	88.9	16.0	6.5	91.0	100.0	98.0	100.0
(pri,non)	91.5	80.0	77.1	11.2	93.7	100.0	100.0	100.0
(pub,non)	95.9	90.6	83.1	62.3	98.5	100.0	100.0	100.0

*N/A: not applicable because of zero applicant.

With less information available, elite colleges experience a drop in their enrollee ability, while the non-elite ones get more high-ability students (Table 16.4). As shown in Table 16.5, average student welfare decreases by $600. Both middle-ability and high-ability students lose. The only winners are low-ability students, who gain because colleges find it harder to distinguish these students from others.

Table 16.4 Ignore Signals: % of High-Ability Students

%	(pri,top)	(pub,top)	(pri,non)	(pub,non)
Baseline	94.7	80.2	11.0	15.9
New	86.4	79.0	12.7	16.1

Table 16.5 Ignore Signals: Student Welfare

$1,000	Baseline	New
All	67.8	67.2
$A = 1$	10.4	11.3
$A = 2$	111.0	109.0
$A = 3$	150.4	148.8

7   Conclusion

In this paper, I have developed and structurally estimated an equilibrium model of the college market. It provides a first step toward a better understanding of the college market by jointly considering tuition setting, applications, admissions and enrollment. In the model, students are heterogeneous in their abilities and preferences. They face uncertainty and application costs when making their application decisions. Colleges, observing only noisy measures of student ability, compete for more able students via tuition and admissions policies. I have estimated the structural model via a three-step estimation procedure to cope with the complications caused by potential multiple equilibria. The empirical results suggest that the model closely replicates most of the patterns in the data.

My empirical analyses suggest that, first, there is substantial heterogeneity in students' preference for colleges. Expanding college capacities has very limited effects on college attendance: neither tuition cost nor college capacity is a major obstacle to college access; a large fraction of students, mainly low-ability students, prefer the outside option over any college option. Second, there are significant amounts of noise in various types of ability measures. When colleges lose access to one measure of student ability, elite colleges draw on higher tuition to help screen students, while non-elite colleges lower their tuition to compete for high-ability students who apply for them as insurance.

The methods developed in this paper and the main empirical findings are promising for future research. Building on Epple, Romano and Sieg (2006) and this paper, a model that endogenizes applications, admissions and financial aid would provide a more comprehensive view of the college market. Building on Arcidiacono (2005) and this paper, a model that studies the strategic interactions between colleges and students and links them to students' labor market outcomes would also be an important extension. The latter will become feasible as more information of students' labor market outcomes becomes available from future surveys of the NLSY97.

Another important direction for future research is to study the long-run equilibrium in order to obtain a better understanding of the trend of college tuition and attendance. In a long run equilibrium model, one can be more explicit about why colleges value student ability. For example, higher-ability students are more likely to do better in the job market, which enhances the college's prestige and attractiveness to future applicants. Moreover, building on this paper and the studies

on childhood human capital formation, such as Cunha, Heckman and Schennach (2010), one can also be more specific about how student preferences for colleges are formed and how childhood investment decisions might affect and be affected by the college market. That is, although students' evaluations of colleges are taken as given in the short run, in the long run, these evaluations will evolve as an equilibrium outcome.

Finally, one can also endogenize college capacities in the long run. One approach to implement this extension is to introduce a cost function for college education, assuming free entry to the market. Equilibrium of the model would then depend on the form of the cost function. Estimation of such a model would require additional data on college expenses and non-tuition revenues, as well as application and admissions data over multiple years.

References

Arcidiacono, Peter. 2005. "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica*, 73(5): 1477-1524.

Arcidiacono, Peter and Sean Nicholson. 2005. "Peer effects in medical school." *Journal of Public Economics*, 89(2-3): 327-350

Avery, Christopher, Andrew Fairbanks and Richard Zeckhauser. 2003. *The Early Admissions Game: Joining the Elite*. Cambridge, MA: Harvard University Press.

Avery, Christopher, Jonathan Levin. 2010. "Early Admissions at Selective Colleges." *American Economic Review*, 100(6): 1-34.

Bajari, Patrick, C. Lanier Benkard and Jonathan Levin. 2007. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, 75(5): 1331–1370.

Bajari, Patrick, Han Hong, John Krainer and Denis Nekipelov. 2010. "Estimating Static Models of Strategic Interactions." *Journal of Business and Economic Statistics*, 28(4): 469-482.

Berry, Steven, James Levinsohn and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841-890.

Berry, Steven, James Levinsohn and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy,* 112 (1): 68-104.

Bowen, William and Derek Bok. 1998. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions*. Princeton, NJ: Princeton University Press.

Bresnahan, Timothy and Peter Reiss. 1990. "Entry in Monopoly Markets." *The Review of Economic Studies*, 57: 531-553.

Chade, Hector, Gregory Lewis and Lones Smith. 2011. "Student Portfolios and the College Admissions Problem." *Social Science Research Network Working Paper* 135834.

Cameron, Stephen and James Heckman. 1998. "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males." *Journal of Political Economy,* 106(2): 262-333.

Cameron, Stephen and James Heckman. 2001. "The Dynamics of Educational Attainment for Black, Hispanic, and White Males." *Journal of Political Economy,* 109(3): 455-499.

Ciliberto, Federico and Elie Tamer. 2009. "Market Structure and Multiple Equilibria in Airline Markets." *Econometrica*, 77(6): 1791-1828.

Cunha, Flavio, James Heckman and Salvador Navarro. 2005. "Separating Heterogeneity from Uncertainty in Life Cycle Earnings." *Oxford Economic Papers*, Vol. 57(2): 191-261.

Cunha, Flavio, James Heckman, and Susanne Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica,* 78(3): 883–931.

Dale, Stacy and Alan Krueger. 2002. "Estimating The Payoff To Attending A More Selective College: An Application Of Selection On Observables And Unobservables." *The Quarterly Journal of Economics*, 117(4): 1491-1527.

de Paula, Aureo. 2013 "Econometric Analysis of Games with Multiple Equilibria." *Annual Review of Economics*, 5: 107-131.

Epple, Dennis, Richard Romano and Holger Sieg. 2006. "Admission, Tuition, And Financial Aid Policies in The Market For Higher Education." *Econometrica,* 74(4): 885-928.

Howell, Jessica. 2010. "Assessing the Impact of Eliminating Affirmative Action in Higher Education." *Journal of Labor Economics*, 28(1): 113-166.

Kane, Thomas. 1998. "Racial and Ethnic Preferences in College Admissions." in *The Black-White Test Score Gap,* ed. Christopher Jencks and Meredith Phillips, 971-996. Washington D.C.: Brookings Institution Press.

Keane, Michael and Kenneth Wolpin. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review,* 42(4): 1051-1103.

Kennan, John and James Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica,* 79(1): 211–251.

Leslie, Larry, and Paul Brinkman. 1998. *The Economic Value of Higher Education.* New York: Macmillan.

Lewbel, Arthur. 2000. "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics,* 97: 145-177.

Light, Audrey and Wayne Strayer. 2002. "From Bakke to Hopwood: Does Race Affect College Attendance and Completion?" *The Review of Economics and Statistics,* 85: 34-44.

Lochner, Lance and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital" *American Economic Review,* 101(6): 2487-2529.

Mailath, George, Masahiro Okuno-Fujiwara and Andrew Postlewaite. 1993. "Belief-Based Refinements in Signalling Games." *Journal of Economic Theory,* 60: 241-276.

Manski, Charles and David Wise. 1983. *College Choice in America.* Cambridge, MA: Harvard University Press

McFadden, Daniel. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica,* 57(5): 995-1026.

Meijer, Erik and Jelmer Ypma. 2008. "A Simple Identification Proof for a Mixture of Two Univariate Normal Distributions." *Journal of Classification,* 25: 113-123.

Moro, Andrea. 2003. "The Effect of Statistical Discrimination on Black-White Wage Differentials: Estimating a Model with Multiple Equilibria." *International Economic Review,* 44(2): 467-500.

Nagypál, Eva. 2004. "Optimal Application Behavior with Incomplete Information." mimeo, Northwestern University.

Rothschild, Michael and Lawrence White. 1995. "The Analytics of the Pricing of Higher Education and Other Services in Which the Customers Are Inputs." *Journal of Political Economy,* 103(3): 573-586.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results from Dartmouth Roommates." *Quarterly Journal of Economics,* 116: 681–704.

Tamer, Elie. 2003. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *The Review of Economic Studies,* 70(1): 147-165.

Winston, Gordon. 1999. "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education." *Journal of Economic Perspectives,* 13(1): 13-36.

Zimmerman, David. 2003. "Peer Effects in Higher Education: Evidence from a Natural Experiment." *The Review of Economics and Statistics,* 85: 9–23.

APPENDIX:

A. Model Details:

A1. College Admissions Problem

The following formally derives a college's optimal admissions policy without discrimination based on student origin.[71] Given tuition profile t, students' strategies $Y(\cdot), d(\cdot)$ and other colleges' admissions policies e_{-j}, college j solves the following problem:

$$\max_{e_j(s,SAT|t)} \left\{ \sum_{s,SAT} e_j(s,SAT|t)\alpha_j(s,SAT|t,e_{-j},Y,d)\mu_j(s,SAT|\cdot)\gamma_j(s,SAT|\cdot) \right\} \tag{12}$$

$$s.t. \sum_{s,SAT} e_j(s,SAT|t)\alpha_j(s,SAT|t,e_{-j},Y,d)\mu_j(s,SAT|\cdot) \le \kappa_j$$

$$e_j(s,SAT|t) \in [0,1],$$

where $e_j(s,SAT|t)$ is college j's admissions policy for its applicants with (s,SAT), $\alpha_j(s,SAT|t,e_{-j},Y,d)$ is the probability that such an applicant will accept college j's admission. $\gamma_j(s,SAT|t,e_{-j},Y,d)$ is the expected ability of such an applicant conditional on her accepting j's admission. $\mu_j(s,SAT|t,e_{-j},Y,d)$ is the measure of j's applicants with (s,SAT). The first order condition for problem (12) is

$$\gamma_j(s,SAT|\cdot) - \nu_j + \nu_a - \nu_b = 0,$$

where ν_j is the multiplier associated with the capacity constraint, i.e., the shadow price of a slot in college j. ν_a and ν_b are adjusted multipliers associated with the constraint that $e_j(s,SAT|t) \in [0,1]$.[72]

If it admits an applicant with (s,SAT) and the applicant accepts the admission, college j must surrender a slot from its limited capacity, thus inducing the marginal cost ν_j. The marginal benefit is the expected ability of such an applicant conditional on her accepting j's admission. Balancing between the marginal benefit and the marginal cost, the solution to college j's admissions problem is characterized by:

$$e_j(s,SAT|t) \begin{cases} = 1 & \text{if } \gamma_j(s,SAT|\cdot) - \nu_j > 0 \\ = 0 & \text{if } \gamma_j(s,SAT|\cdot) - \nu_j < 0 \\ \in [0,1] & \text{if } \gamma_j(s,SAT|\cdot) - \nu_j = 0 \end{cases}, \tag{13}$$

$$\sum_{s,SAT} e_j(s,SAT|t)\alpha_j(s,SAT|\cdot)\mu_j(s,SAT|\cdot) \le \kappa_j, \tag{14}$$

and

$$\nu_j \begin{cases} \ge 0 \text{ if } (14) \text{ is binding} \\ = 0 \text{ if } (14) \text{ is not binding} \end{cases}.$$

---

[71] The derivation of policy with origin-based discrimination is similar: a student's origin will be observed and the argument in $e_j(\cdot)$, $\alpha_j(\cdot)$, $\mu_j(\cdot)$ and $\gamma_j(\cdot)$ will extend from (s,SAT) to $(s,SAT,I(l_i = l_j))$.

[72] ν_a, ν_b are the multipliers associated with $\alpha_j(s,SAT|\cdot)\mu_j(s,SAT|\cdot)e_j(s,SAT|t) \in [0,1]$.

To implement its admissions policy, college j will first rank its applicants with different (s, SAT) by their expected ability conditional on their acceptance of j's admissions. All applicants with the same (s, SAT) are identical to the college and hence are treated equally. Everyone in an (s, SAT) group will be admitted if 1) this (s, SAT) group is ranked highest among the groups whose admissions are still to be decided, 2) their marginal contribution to the college is positive, and 3) the expected enrollment of this group is no larger than college j's remaining capacity, where j's remaining capacity equals κ_j minus the sum of expected enrollment of groups ranked above. A random fraction of an (s, SAT) group is admitted if 1) and 2) hold but 3) fails, where the fraction equals the remaining capacity divided by the expected enrollment of this group. As a result, a typical set of admissions policies for the ranked (s, SAT) groups, $\{e_j(s, SAT|t)\}$, would be $\{1, ..., 1, \varepsilon, 0, ..., 0\}$, with $\varepsilon \in (0, 1)$ if the capacity constraint is binding, and $\{1, ..., 1\}$ if the capacity constraint is not binding or just binding.

A1.1 Calculating $\alpha_j(s, SAT|t, e_{-j}, Y, d)$ and $\gamma_j(s, SAT|t, e_{-j}, Y, d)$

All objects depend on $\{t, e_{-j}, Y, d\}$. To save notation, the dependence is suppressed. Let $\Pr(accept|X, SAT, \eta, j)$ be the probability that a student with characteristics (X, SAT, η) who applies to College j accepts j's admission. Let $F(X, \eta|s, SAT, j)$ be the distribution of (X, η) conditional on (s, SAT) and application to j. The probability that an applicant with (s, SAT) accepts j's admission is:

$$\alpha_j(s, SAT|\cdot) = \int \Pr(accept|X, SAT, \eta, j) dF(X, \eta|s, SAT, j).$$

Let $\Pr(O_{-j}|A, SAT) \equiv \prod_{l \in O \setminus j} p_l(A, SAT) \prod_{j' \in Y \setminus O} (1 - p_{j'}(A, SAT))$ be the probability of admission set O for a student with (A, SAT), with college j admitting her for sure,

$$\Pr(accept|X, SAT, \eta, j) = \sum_{O_{-j} \subseteq Y(X, SAT) \setminus \{j\}} \Pr(O_{-j}|A, SAT) I(j = d(X, SAT, \eta, O)).$$

That is, the student will accept j's admission if j is the best post-application choice for her. The distribution $F(X, \eta|s, SAT, j)$ is given by

$$
\begin{aligned}
dF(X, \eta|s, SAT, j) &= \frac{P(s|A) I(j \in Y(X, SAT)) dH(X, \eta|SAT)}{\mu_j(s, SAT|\cdot)}, \\
\mu_j(s, SAT|\cdot) &= \int P(s|A) I(j \in Y(X, SAT)) dH(X, \eta|SAT),
\end{aligned}
$$

where $H(X, \eta|SAT)$ is exogenous and equal to the product of type distribution, the distribution of ex-post shocks and the distribution of family backgrounds conditional on SAT.

The expected ability of applicant (s, SAT) conditional on acceptance is

$$\gamma_j(s, SAT|\cdot) = \frac{\int A \Pr(accept|X, SAT, \eta, j) dF(X, \eta|s, SAT, j)}{\alpha_j(s, SAT|\cdot)}.$$

A2. Proof of existence in a simplified model.

Assume there are two colleges $j \in \{1, 2\}$ and a continuum of students divided into two ability levels. The utility of the outside option is normalized to 0. The utility of attending college 1 is $u_1(A)$ for all with ability A, and that of attending college 2 is $u_2(A) + \epsilon$, where ϵ is i.i.d. idiosyncratic taste. There are two SAT levels and two signal levels. There is no ex-post shock. Some notations to be used: for an (A, SAT) group, let the fraction of students that do not apply to any college be $\theta_{A,SAT}^0$, the fraction of those applying to college j only be $\theta_{A,SAT}^j$ and the fraction applying to both be $\theta_{A,SAT}^{12}$. For each (A, SAT) group, $\theta_{A,SAT} \in \Delta$, a 3-simplex. For all four (A, SAT) groups, $\theta \in \Theta \equiv \Delta^4$. On the college side, each college chooses admissions policy $e_j \in [0, 1]^4$, where 4 is the number of (s, SAT) groups faced by the college.

Proposition 1 *For any given tuition profile t, an application-admission equilibrium exists.*

Proof. Step 1: The application-admission model can be decomposed into the following sub-mappings:

Taking the distribution of applicants, and the admissions policy of the other college as given, college j's problem (12) can be viewed as the sub-mapping

$$M_j : \Theta \times [0, 1]^4 \rightrightarrows [0, 1]^4,$$

for $j = 1, 2$. Taking college admissions policies as given, the distribution of students is obtained via the sub-mapping

$$M_3 : [0, 1]^4 \times [0, 1]^4 \to \Theta.$$

An equilibrium is a fixed point of the mapping:

$$
\begin{aligned}
M \quad &: \quad \Theta \times [0, 1]^4 \times [0, 1]^4 \rightrightarrows \Theta \times [0, 1]^4 \times [0, 1]^4 \\
s.t. \quad \theta \quad &\in \quad M_3(e_1, e_2) \\
e_j \quad &\in \quad M_j(\theta, e_{j'}) \ j, j' \in \{1, 2\}, j \neq j'.
\end{aligned}
$$

Step 2: Show that Kakutani's Fixed Point Theorem applies in mapping M and hence an equilibrium exists.

1) The domain of the mapping, being the product of simplexes, is compact and non empty.

2) It can be shown that the correspondence $M_j(\cdot, \cdot)$ is compact-valued, convex-valued and upper-hemi-continuous, for $j = 1, 2$. In particular, the (s, SAT)'th component of $M_j(\theta, e_{j'})$ is characterized by (13) and (14), where $\gamma_j(s, SAT) + M(t_j; m_j) - \lambda_j$ is continuous in $(\theta, e_{j'})$.

3) Aggregate individual optimization into distribution of students θ.

Generically, each student has a unique optimal application portfolio as the solution to (6). For given (A, SAT), there exist $\epsilon^*(e) \geq \epsilon^{**}(e)$, both continuous in e, such that:

For $\epsilon \geq \epsilon^*(e)$, $Y(A, SAT, \epsilon) = \begin{cases} \{2\} \text{ if } C(2) - C(1) > k_1(e) \\ \{1, 2\} \text{ otherwise} \end{cases}$;

for $\epsilon \in [\epsilon^{**}(e), \epsilon^*(e))$, $Y(A, SAT, \epsilon) = \{1, 2\}$; and

for $\epsilon < \epsilon^{**}(e)$, $Y(A, SAT, \epsilon) = \begin{cases} \{1\} \text{ if } C(1) \leq k_2(e) \\ \emptyset \text{ otherwise} \end{cases}$,

where $k_1(e)$ and $k_2(e)$ are continuous in e. Therefore, the (A, SAT) population can be mapped into a distribution $\theta_{A,SAT} \in \Delta$, and this mapping is continuous in e. Because the mapping from $[0, 1]^4 \times [0, 1]^4$ into the individual optimal portfolio is a continuous function, and the mapping from the individual optimization to Θ is continuous, the composite of these two mappings, M_3, is single-valued and continuous.[73]

Given 1)-3), Kakutani's Fixed Point Theorem applies.[74] ∎

Since for every t, $AE(t)$ exists in the subsequent game, an SPNE exists if a Nash equilibrium exists in the tuition setting game. Let $\bar{t}_j$ denote some large positive number, such that for any t_{-j}, the optimal $t_j < \bar{t}_j$. $\bar{t}_j$ exists because the expected enrollment, hence college j's payoff goes to 0, as t_j goes to ∞. Define the strategy space for college j as $[0, \bar{t}_j]$, which is nonempty, compact and convex. The objective function of college j is continuous in t, since the distribution of applicants, and hence the total expected ability, is continuous in t. Given certain regularity conditions, the objective function is also quasi-concave in t_j. The general existence proof for Nash equilibrium applies.

B. Data Details

B1. The NLSY97 consists of a sample of 8984 youths who were 12 to 16 years old as of December 31, 1996. There is a core nationally representative random sample and a supplemental sample of blacks and Hispanics. Annual surveys have been completed with most of these respondents since 1997.

B2. Empirical Definition of Early Admission:

1) Applications were sent earlier than Nov. 30th, for attendance in the next fall semester and

2) The intended college has early admissions/ early decision/ rolling admissions/ priority admissions policy,[75] and

3) Either a. one application was sent early and led to an admission or

b. some application(s) was (were) sent early but rejected, and other application(s) was (were) sent later.

---

[73] In the case of four schools, ϵ becomes a 3-dimension vector, as are the cutoff tastes. To show continuity, we change one dimension of ϵ at a time while keeping the other dimensions fixed.

[74] When there are $J > 2$ schools, Step 1 of the proof can be easily extended. In Step 2, ϵ, and hence the cutoffs, will be of $J - 1$ dimensions. Obtaining an analytical solution to these cutoffs is much more challenging.

[75] The data source for college early admission programs is 1) Christopher et al. (2003), and 2) web information posted by individual colleges.

B3. Since 1983, U.S. News and World Report has been publishing annual rankings of U.S. colleges and is the most widely quoted of its kind in the U.S.[76] Each year, seven indicators are used to evaluate the academic quality of colleges for the previous academic year.[77]

B4. Empirical Definition of Applications, Admissions and Enrollment

A student is said to have applied/been admitted once (twice) to Group (pri,elite) if she applied to/was admitted to one (more than one) college within this group; and is said to have enrolled in this group if he enrolled in any college in this group. For other three college groups, the definitions are similar, but with in-state and out-of-state distinctions.

B5. Interpretation of the Number of Colleges

The number of colleges per group can be interpreted as follows: There is one elite public college per state. There are $n \geq 2$ elite private colleges nationwide, each with $\frac{1}{n}$ of the total capacity of the (private, elite) group. For the non-elite colleges, there are n_l private and n'_l public colleges in state l, and each shares $\frac{1}{n_l} \left(\frac{1}{n'_l} \right)$ of the total capacity of its group in state l. n_l, $n'_l \geq 2$ are both proportional to the population of students in state l.[78] To reduce computation, I assume that after seeing all colleges, a student will include, as final alternatives to consider, two out-of-state elite public colleges, two elite private colleges, two in-state and two out-of-state non-elite private colleges, two in-state and two out-of-state non-elite public colleges, together with the one in-state elite public college. Given the i.i.d. tastes for colleges across students, each college within a group-state combination will be faced with the same distribution of students.

C. Details on Estimation

C1. Details on SMLE:

(1) Approximate the following integration via a kernel smoothed frequency simulator [79]

$$\int I(Y_i|K, SAT_i, B_i, \epsilon) I(d_i|O_i, K, B_i, \epsilon, \zeta, \eta) dG(\epsilon, \zeta, \eta). \tag{15}$$

---

[76] The exception is 1984, when the report was interrupted.

[77] These indicators include: assessment by administrators at peer institutions, retention of students, faculty resources, student selectivity, financial resources, alumni giving, and (for national universities and liberal arts colleges) "graduation rate performance", the difference between the proportion of students expected to graduate and the proportion who actually do. The indicators include input measures that reflect a school's student body, its faculty, and its financial resources, along with outcome measures that signal how well the institution does its job of educating students.

[78] The total number of seats in all colleges within a group-state is assumed to be proportional to the number of students in that state. As such, the fraction of in-state students that can be accommodated by in-state colleges is the same across states. Since n_l and n'_l are proportional to the state population, in equilibrium, there will be fewer out-of-state applications sent to a less-populated state (with fewer colleges) than to a more-populated state (with more colleges). Moreover, in equilibrium, both the in-state and out-of-state applications to a group-state will be evenly distributed across the n_l colleges within a state-college-group. As such, in equilibrium, each college within a college group will be able to accommodate the same fraction of its *applicants*, regardless of its state population.

[79] I describe the situation where I do not observe any information about the student's financial aid. For students with some financial aid information, the observed financial aid replaces the random draw of the corresponding financial aid shock.

For each student (SAT_i, B_i), I draw shocks $\{(\epsilon_{ir}, \eta_{ir})\}_{r=1}^R$ from their joint distribution $G(\cdot)$. These shocks are the same across K for the same student i, but are i.i.d. across students. All shocks are fixed throughout the estimation. Let u_{jir} be the ex-post value of college j for student$_{ir}$ with $(K, SAT_i, B_i, \epsilon_{ir}, \eta_{ir})$, let $v_{ir} = \max\{0, \{u_{jir}\}_{j \in O_i}\}$, let $V_{ir}(Y)$ be the ex-ante value of portfolio Y for this student, and $V_{ir}^* = \max_{Y \subseteq J}\{V_{ir}(Y)\}$. (15) is then approximated by:

$$\frac{1}{R} \sum_{r=1}^R \frac{\exp[(V_{ir}(Y_i) - V_{ir}^*)/\tau_1]}{\sum_{Y \subseteq J} \exp[(V_{ir}(Y) - V_{ir}^*)/\tau_1]} \frac{\exp[(u_{d_i ir} - v_{ir})/\tau_2]}{\sum_{j \in O_i} \exp[(u_{jir} - v_{ir})/\tau_2]},$$

where τ_1, τ_2 are smoothing parameters. When $\tau \to 0$, the approximation converges to the frequency simulator.

(2) Solving the optimal application problem for student $(K, SAT_i, B_i, \epsilon_{ir})$:

$$V_i(Y) = \sum_{O \subseteq Y} \Pr_i(O) E_{(\eta, \zeta)} \max\{u_{0ir}, \{u_{jir}\}_{j \in O}\} - C(|Y|).$$

The Emax function has no closed-form expression and is approximated via simulation. For each $(K, SAT_i, B_i, \epsilon_{ir})$, draw M sets of shocks $\{\eta_m\}_{m=1}^M$. For each of the M sets of $(K, SAT_i, B_i, \epsilon_{ir}, \eta_m)$, calculate $\max\{u_{0irm}, \{u_{jirm}\}_{j \in O}\}$, where u_{jirm} denotes u_{jir} evaluated at the shock η_m. The Emax is the average of these M maximum values.

C2. Details on the Second-Step SMDE:

(1) Targets to be matched: for each of the Groups $2, 3$ and 4, there are 9 admissions probabilities to be matched $\{p_j(A, SAT)\}_{(A, SAT) \in \{1,2,3\} \times \{1,2,3\}}$. For (pri,elite), there are 6 admissions probabilities to be matched. Since no one in $SAT = 1$ group applied to (pri,elite), $\{p_1(A, SAT = 1)\}_{A \in \{1,2,3\}}$ are fixed at 0. The other four targets are the equilibrium enrollments simulated from the first step. In all, there are 37 targets to be matched using college-side parameters: $\{P(s|A)\}, \{\kappa_j\}_j$, ten of which are free.

(2) Optimal Weighting Matrix:

Let Θ^* be the true parameter values. The first-step estimates $\widehat{\Theta}_1$, being MLE, are asymptotically distributed as $N(0, \Omega_1)$. It can be shown that the optimal weighting matrix for the second-step objective function (11) is $W = [Q_1 \Omega_1 Q_1']^{-1}$, where Q_1 is the derivative of $q(\cdot)$ with respect to $\widehat{\Theta}_1$, evaluated at $(\widehat{\Theta}_1, \Theta_2^*)$. The estimation of W involves the following steps:

1) Estimate the variance-covariance matrix $\widehat{\Omega}_1$: in the case of MLE, this is minus the outer product of the score functions evaluated at $\widehat{\Theta}_1$. The score functions are obtained via numerically taking partial derivatives of the likelihood function with respect to each of the first step parameters evaluated at $\widehat{\Theta}_1$.

2) Obtain preliminary estimates $\widetilde{\Theta}_2 \equiv \arg \min_{\Theta_2}\{q(\widehat{\Theta}_1, \Theta_2)'\widetilde{W}q(\widehat{\Theta}_1, \Theta_2)\}$, where $\widetilde{W}$ is any positive-definite matrix. The resulting $\widetilde{\Theta}_2$ is a consistent estimator of Θ_2^*.

3) Estimate Q_1 by numerically taking derivative of $q(\cdot)$ with respect to $\widehat{\Theta}_1$, evaluated at $(\widehat{\Theta}_1, \widetilde{\Theta}_2)$.

In particular, let Δ_m denote a vector with zeros everywhere but the m'th entry, which equals a small number ε_m. At each $\left(\widehat{\Theta}_1 + \Delta_m, \widetilde{\Theta}_2\right)$, I simulate the student decision model and calculate the targets for the second-step estimation. Then holding student applications fixed, I solve for college optimal admissions and calculate the distance vector $q\left(\widehat{\Theta}_1 + \Delta_m, \widetilde{\Theta}_2\right)$. The m'th component of Q_1 is approximated by $[q\left(\widehat{\Theta}_1 + \Delta_m, \widetilde{\Theta}_2\right) - q\left(\widehat{\Theta}_1, \widetilde{\Theta}_2\right)]/\varepsilon_m$.

C3. Details on the Third-Step: Solving College j's Tuition Problem

Given $\widehat{\Theta}$, t^*_{-j} and some m, I examine college j's expected payoff at each trial tuition level t'_j and obtain the optimal tuition associated with this m. This procedure requires computing the series of application-admission equilibria $AE\left(\cdot, t^*_{-j}\right)$, which can only be achieved through simulation. To do so, I use an algorithm motivated by the rule of "continuity of equilibria," which requires, intuitively, that $AE(t'_j, t^*_{-j})$ be close to $AE(t_j, t^*_{-j})$ when t'_j is close to t_j. Specifically, I start from the equilibrium at the data tuition level $\left(t^*_j, t^*_{-j}\right)$, which is numerically unique for nontrivial initial beliefs $(p >> 0)$. $AE(t^*)$ is found to be unique numerically in my search for equilibrium starting from 500 different combinations of nontrivial initial beliefs. Then, I gradually deviate from t^*_j, for $\left(t^{"}_j, t^*_{-j}\right)$, I start the search for new equilibrium, i.e., the fixed point of admissions policies $e\left(.| \left(t^{"}_j, t^*_{-j}\right)\right)$, using, as the initial guess, the equilibrium $e\left(.| \left(t'_j, t^*_{-j}\right)\right)$ associated with the most adjacent $\left(t'_j, t^*_{-j}\right)$. The resulting series of $AE\left(., t^*_{-j}\right)$ is used to solve college j's tuition problem.

D. Detailed Functional Forms:

D1. Type Distribution:

$P(K|SAT, B) = \Pr(A = a|SAT, B)P(z|A) = \Pr(A = a|SAT, y)P(z|A)$, where y is family income, $\Pr(A = a|SAT, y)$ is an ordered logistic distribution and $P(z|A)$ is non-parametric. For $a = 1, 2, 3$

$$\Pr(A = a|SAT, y) = \frac{1}{1 + e^{-cut_a + \alpha_1 y_i + \alpha_2 I(SAT_i = 2) + \alpha_3 I(SAT_i = 3) + \alpha_4 y_i^2}}$$
$$- \frac{1}{1 + e^{-cut_{a-1} + \alpha_1 y_i + \alpha_2 I(SAT_i = 2) + \alpha_3 I(SAT_i = 3) + \alpha_4 y_i^2}}$$

where $cut_0 = -\infty$ and $cut_3 = +\infty$.

D2. Financial Aid Functions:

$$f_0(SAT_i, B_i) = \beta_0^0 + \beta_1^0 I(race_i = black) + \beta_2^0 I(SAT_i = 2) + \beta_3^0 I(SAT_i = 3)$$
$$+ \beta_4^0 y_i + \beta_5^0 asset_i + \beta_6^0 I(nsib > 0) + \beta_7^0 I(4\text{yr college})$$
$$f_{0i} = \max\{f_0(SAT_i, B_i) + \eta_{0i}, 0\},$$

where $nsib$ denotes the number of siblings in college at the time of i's application and $\eta_{0i} \sim$ i.i.d. $N(0, \sigma_{f_0}^2)$. For $j \geq 1$

43

$$f_j(SAT_i, B_i) =$$

$$\beta_0^1 \quad +\beta_1^1 I(race_i = \quad black) + \beta_2^1 I(SAT_i = 2) + \beta_3^1 I(SAT_i = 3) + \beta_4^1 y_i + \beta_5^1 asset_i$$
$$+\beta_6^1 I(nsib > \quad 0) + \beta_7^1 I(SAT_i = 2)I(j \in pri) + \beta_8^1 I(SAT_i = 3)I(j \in pri)$$
$$+\beta_9^1 I(j \in \quad (pri, elite)) + \beta_{10}^1 I(j \in 2yr)$$

$$f_{ji} = \max\{f_j(SAT_i, B_i) + \eta_{ji}, 0\}$$

where $\eta_{ji} \sim i.i.d.N(0, \sigma_{f_1}^2)$.

E. Identification

E1. Identification of type distribution, application cost, type-specific admissions probabilities and utilities

In the following, I will prove the identification of a model with two types.[80] The logic can be extended to a more general model with multiple types. I omit the more general proof as it requires much more complicated algebraic analyses that are very cumbersome to show.

To give the intuition, assume there is only one college and a student decides whether or not to apply. The intuition applies to the case with more colleges.[81] Let (SAT, y, h) be observable characteristics of a student, who is one of the two unobserved types $A \in \{1, 2\}$. Define $\lambda(SAT, y) \equiv \Pr(A = 1 | SAT, y)$. Each student has an idiosyncratic taste for the college $\epsilon \sim i.i.d.N(0, 1)$.[82] Let the type-specific (gross) utility from attending college be u_A^* and the value of the outside option be normalized to zero. The admission probabilities are (A, SAT)-specific, denoted as $p_{A,SAT}$. Let c be the application cost, and $f(SAT, y, h)$ the financial aid function net of tuition, which is differentiable in $h \in H = R$. Following (IA1) and (IA2) in the text, assume that h is independent of A conditional on (SAT, y) and ϵ is independent of (A, SAT, y, h).

E1.1 Identification of $\lambda(SAT, y)$, $u_1^* - \frac{c}{p_{1,SAT}}$ **and** $u_2^* - \frac{c}{p_{2,SAT}}$

Let $d \in \{0, 1\}$ be the decision of whether or not to apply, which relates to the latent variable d^* in the following way:

$$d(SAT, y, h, \epsilon, A) = 1 \text{ if only if}$$
$$d^*(SAT, y, h, \epsilon, A) \equiv \left[\begin{array}{c} p_{1,SAT}\left(f(SAT, y, h) + u_1^* - \epsilon\right)I(A = 1) \\ +p_{2,SAT}\left(f(SAT, y, h) + u_2^* - \epsilon\right)I(A = 2) \end{array} - c \right] > 0.$$

Let $u_{A,SAT} = u_A^* - \frac{c}{p_{A,SAT}}$. The model implies that the probability of observing the decision to

---

[80] The proof builds on Meijer and Ypma (2008), who show the identification for a mixture of two continuous univariate distributions that are normal.

[81] In my model, there are multiple colleges, but I also observe the whole application portfolio made by a student, which gives me more information to identify their preferences for multiple colleges.

[82] Given that net financial aid enters the utility function with coefficient one, the standard deviation of ϵ is identified from the variation in financial aid within (SAT, y) group. To simplify the notation, I will present the case where σ_ϵ is normalized to 1.

apply by someone with (SAT, y, h) is

$G(SAT, y, h)$

$$= \lambda\,(SAT, y)\,\Phi(f(SAT, y, h) + u_{1,SAT}) + (1 - \lambda\,(SAT, y))\Phi(f(SAT, y, h) + u_{2,SAT}). \tag{16}$$

Fix (SAT, y), (16) varies only with h, so we can suppress the dependence on (SAT, y), i.e., within a fixed (SAT, y)

$$G(h) = \lambda\Phi(f(h) + u_1) + (1 - \lambda)\Phi(f(h) + u_2). \tag{17}$$

The following theorem shows that for any given (SAT, y), $\lambda\,(SAT, y)$, $u_1^* - \frac{c}{p_{1,SAT}}$ and $u_2^* - \frac{c}{p_{2,SAT}}$ are identified.

Theorem 1 *Assume that 1)* $\lambda \in (0, 1)$, *2) there exists an open set* $H^* \subseteq H$ *such that for* $h \in H^*$, $f'(h) \neq 0$. *Then the parameters* $\theta = (\lambda, u_1, u_2)'$ *in* (17) *are locally identified from the observed application decisions.*

Proof. The proof draws on the well-known equivalence of local identification with positive definiteness of the information matrix. In the following, I will show the positive definiteness of the information matrix for model (17).

Step 1. Claim: *The information matrix* $I(\theta)$ *is positive definite if and only if there exist no* $w \neq 0$, *such that* $w'\frac{\partial G(h)}{\partial \theta} = 0$ *for all* h.

The log likelihood of an observation (y, h) is

$$L(\theta) = d \ln(G(h)) + (1 - d) \ln(1 - G(h)).$$

The score function is given by

$$\frac{\partial L}{\partial \theta} = \frac{d - G(h)}{G(h)(1 - G(h))} \frac{\partial G(h)}{\partial \theta}.$$

Hence, the information matrix is

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

Given $G(h) \in (0, 1)$, it is easy to show that the claim holds.

Step 2. Show $w'\frac{\partial G(h)}{\partial \theta} = 0$ for all $h \implies w = 0$.

$\frac{\partial G(h)}{\partial \theta}$ is given by:

$$\frac{\partial G(h)}{\partial \lambda} = \Phi(f(h) + u_1) - \Phi(f(h) + u_2)$$

$$\frac{\partial G(h)}{\partial u_1} = \lambda\phi(f(h) + u_1)$$

$$\frac{\partial G(h)}{\partial u_2} = (1 - \lambda)\phi(f(h) + u_2)$$

Suppose for some w, $w'\frac{\partial G(h)}{\partial \theta} = 0$ for all h:

$$w_1[\Phi(f(h) + u_1) - \Phi(f(h) + u_2)] + w_2\lambda\phi(f(h) + u_1) + w_3(1 - \lambda)\phi(f(h) + u_2) = 0$$

Take derivative with respect to h evaluated at some $h \in H^*$

$$w_1[\phi(f(h) + u_1) - \phi(f(h) + u_2)]f'(h) + w_2\lambda\phi'(f(h) + u_1)f'(h) \qquad (18)$$
$$+w_3(1 - \lambda)\phi'(f(h) + u_2)f'(h) = 0.$$

Define $\rho(h) = \frac{\phi(f(h) + u_1)}{\phi(f(h) + u_2)}$, divide (18) by $\phi(f(h) + u_2)$:

$$w_1[\rho(h) - 1] - w_2\lambda(f(h) + u_1)\rho(h) - w_3(1 - \lambda)(f(h) + u_2) = 0$$

$$\rho(h)[w_1 - w_2\lambda(f(h) + u_1)] - [w_1 + w_3(1 - \lambda)(f(h) + u_2)] = 0 \qquad (19)$$

Since $\rho(h)$ is a nontrivial exponential function of h, (19) hold for all $h \in H^*$ only if both terms in brackets are zero for each $h \in H^*$, i.e.

$$w_1 - w_2\lambda(f(h) + u_1) = 0 \qquad (20)$$
$$w_1 + w_3(1 - \lambda)(f(h) + u_2) = 0.$$

Take derivative of (20) again with respect to h, evaluated at $h \in H^*$:

$$w_2\lambda f'(h) = 0$$
$$w_3(1 - \lambda)f'(h) = 0.$$

Since $\lambda \in (0, 1)$ and $f'(h) \neq 0$ for some h, $w = 0$. $\blacksquare$

E1.2 Identification of $p_{A,SAT}$:

In the data, we observe the admission rates for students given their SAT, y and h, $\Pr(Admssion|SAT, y, h)$,

which is generated via the following equation:

$$\Pr(Admssion|SAT, y, h) = \frac{\left[\begin{array}{c} \lambda\left(SAT, y\right) p_{1,SAT} \Pr(\text{apply}|SAT, y, h, A = 1) \\ + \left(1 - \lambda\left(SAT, y\right)\right) p_{2,SAT} \Pr(\text{apply}|SAT, y, h, A = 2) \end{array} \right]}{\Pr(\text{apply}|SAT, y, h)}. \quad (21)$$

Given the identification of $\lambda\left(SAT, y\right)$ and $u_A = u_A^* - \frac{c}{p_{A,SAT}}$ as in E1.1, $\Pr(\text{apply}|SAT, y, h, A) = \Phi\left(f(h, SAT, y) + u_A\right)$ is also identified. The denominator is directly from the data. Define the following known objects

$$\begin{aligned} \Upsilon\left(SAT, y, h, A = 1\right) &= \frac{\lambda\left(SAT, y\right) \Phi\left(f(h, SAT, y) + u_1\right)}{\Pr(\text{apply}|SAT, y, h)}, \\ \Upsilon\left(SAT, y, h, A = 2\right) &= \frac{\left(1 - \lambda\left(SAT, y\right)\right) \Phi\left(f(h, SAT, y) + u_2\right)}{\Pr(\text{apply}|SAT, y, h)}. \end{aligned}$$

Equation (21) can be written as

$$\Pr(Admssion|SAT, y, h) = \sum_A \Upsilon\left(SAT, y, h, A\right) p_{A,SAT}$$

The variations in y and/or h move both $\Pr(Admssion|SAT, y, h)$ and $\Upsilon\left(SAT, y, h, A\right)$, which identifies $p_{A,SAT}$ for $A = 1, 2$.

E1.3 Identification of u_A^* **and** c

Recall that

$$u_{A,SAT} = u_A^* - \frac{c}{p_{A,SAT}}.$$

Given that $u_{A,SAT}$ is identified from E1.1 and $p_{A,SAT}$ from E1.2, fix A, both $u_{A,SAT}$ and $p_{A,SAT}$ vary with SAT, which identifies u_A^* and c.

E1.4 Relating the Identification to the Model

In practice, the variable y corresponds to family income (a 5-year average) and h other family background variables. The identification argument above assumes that 1) only SAT and family income enter the type distribution, i.e., SAT and family (permanent) income summarize all information that correlates with ability; and 2) the expected financial aid depends on SAT and all family-background variables. For example, conditional on family permanent income, family assets (which serves the role of the h variable in the previous proof) vary with factors, such as housing prices and stock prices, that are not correlated with ability.[83]

E2. Ability values ω :

---

[83] This exclusion restriction is sufficient but not necessary for identification. For example, I could allow family assets to enter type distribution as a categorical variable, and to enter the financial aid function as a continuous variable. The within-category variation in assets would be enough for identification.

ω is not point identified, even after normalizing ω_1. The reasoning is as follows: each college j faces discrete (s, SAT) groups of applicants and its admissions policy depends on the rankings of these groups in terms of their conditional expected abilities. These relative rankings remain unchanged for a range of ω's, as do colleges' decisions and the model implications. Knowing that ω is not point identified, I fix $\widehat{\omega} = [e, e^2, e^3]'$. At other values of ω around $[e, e^2, e^3]'$, the estimates for the other parameters in steps two and three will change accordingly. However, the counterfactual experiment results are robust.[84]

F. Tuition Elasticity

This exercise examines students' responsiveness to tuition changes. A counterfactual experiment based on the SPNE model is not directly comparable to previous studies because SPNE model endogenizes tuition and admissions policies. Instead, I simulated the student decision model, holding admissions probabilities at the baseline levels. In response to a $\$1,000$ (2003 dollars) tuition reduction, college enrollment will increase by 1%, which is lower than findings from previous studies.[85] The discrepancy may be explained by the different cohorts of students studied in various papers. This paper studies a cohort of students who enter college around 2002, much younger than those studied in previous literature. As shown in Figures 2 and 3, which are based on data from National Center for Education Statistics, state and local government spending on higher education has been rising over time, suggesting some expansion on the supply side. Such a pattern was accompanied by steady growth in college enrollment rate in earlier years. However, college attendance has been stagnant since 1998. These figures are consistent with the hypothesis that a lower fraction of students in the later cohorts are at the margin.[86]

G. Additional Tables
G1. Parameter Estimates

Table G1.1 Ordered Logit Ability Distribution

cut_1	cut_2	$y/1000$	$(y/1000)^2$	$SAT=2$	$SAT=3$
2.56	6.38	0.02	-0.00003	3.72	5.02
(0.21)	(0.37)	(0.005)	(0.00002)	(0.24)	(0.36)

cut_1, cut_2 are the cutoff parameters for the ordered logit.

---

[84]Results available upon request.

[85]For example, Leslie and Brinkman (1998) and Cameron and Heckman (2001).

[86]Although different, this hypothesis is not in conflict with Lochner and Monge-Naranjo (2011). In this hypothesis, there are more students from the later cohorts that are constrained, but instead of at college age, in their childhood when their pre-college human capital is formed. By college age, these "long-term-constrained" students are not at the margin and therefore are not easily attracted to colleges by moderate tuition reductions.

	$A = 1$	$A = 2$	$A = 3$
$\Pr\left(z = 1 \middle| A\right)$	1.0	0.75	0.70
	-	(0.04)	(0.05)

[a] $\Pr(z = 1 | A = 1) = 1$ cannot be rejected at 10% significance level.

[b] 86% of all students are of z-1 type.

Table G1.3 Ability Distribution

%	$A = 1$	$A = 2$	$A = 3$
All	48.6	38.8	13.2
$z = 1$	55.6	33.7	10.7
$z = 2$	0.0	71.0	29.0

Simulation based on the estimates in Tables G1.1 and G1.2.

Ability distribution among all students and by z type.

Table G1.4 Ability, SAT and Signals Among Applicants

	$A = 1$	$A = 2$	$A = 3$
% Applicants	3.5	68.5	28.0
	$SAT = 1$	$SAT = 2$	$SAT = 3$
% Applicants	15.4	60.4	24.2
	$s = 1$	$s = 2$	$s = 3$
% Applicants	8.6	73.2	18.2

Simulation based on the estimates in Tables G1.1 and G1.2.

Distribution among all applicants: each row adds up to 100%.

Table G1.5 Financial Aid

	General aid		College-Specific Aid	
	Coefficient	Std. Error	Coefficient	Std. Error
Constant	−6087.6	(848.2)	−12641.8	(2297.5)
Black	921.6	(927.5)	6774.1	(1819.9)
Family Income/1000	−34.7	(10.7)	−71.6	(16.5)
Family Assets/1000	−4.4	(2.5)	−5.9	(4.8)
$SAT=2$	2334.4	(987.8)	3747.9	(2250.9)
$SAT=3$	4366.2	(1240.8)	7342.0	(2498.1)
Sibling in College[a]	944.6	(832.1)	2958.2	(1751.9)
$(SAT=2) \times private$			12169.7	(1802.3)
$(SAT=3) \times private$			15130.8	(2391.4)
4-Yr Colleges	4123.0	(960.6)		
(pri,elite)			−11764.6	(4845.9)
2-Yr Colleges			−4281.5	(2077.3)
σ_η (aid shock)	498.0	(210.4)	6796.6	(193.6)
σ_{err} (measurement)	7047.3	(1284.0)	10466.0	(612.3)

a. Whether the student has some siblings in college at the time of application.

Table G1.6 Capacities (%)

κ_1	κ_2	κ_3	κ_4
1.1 (0.1)	8.2 (0.2)	10.8 (0.03)	20.0 (0.07)

Table G1.7 Tuition Weights

	(pri,elite)	(pub,elite)		(pri,non)	(pub,non)	
		In-State	Out-of-State		In-State	Out-of-State
m_{2j}	−0.0012	−0.0069	−0.0004	−0.0012	−0.0085	−0.0015
	(0.0002)	(0.002)	(0.0001)		(0.003)	(0.0009)
m_{1j}			0.0701 (0.002)			

Tuition is measured in thousands of dollars. m_{1j} are restricted to be the same across j's.

Allowing m_{2j} to differ between the two private college groups does not improve the fit.

G2. Model Fit

Table G2.1.1 Num. of Applications: by SAT

%	SAT=1*		SAT=2		SAT=3*	
n	Data	Model	Data	Model	Data	Model
0	85.1	86.3	24.9	24.4	13.0	12.4
1	12.1	8.3	45.2	44.1	43.0	57.6
2 or more	2.8	5.4	29.9	31.5	44.0	30.0

*$\chi^2 > \chi^2_{2,0.05}$

Table G2.1.2 Num. of Applications: by Family Inc

%	Income < Median*		Income $\geq$ Median	
n	Data	Model	Data	Model
0	68.2	68.4	40.2	40.5
1	22.1	19.0	33.9	36.5
2 or more	9.7	12.6	25.9	22.9

*$\chi^2 > \chi^2_{2,0.05}$

Table G2.2.2 Application Rates: by Family Inc

%	Income < Median		Income $\geq$ Median	
	Data	Model	Data	Model
(pri,elite)	7.6	5.8	10.8	11.4
(pub,elite)	21.8	20.8	37.2	32.6
(pri,non)	39.7	40.6	47.2	51.5
(pub,non)	76.0	72.3	69.1	65.1

Table G2.3.1 Admission Rates: by SAT

%	SAT=1		SAT=2		SAT=3	
	Data	Model	Data	Model	Data	Model
(pri,elite)	n/a	22.1	38.5	41.3	61.7	76.2
(pub,elite)	53.8	16.0*	80.0	91.0	92.8	98.0
(pri,non)	83.9	77.1	93.0	93.7	94.0	100.0
(pub,non)	83.9	83.1	95.0	98.5	99.0	100.0

* $\chi^2 > \chi^2_{1,0.05}$

Table G2.3.2 Admission Rates: by Family Inc

%	Income < Median		Income ≥ Median	
	Data	Model	Data	Model
(pri,elite)	55.0	51.1	52.8	60.6
(pub,elite)	91.2	86.9	80.3	92.6*
(pri,non)	86.5	87.8	93.5	92.9
(pub,non)	91.0	93.7	95.6	97.2

* $\chi^2 > \chi^2_{1,0.05}$

Table G2.4.1 Final Allocation: by SAT

%	SAT=1		SAT=2*		SAT=3	
	Data	Model	Data	Model	Data	Model
(pri,elite)	0.0	0.1	0.2	1.8	7.3	6.7
(pub,elite)	0.4	0.2	13.0	13.0	22.2	25.8
(pri,non)	4.3	3.4	16.9	18.7	25.6	18.2
(pub,non)	6.9	6.3	39.7	35.2	31.0	32.2
2-yr college	25.8	27.0	22.4	20.7	11.1	13.8
Outside	62.6	63.1	7.9	10.5	2.9	3.3

* $\chi^2 > \chi^2_{5,0.05}$

Table G2.4.2 Final Allocation: by Family Inc

%	Income < Median		Income ≥ Median*	
	Data	Model	Data	Model
(pri,elite)	0.6	0.6	1.3	2.5
(pub,elite)	3.9	3.5	11.5	12.5
(pri,non)	7.2	8.5	15.9	13.2
(pub,non)	15.8	14.4	28.0	25.8
2-yr college	22.7	25.0	22.7	21.1
Outside	49.8	48.0	20.5	24.9

* $\chi^2 > \chi^2_{5,0.05}$

Table G2.5.1 Home Bias: by SAT

(%)	SAT=1		SAT=2		SAT=3	
	Data	Model	Data	Model	Data	Model
Home-Only Applicants[a]	79.2	63.8*	68.0	68.6	50.3	62.6*
Home-State Attendees[b]	85.6	78.8	78.7	79.1	65.2	74.8

* $\chi^2 > \chi^2_{1,0.05}$

[a] % students who apply only within home states among all 4-yr applicants.

[b] % students who attend home-state 4-yr colleges among all 4-yr attendees.

Table G2.5.2 Home Bias: by Family Inc

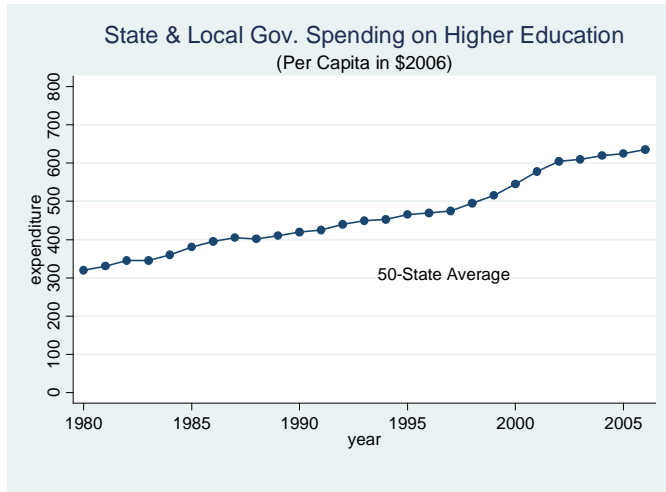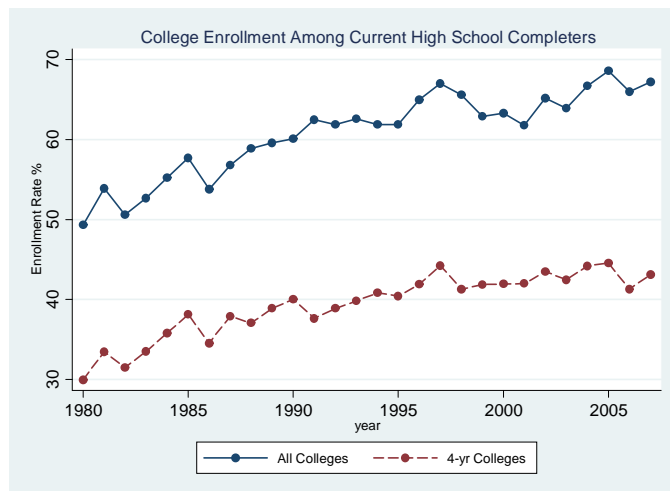| (%)	Income < Median		Income ≥ Median	
	Data	Model	Data	Model
Home-Only Applicants	72.1	67.5	62.2	67.5
Home-State Attendees	80.5	79.1	74.1	77.4

Figure 2: Government Spending on Higher Education

Figure 3: College Enrollment Trend