Competing Campuses:

An Equilibrium Model of the U.S. Higher Education Market

Emily E. Cook*

November 14, 2019

Abstract

Universities compete through price and non-price characteristics, including the number and type of undergraduate majors offered. Undergraduate majors are a key source of differentiation in U.S. higher education, and survey evidence suggests that students consider a university's offered majors when they make their college choice. To further an understanding of the effects of competition on prices, selectivity and majors offered, I answer three specific questions. First, to what extent do majors offered affect applications and enrollment? Second, what are the preferences and costs that drive universities' choice of price, selectivity, and majors offered? Finally, what are the effects of subsidies? I use state-level data from ACT and College Board and information on university characteristics to estimate a model that features four stages: first, universities choose majors, admission criteria and price; second, students choose an application portfolio; third, universities make offers of admission; and finally, students make enrollment decisions. The estimates show that students are willing to pay over $100 per year for each major, with heterogeneity by type of major. Universities place substantial weight on selectivity, and the cost of supplying different majors varies by type. In counterfactual simulations, I discuss the effects of a state-level grant for students to attend public universities.

JEL Codes: I23, I28, L13, L31

Keywords: College Choice, College Majors, College Admission, Non-profit Objective

*Department of Economics, University of Virginia, emilycook@virginia.edu. I thank Federico Ciliberto, Sarah Turner, Gaurab Aryal, Simon P. Anderson, Ekaterina Khmelnitskaya, Dan Savelle, Cailin Slattery and many other students and faculty at the University of Virginia for their advice and support. The work also benefited from input from participants in the International Industrial Organization Conference 2019 and the D.C. IO Day 2019. Ed Hall and the Advanced Research Computing Services at the University of Virginia provided computational resources and technical support. Victoria Roloff provided excellent research assistance. I also wish to acknowledge the Bankard Fund for Political Economy and the National Science Foundation for financial support. provision of data. The results contained herein were derived in part from data provided by ACT Research, and in part from data provided by the College Board. Copyright © 2000-2015 The College Board. www.collegeboard.org. All errors are mine.

1 Introduction

Survey evidence suggests that the menu of majors offered at a university is an important determinant of where students apply, and ultimately, enroll. Among students who were high school seniors in 2012, 74.9% indicated that the majors and programs offered would be a very important factor in their college choice.[1] One reason for this is that a student's choice of major may be limited by the selection of majors offered at their university. There are notable differences across universities in the majors that they offer—the average university offers only 55 majors out of approximately 1,500 possible undergraduate majors that are classified by the National Center for Education Statistics. The standard deviation of majors offered is 26.

Recognizing that the majors and programs offered may affect student demand, universities compete against each other through prices, selectivity, and the number and type of majors offered.[2] For example, when the University of Wisconsin at Stevens Point (UWSP) proposed to cut 13 majors (including French, German, history, geology, geography and art) and add new programs in high-demand majors like environmental science, the provost cited "rising competition among public and private universities" as one of the motivating factors (Strauss, 2018).[3]

This research studies competition between four-year universities in a context in which universities make decisions about prices, selectivity and the number and type of majors offered. Foundational to the study of competition in this market is a baseline understanding of universities' incentives. Thus the first step in this research is to establish the sign and magnitude of the effect of prices, selectivity, and majors offered on demand, conditional on other university characteristics. Given these demand estimates, I then estimate the preferences and costs that drive universities' choices. Finally, I use these results to study the equilibrium effects of subsidies in the form of financial aid.

The tool for analysis is an equilibrium model that features four stages: first, universities choose majors, admission criteria and price; second, high-school seniors observe universities' choices and choose a set of universities to apply to; third, universities offer admission to a subset of applicants; and finally, students make their enrollment decisions. Modeling the application, admission, and enrollment stages of the process produces college choice parameters that have a meaningful interpretation in an expected

[1] The survey evidence is from the National Center for Education Statistics' High School Longitudinal Study of 2009.

[2] Prices and selectivity are likely some of the most flexible choices of universities, so prices and admissions selectivity are endogenous responses in addition to the number and type of majors.

[3] UWSP ultimately decided to maintain these majors, saying that they had found other ways to cut costs.

utility-maximization framework. Modeling these three stages allows me to compute measures of selectivity and total enrollment and study how these separately enter the university's problem. The model of university behavior features a trade-off between selectivity and revenue net of costs, with heterogeneous preferences for selectivity. The setting is strategic because universities compete with each other for students, and generally in limited geographic areas.[4]

To estimate the model I use state-level data from ACT and College Board combined with information on universities from the Integrated Postsecondary Education Data System (IPEDS). IPEDS is a set of surveys of the population of universities in the United States, and is maintained by the National Center for Education Statistics. These surveys provide information on tuition and financial aid, location, and many other university characteristics. IPEDS also provides enrollment to each university from each state, which I use to construct state-level enrollment shares. The ACT and College Board data provide the number of students in each state who sent score reports to each university, which I use as a measure of applications. The application and enrollment data together allow estimation of student preferences. In addition to the state- and university-level data, individual student data from the NCES' High School Longitudinal Study of 2009 (HSLS:09) and the Education Longitudinal Study of 2002 (ELS:2002) inform pieces of the estimation procedure.

The paper innovates on both the demand- and supply-sides of the market. On the demand-side, the primary contribution is to demonstrate that variation in majors offered affects college choice. On the supply-side, I present a model that allows estimation of university-specific strategies. Incorporated in the model is the strategic and regional nature of competition in this market.

The college choice model follows several other structural models in its basic outline. Arcidiacono (2005), Howell (2010), and Kapor (2016) each model the application, admissions, and enrollment stages of the college choice problem in order to study affirmative action and related policies. Fu (2014) uses a structural model to study the effect of increasing college capacity on enrollment, and the effect of basing admission decisions only on test scores. In each of these papers, a student's application decision is modeled as a choice between application portfolios, where the value of an application portfolio depends upon the probability of admission and the expected value of enrollment.[5]

[4] Hoxby (1997) and Smith et al. (2018) both assert that universities compete for students, and that universities compete most strongly with universities that are located nearby.

[5] A multiple-discrete choice problem arises in the application stage because a student can choose more than one school to apply to at a time. Modeling the application stage as a choice between application portfolios converts the multiple-discrete choice problem to a single-discrete choice problem. This solution has been used in other contexts. For example, Gentzkow (2007) estimates demand

There are several key differences between my college choice framework and the models referenced above. First, enrollment utility is a function of majors offered. Second, the model takes into account the regional nature of markets; each student will choose from a set of universities that depends upon the student's location. Third, I utilize aggregate data and a simulated method of moments estimation strategy, which allows me to estimate demand for over 800 individual universities with a rich set of controls for university characteristics. The method is complementary to other methods with individual data, as those methods allow richer controls for individual characteristics.

Another way in which this college choice model differs from previous structural models is that the model implies correlation between universities' choices (prices, majors and admission thresholds) and university characteristics that are unobserved to the econometrician. This feature is typical of the extensive literature in Industrial Organization on static equilibrium models of oligopoly. The most influential paper in this literature, Berry et al. (1995) (BLP), developed an equilibrium model in which consumers make a single-discrete choice and firms set prices in Bertrand competition. Their paper provides an estimation strategy for handling correlation between unobserved firm quality and prices. While my model does not generate the same demand-side expressions, their estimation method is applicable in this context.[6]

This paper also contributes to a better understanding of the supply-side and equilibrium in higher education, which is a complex and underdeveloped area of research. Epple et al. (2006) and Fu (2014) provide the first empirical models of the higher education market. These models are insightful but have been limited by a need to aggregate universities at a high level. Unlike the previous equilibrium models, universities in the present model choose individual strategies and are differentiated along a variety of endogenous and exogenous dimensions.

The question of which university objective function to use is crucial. There is no single objective function that has been developed to capture all of the incentives for all activities a university engages in, yet there are reasonable ways to think about the tradeoffs involved in certain decisions. Typically, the literature assumes that monetary inputs and student ability enter the objective function. In Epple et al. (2006), universities maximize quality, which is a function of the average student ability, expenditures that

for newspapers, where consumers can choose to read more than one newspaper. In his model, consumers choose between "bundles" of the available newspapers. See his paper for a more thorough review of multiple-discrete choice models.

[6]While I borrow ideas from work on consumer product markets, notable differences between the higher education market and a typical consumer product market motivate substantial deviations from a typical model of Betrand competition with discrete choice demand. The institutional details that drive these differences include: 1) universities are not profit maximizing, 2) consumers (students) must apply and be accepted in order to "purchase" the product, and 3) universities charge one price across all markets (with the exception of public universities that charge an in-state and out-of-state price).

contribute to educational quality, and the average income of the student body. Each of these are assumed to enter a Cobb-Douglas quality function, the parameters of which are estimated. Average income was included so that the model would better reflect the observed degree of price discrimination. In Fu (2014), universities value net tuition and student ability.

The model proposed in Section 3 also combines selectivity and monetary considerations. Universities value their students' academic fit at the university, which is the student test score plus an idiosyncratic measure of the student's ability to succeed at the university. Universities also value revenue net of costs ("net revenue"). The relative value of each input is estimated, so one of the outcomes of this research is to determine the extent to which universities behave as selectivity or net-revenue maximizers.[7] In Section 3, I provide more detail regarding the choice of objective function.

The remaining sections of this paper are as follows: Section 2 describes the data sources and provides motivating empirical evidence. Section 3 outlines the model in several stages and derives equations that are later used in estimation. Sections 4 and 5 present the identification argument, empirical strategy and resulting model parameters. Section 6 presents the counterfactuals, and Section 7 concludes with a discussion of the limitations of this research and areas for future work.

2 Data

This section describes the data sources along with the empirical evidence that motivates the equilibrium model. In the first subsection, the objective is to describe the population of universities used in this study, with particular attention to the cross-sectional variation in majors offered, prices, and selectivity. In the second subsection, I define the regional structure of universities' markets. The third subsection describes the student data, both individual and aggregate.

2.1 University Data

The university characteristics come from the National Center for Education Statistics' (NCES) Integrated Postsecondary Education Data System (IPEDS) for the academic year 2013-14. IPEDS is a collection of university surveys on a wide range of topics such as admissions and enrollment, finances, financial aid, and

[7] Because this is a static framework, I can only estimate to how universities exhibit these preferences on a year-to-year basis. It may be that by remaining selective, universities increase their long-run revenue through gifts (Hoxby, 2009). This model will not be able to determine the extent of this dynamic incentive.

degrees granted. These surveys include the entire population of postsecondary institutions in the U.S. that participate in Federal student aid programs under Title IV (such as Pell Grants and Federal student loans).

Four-year public or private non-profit institutions located in the 48 contiguous states or the District of Columbia are the focus of this analysis. In the model and estimation, two-year colleges and for-profit institutions are included in the outside option, along with the labor market.[8] After imposing some restrictions on size and selectivity, the population for analysis includes 428 public universities and 466 private universities.[9] Enrollment at the institutions in this analysis makes up 79% of first-time enrollment of recent high school graduates at all public and private non-profit four-year institutions.

Information on offered majors is available from the IPEDS survey on degrees conferred. For each offered major and degree level, institutions report the number of degrees conferred in a given academic year.[10] A college major is defined using the NCES' 2010 Classification of Instructional Programs (CIP) taxonomy. This system classifies all major programs using a six-digit code, which represents a significant degree of disaggregation (see Appendix Table 1 for examples). Majors that are highly related share the first four digits of the six-digit code, and majors are further classified into broad areas using the first two digits. It is infeasible to model hundreds of these individual six-digit majors separately, so I combine CIP codes to form these groups: 1) Arts, Social Science and Humanities, 2) Business and Communications, 3) Engineering, Math, and Science, and 4) Education, Health, Criminal Justice, and Other.[11] I then measure majors offered using the total count of six-digit codes, and the proportion of majors that the university offers in each group. Appendix Table 2 shows the major groups, the total number of majors that exist in the CIP system within each group, and the two-digit CIP codes that were aggregated to generate each group.

The resulting population of universities and key variables including majors offered and tuition are summarized in Table 1. Public universities offer nearly 60 majors on average, while private universities offer approximately 51. One key ingredient in the model and estimation strategy—and indeed, one reason we might think that the number and type of majors offered might matter to students—is the heterogeneity

[8] While there is some substitution the for-profit four-year and traditional public and private non-profit four-year sectors, the substitution is limited. For-profit institutions comprised approximately 3 percent of first-time four-year enrollment among recent high school graduates in the 2013-14 academic year. Including two-year colleges as a separate option would be an interesting way to extend this model, which I leave for future research.

[9] I restrict analysis to institutions that are classified in the Carnegie Classification system as Baccalaureate, Masters, or Doctoral/Research level universities. Finally, I restrict the sample to institutions of sufficient size or selectivity: institutions included have either: total undergraduate enrollment greater than 2,000, a Barron's selectivity rank of "Competitive" or higher, or receive applications from more than 2% of students in at least one state.

[10] Institutions report a zero when the degree is offered but no degrees are conferred, and the data are reported as missing if the degree is not available.

[11] The groups were chosen to combine areas of study that are loosely similar in content.

Table 1: Summary of Key University Characteristics

	Public (In-State)		Private	
Variable	Mean	St. Dev.	Mean	St. Dev.
Number of Majors	59.85	29.55	51.27	21.29
Share Arts, Social Science and Humanities	0.37	0.12	0.47	0.15
Share Business and Communications	0.13	0.06	0.14	0.08
Share Engineering, Math, and Science	0.22	0.10	0.19	0.10
List Tuition and Fees, $1,000s$	8.45	2.47	32.83	8.87
Net Tuition and Fees, $1,000s$	2.71	2.80	14.33	7.17
FTE Undergrad Enrollment (1,000s)	11.45	8.56	3.33	3.36
25th Percentile ACT	20.10	2.67	22.21	3.69
N	428		466	

NOTE: The table summarizes several key characteristics of 4-year degree-granting private non-profit or public institutions of higher education in the U.S. that are included in this analysis. See Section 2 for the inclusion criteria. The share of majors offered by type is calculated as the number of majors the university offers in each category divided by the total number of majors offered at the university. Net tuition and fees is list tuition and fees minus average grants and aid (aid that is not required to be repaid by students). In the model and estimation, tuition for public universities will be state-residence specific, although only in-state tuition is summarized here. All monetary values are in 2015 $.

in the number and type of majors offered. The second column in Table 1 shows that the standard deviation of the number of majors offered is nearly 30 majors for public universities and 21 for public universities.

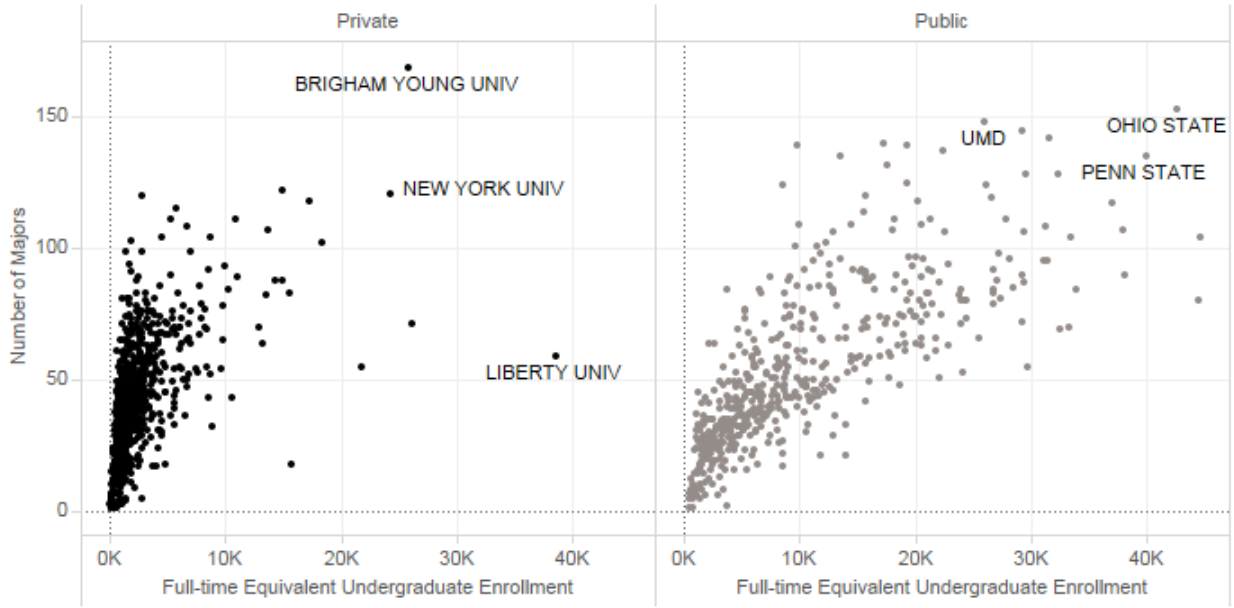
The primary source of price variation in the data is between the in-state public tuition rate and the out-of-state and private rates. Table 1 shows that average net tuition (tuition minus average grants and aid, excluding loans) is approximately $2,710 per year at in-state public universities and approximately $14,330 at private universities.

Universities are differentiated in other notable ways, including selectivity. The average of the 25th percentile ACT score within public universities is 20 (on a scale from 1 to 36), while in private universities it is 22. There is some variation within type, with a standard deviation of 2.7 for public universities and 3.7 for private universities.

Figure 1 shows that the number of majors offered is correlated with total enrollment. Because the number of majors offered is an equilibrium outcome, it is not clear at the outset to what extent the positive correlation with enrollment reflects demand- or supply-side factors. The model presented in this paper will generate the supply and demand functions that generate the equilibrium number of offered majors.

While Table 1 showed that private universities offer fewer majors on average, Figure 1 shows that the number of majors offered increases with enrollment more rapidly at private universities than at public

Figure 1: Number of Majors versus Enrollment



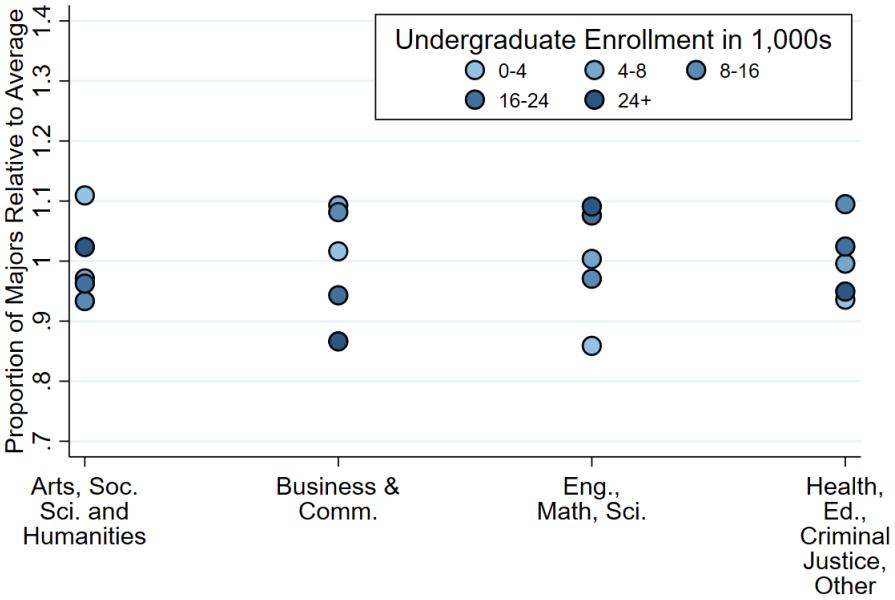
NOTE: This graph shows the total number of majors on the vertical axis and the total undergraduate enrollment on the horizontal axis. Public universities are in gray and private universities in black. This was generated using the 2013 IPEDS data.

universities. The fact that private universities offer more majors for a given level of enrollment suggests that public universities are able to obtain higher enrollment levels (through subsidized prices) without offering as many additional majors.

The share of majors offered within each category is also correlated with enrollment. Figure 2 shows, by undergraduate enrollment, the proportion of majors that fall into each category relative to the average. Enrollment is positively correlated with the relative share of majors in Engineering, Math and Science, while smaller universities tend to have the greatest proportion of majors offered in Arts, Social Science, and the Humanities. Universities with lower enrollment tend to have a greater share in Business, although this pattern does not hold true for the smallest institutions.

In summary, the descriptive analysis evidences substantial differentiation across institutions in net tuition, the number and type of majors offered, selectivity, and other factors. The differentiation in prices, majors offered, and selectivity is likely explained by market forces, as universities choose these in response to market conditions.

Figure 2: Type of Majors Offered within University, by Enrollment



NOTE: This graph shows the institution's proportion of majors offered in each category relative to the average proportion in each category. This was generated using the 2013-14 IPEDS data as described in the text.

2.2 Regional Competition

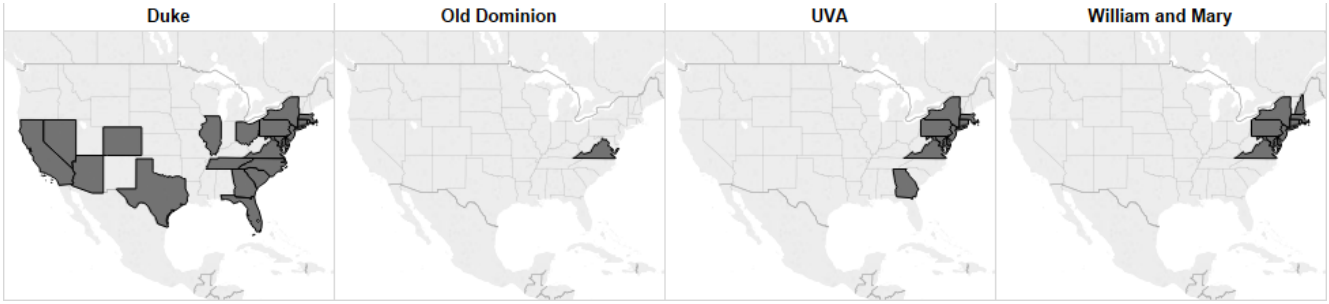
In principle, universities may compete for students across the U.S., but enrollment data by state shows that most universities draw the majority of their enrollment from only a few states. This point has been noted as a key feature of U.S. higher education by Hoxby (1997) and Smith et al. (2018). Importantly, the set of competitors that each institution faces differs depending upon where the institution is located and how widely it draws its enrollment.[12]

To capture the regional structure of competition in the model, I restrict universities to compete in a set of states defined by the empirical distribution of enrollment. This implies that students' choice sets are in part determined by their geographic location. To determine which universities compete in each state, I rank the states (within university) by the proportion of the state's high school graduates that attend that university.[13] Then, starting from the top of the list, I select the states that together account for at least 85%

[12] Another implication of the localized draw of enrollment is that universities face different pools of potential students, with differing distributions of income, race, and academic ability, as noted in Hoxby and Turner (2019).

[13] The total number of high school graduates by state (market size) is obtained from NCES data as collected in the *Knocking at the College Door* reported produced by the Western Interstate Commission for Higher Education (WICHE).

Figure 3: University Market Presence



NOTE: This figure shows the states in which each institution competes. To determine this set of states, states are ranked (within university) by the proportion of the state's high school graduates that attend that university. Then starting from the top of the list, states are included in the university's market until the included states make up 85% of the university's enrollment.

of the university's total enrollment. In these computations I used ten years of data (2005-06 through 2015-16) to smooth out year-to-year fluctuations. I further restrict the set of states a university competes in by requiring that at least 0.15% of students in the state apply to that university.

Figure 3 demonstrates the outcome of this method for the College of William and Mary, Duke University, Old Dominion University, and the University of Virginia. Duke is well-known nationwide and highly prestigious, so its enrollment pulls from a large number of states across the nation. The University of Virginia, while a public institution, is one of the top-ranked public institutions in the nation, so it also draws a large number of students from other states, particularly in the Northeast. The College of William and Mary, a small but highly-ranked college, draws most of its students from the Northeast as well, while Old Dominion University (a less selective public university located in Virginia) draws nearly all of its enrollment from Virginia. All four institutions are competitors in Virginia, while of these four, only Duke competes for students from Texas.

2.3 Student Data

Estimation of the college-choice model utilizes data on applications and enrollment by state. The IPEDS Residence and Migration Survey provides enrollment in each institution by student state of residence. I use these data to generate enrollment shares, defined as the enrollment in an institution as a proportion of the state's high school graduating class. A distribution of enrollment shares (across all states and institutions) is summarized in the second column of Appendix Table 4. The distribution of enrollment shares is heavily right-skewed, with the enrollment share reaching 1% only at the 90th percentile, and then 16% at the

maximum.

While IPEDS provides enrollment data by state, it provides no information on applications by state. Data from the SAT and the ACT admissions tests help fill this gap. When students take the SAT or ACT, they have the option to send their score reports to institutions to supplement their college applications. Because the colleges in my study require an SAT or ACT score report to accompany each application, these reports provides a proxy for applications.[14] The College Board and the ACT each provided a table showing the total score reports sent to each institution in the country, by student state of residence and graduation year. I link the data from the ACT and SAT by university name and then sum the total score reports across the two tests. An example of these data for Connecticut is presented in Table 3. The score-send data are summarized in the first column of Table 4. As is evident in this table, the distribution of score reports sent is also heavily right-skewed; on average, the top 10% of universities competing in each state receive 55% of all score reports sent.

The ACT and College Board also provide a distribution of test scores among test-takers in each state. Estimation of the model requires the distribution of test scores among all high school graduates. I estimate the test score distribution among all high school graduates utilizing variation in admission testing policies by state over time. The details are described in Appendix A.

In addition to these sources of aggregate data, I have access to a sample of high school students from the High School Longitudinal Study 2009 (HSLS:09). This survey provides information on the application portfolios, admission outcomes, enrollment decisions, and student characteristics for a sample of students who were freshmen in high school in 2009. I use this sample to estimate a parameter of the admissions framework. In addition, I utilize the Education Longitudinal Study of 2002 (ELS:02) for summary statistics that inform the estimation.

In the section that follows, I describe a model in which prices, the number and type of majors offered, and selectivity are determined in equilibrium. From the model I derive estimating equations that are matched to the aggregate data on score-sends and enrollment, the individual data on admissions outcomes conditional on application, and the university-level data on prices, admission selectivity, the number and type of majors offered, and other characteristics.

[14] Card and Krueger (2005) show a high correlation between test-score sending and applications, so the SAT and ACT report-sending data may be a close approximation for applications. The universities in my study require the ACT or SAT for admission.

3 Model

There are two types of agents in the model: high school graduating seniors (students) and universities. The timing of the model is illustrated in Figure 4. First, universities simultaneously choose the number of majors and the proportion in each subject category, the price ("net tuition," which is tuition minus average grants and aid), and an admission threshold, which is a minimum measure of a student's academic fit that the university will accept.

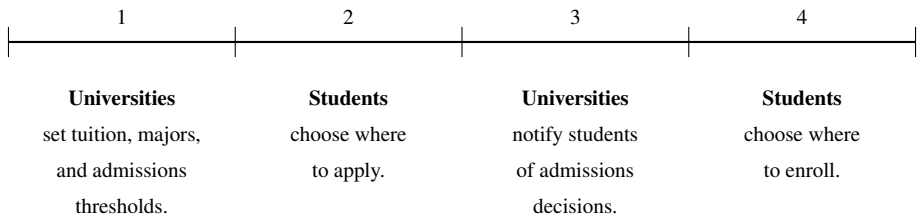
Students then observe the majors offered, prices, admission thresholds, and exogenous characteristics of institutions. Students do not observe their academic fit with each university, but each student knows his probability of admission given each university's admission threshold and his college admission test score. Each student chooses an application portfolio (a set of universities to which they apply—this can include no university) based on observable university characteristics, the student's economic cost of application, and his likelihood of admission.[15]

After applications are received, universities observe a measure of academic fit for each applicant, compare this measure to the admission threshold, and then notify each student of his admission outcome. When a student is admitted, he receives additional idiosyncratic information about his utility of attending a university. He then chooses a university to attend from the set of universities to which he was accepted, or he may choose not to attend any university.

Details on each stage follow below, in reverse chronological order. I start with the utility of enrollment, then present the admission mechanism and the value function at the application stage. Next I derive expressions for the university's share of applicants and enrollees, both of which will be used to estimate the demand parameters. I then turn to the supply side, presenting the university's objective function and first-order conditions for majors, net tuition, and admission thresholds.

[15] Note that there are several simplifications in the model thus far. First, net tuition for the upcoming year is typically announced after application deadlines. Essentially, I assume that students are able to make an accurate prediction of average net price for the upcoming year based on available information, such as last year's net price. Second, the admission criteria that an individual university uses are unobserved, and are potentially quite complex. However, test scores are likely the primary predictor of admission at the universities in this study, and good information is widely available on the average and percentiles of admission test scores within institution. I assume that students predict their admission probabilities based on this information and their own test scores.

Figure 4: Model Timing



3.1 Student Utility from Enrollment

At the application stage, students choose an application portfolio that maximizes their expected utility of enrollment net of application costs. Thus, the first step toward a fully specified model of college choice is to define the utility of enrolling in a university.

Let $l = 1...L$ denote the high school graduate's state of residence. There are $J + 1$ universities $j = 0, 1, ..., J$, where $j = 0$ represents the outside option of four-year colleges not included here (such as for-profit universities), two-year colleges, other educational certificate programs, the labor market, or other options. Each university offers a number of majors within each of K broad categories. Let $\mathbf{M_j}$ be a $K \times 1$ vector where the first element, M_{j1} is the total number of majors university j offers, and the remaining $K - 1$ elements are the proportion of majors in each group (with Education the excluded category). Universities also have a number of exogenous characteristics, $\mathbf{X_{jl}}$, such as whether the university is under private or public control and the distance to the population center of each state. Universities charge a net tuition, p_{jl}, which is tuition minus average grants and aid. For private universities, this price will be constant across l, while for public universities, the price varies by in-state status (whether l is the state where j is located).

The utility individual i receives from enrolling at university j is:

$$u_{ijl} = \beta_0 + \beta_1 p_{jl} + \boldsymbol{\beta_2}\mathbf{M_j} + \boldsymbol{\beta_3}\mathbf{X_{jl}} + \xi_{jl} + \eta_{ijl} \tag{1}$$

The student-specific error, η_{ijl}, is distributed Type I Extreme Value. This error represents information that becomes known to the student after their acceptance and that may have different value to different students (e.g., information about campus dining options presented during a campus tour). The university-specific

error ξ_{jl} represents the mean value of university characteristics unobserved to the researcher but known to students before application, such as a reputation for the quality of amenities like the gym, libraries, and dorms.

Let $\mathbf{D}$ be a $(J+1) \times 1$ vector, where $D_j = 1$ represents an admission offer at university j and $D_j = 0$ represents no admission offer, whether this is because the student did not apply or because the university rejected the student. Students may always choose not to attend any college ($D_0 = 1$), in which case they receive a mean utility which is normalized to zero.

Let $\nu_j = u_{ij} - \eta_{ij}$, with the state subscripts dropped for notational simplicity. Because of the Type I Extreme Value assumption on η_{ij}, the probability that a student enrolls in a particular university among universities that admitted the student is given by the usual conditional logit expression. If the university does not admit the student, then the probability of enrollment is zero.

$$\mathbb{P}(Enroll_{ij}|\mathbf{D}) = \begin{cases} \frac{e^{\nu_{ij}}}{\sum_{j' \ni D_{j'}=1} e^{\nu_{j'}}}, & \text{if } \mathbf{D}_j = 1 \\ 0, & \text{if } \mathbf{D}_j = 0. \end{cases} \tag{2}$$

3.2 Admission Probabilities

At the admission stage (Stage 3 in Figure 4), universities observe each student's academic fit and compare this value to their pre-determined admission threshold. Universities then inform each student about his admission outcome. In this section I describe the structure of the admission process, which will give rise to a probability of admission for each student. This probability will enter the application value function in Stage 3.

Each student has a test score, s_i, which is known to the student and is observed by the university at the admission stage. Prior to the admission stage, the university knows the distribution of test scores among all high school graduates. At the admission stage, the university also sees an idiosyncratic measure of a student's academic fit, ζ_{ij}, which is distributed $N(0, \sigma^2)$. This value captures idiosyncratic components of a student's fit at the university, which an admissions officer may infer from recommendations, essays or interviews. Students know only its distribution. Universities see only this idiosyncratic measure for their own applicants. Furthermore, universities do not know how many or which universities have received applications from each student.

The admission decision is based upon an admission index, which is the sum of the academic fit and the observed test score: $s_{ij} = s_i + \zeta_{ij}$. Universities compare this value to their admission rule, which is a threshold, $\underline{s}_j$.[16] Given the admission threshold, the student's score, and the distribution of the academic fit, students compute their probability of admission to each university. This is:

$$\mathscr{P}(D_j = 1|s_i, \underline{s}_j) = Prob(s_i + \zeta_{ij} > \underline{s}_j) = \Phi((s_i - \underline{s}_j)/\sigma), \tag{3}$$

where $\Phi(\cdot)$ is the standard normal CDF. This will be the first estimating equation, from which I will estimate σ.

Let $\mathbf{Y}$ be a vector indicating the student's application portfolio. $\mathbf{Y}$ is a $(J+1) \times 1$ vector where the element $Y_j = 1$ if the student applies to school j and 0 otherwise. The student always "applies" to the outside option $j = 0$. Given the admission probabilities and the assumption of independent admission outcomes conditional on test scores, students have the following probability of being admitted to a set $\mathbf{D}$ conditional on their academic skill and application portfolio:

$$\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) = \prod_{j \ni Y_j=1,\, D_j=1} \mathscr{P}(D_j = 1|s_i, \underline{s}_j) \prod_{j \ni Y_j=1,\, D_j=0} [1 - \mathscr{P}(D_j = 1|s_i, \underline{s}_j)]. \tag{4}$$

3.3 Application Value Function

Each student chooses an application portfolio to maximize his expected utility of enrollment net of application costs. To derive this expected value, we can first find the utility i receives from being accepted to $\mathbf{D}$. This is the expected value of the maximum utility among the universities represented by $\mathbf{D}$, which can be expressed using the log-sum formula, as I have assumed the η_{ij} are Type-I Extreme Value (with location and scale parameters normalized to 0 and 1, respectively):

$$v(\mathbf{D}) = ln(\sum_{j\ s.t.\ D_j=1} e^{\nu_{ij}}) + \gamma, \tag{5}$$

where γ is the Euler–Mascheroni constant. This constant cancels out of the choice probabilities as it does not change the relative utility of any option, and will be dropped in the remaining exposition.

[16] While one can conceive of admission rules that are not thresholds (such as a lottery among applicants), I assume that universities must set thresholds. The higher education industry is always under intense public scrutiny, and the threshold rule is consistent with public perception that university admission in the U.S. should be a meritocratic process.

The value of an application portfolio is the expected utility of enrollment given the probability of acceptance at each university, net of cost of application. As discussed in the previous section, the probability of being admitted to $\mathbf{D}$, conditional on the student's skill level, s_i, and application portfolio, $\mathbf{Y}$, is denoted by $\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y})$.

Then the expected value of an application portfolio Y for individual i is:

$$V(\mathbf{Y}|s_i) = \sum_{\mathbf{D}\subseteq\mathbf{Y}} (\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y})v(\mathbf{D})) - c(\mathbf{Y}) + \epsilon_{i\mathbf{Y}}, \tag{6}$$

where $c(\mathbf{Y}) - \epsilon_{i\mathbf{Y}}$ represents i's cost of application to $\mathbf{Y}$. $\epsilon_{\mathbf{Y_i}}$ is a Type I Extreme Value error that captures unobserved elements of the cost of application. This unobserved component could include application completion time associated with different application procedures across schools (everything from different log-in requirements for the online application system to different application essay topics or lengths).[17]

Based on the individual utility of enrollment and value for each application portfolio, the next step is to generate expressions for market-level statistics that can be used to estimate the parameters with aggregate data. In the next section I derive the two estimating equations for the demand-side parameters.

3.4 Aggregate Applications and Enrollment

In this section, I derive two expressions that are matched to the data by market. The first is the market penetration of each university, defined as the proportion of high school graduates in each state who apply to the university. The second expression is the analogous enrollment share by market.

Each student faces a unique set of application portfolios from which to choose (choice set), which I denote Υ_i. The empirical specification of these choice sets is in part determined by the student's geographic location, and is in part random. The limitations on choice sets are discussed in further detail in Section 5.1.

Conditional on the choice set Υ_i and the student's test score s_i, the probability that an individual applies to an application portfolio $\mathbf{Y}$ is:

$$\mathbb{P}(Apply_{\mathbf{Y}}|s_i, \Upsilon_i) = \frac{e^{\tilde{V}(\mathbf{Y}|s_i)}}{\sum_{\mathbf{Y}'\in\Upsilon_i} e^{\tilde{V}(\mathbf{Y}|s_i)}}, \tag{7}$$

[17] I assume this cost error is uncorrelated with the individual preferences for specific universities. I also assume it is independent across application sets. The latter is a weakness that will be addressed in future work through a Generalized Extreme Value assumption on the error term, as in (Bresnahan et al., 1997) and (Arcidiacono, 2005).

where $\tilde{V}(\mathbf{Y}|s_i) = V_i(\mathbf{Y}|s_i) - \epsilon_{i\mathbf{Y}}$. Summing this expression across all application portfolios that include j gives the total probability that i applies to university j:

$$\mathbb{P}(Apply_j|s_i, \Upsilon_i) = \frac{\sum_{\mathbf{Y} \ni Y_j=1} e^{\tilde{V}(\mathbf{Y}|s_i)}}{\sum_{\mathbf{Y}\in\Upsilon_i} e^{\tilde{V}(\mathbf{Y}|s_i)}}. \tag{8}$$

Aggregating to the market-level provides the market penetration function for j:

$$a_j = \mathbb{P}(Apply_j) = \sum_s \sum_\Upsilon \sum_{\mathbf{Y}\in\Upsilon\ni Y_j=1} \mathbb{P}(Apply_{\mathbf{Y}}|s, \Upsilon)P(\Upsilon)P(s) \tag{9}$$

Equation 9 becomes the second estimating equation. The model market penetration, a_j, will be matched to the observed proportion of students who apply to each university, A_j, at the market level.

Here we can observe a key distinction between this multiple discrete choice framework and the models of demand typical of the single discrete choice literature following Berry et al. (1995). The sum of the market penetration of applications, $\sum_j a_j$, yields a number greater than one, because each application portfolio will be counted as many times as universities it contains. In other words, if a student applies to three universities, that student is counted in the market penetration of applications for each of the three universities.

The second equation used in estimation is the first-year enrollment share within each market. For an individual, the probability of enrolling in university j, unconditional on admission decisions, is:

$$\mathbb{P}(Enroll_j|s_i, \Upsilon_i) = \sum_{\mathbf{Y}\in\Upsilon_i\ni Y_j=1} \mathbb{P}(Apply_{\mathbf{Y}}|s_i, \Upsilon_i) \sum_{\mathbf{D}\in\mathbf{Y}} \mathscr{P}(\mathbf{D}|s_i, \mathbf{Y})\mathbb{P}(Enroll_j|\mathbf{D}). \tag{10}$$

Each component of this expression has been described before. They are, respectively, the probability of application to a specific application portfolio conditional on the student's test score and choice set (equation 7), the probability of acceptance to a set of universities conditional on the student's applications and test score, (equation 4), and the probability of enrollment conditional on the student's admissions outcomes (equation 2).

Aggregating to the market-level yields the following expression for the enrollment share, e_j, which is

the third estimating equation:

$$e_j = \mathbb{P}(Enroll_j) = \sum_s \sum_\Upsilon \mathbb{P}(Enroll_j|s, \Upsilon)P(\Upsilon)P(s). \tag{11}$$

In this section so far, I have described the college choice model and admissions mechanism and derived the estimating equations for the college choice and admissions parameters. To recap, the estimating equations are: 1) The probit expression for each student's admission probability in Equation 3, 2) the market penetration function (or applicant share) in Equation 9, and 3) the market-level enrollment share in Equation 11. In the next two subsections, I turn to the supply-side. I first present the university objective function and then the first-order conditions for the setting of majors offered, net tuition, and admission thresholds.

3.5 The University Objective Function

A university is a complex institution, with potentially many factors in an objective function that is not well-understood. The universities in this study are not for-profit institutions, so the standard models of firm behavior may be of limited applicability. This research moves toward a better understanding of university behavior, although I do not claim to have modeled every aspect of a university's operations.

In this model, universities choose majors, net tuition, and admission thresholds to maximize their objective subject to student and competitor behavior. Both public and private universities potentially value selectivity and revenue net of cost of providing undergraduate education ("net revenue"). Selectivity is measured as the average student's rank among all high school graduates. The inclusion of the selectivity term is required to reflect the fact that universities reject students — if universities only valued profit, prices would adjust upwards and no student would be rejected. The inclusion of net revenue in the objective captures the idea that universities may value having financial resources which can be used to support research, graduate education, or other programs.

Universities make their decisions before the heterogeneous components of application costs, academic fit, and enrollment utility are realized, thus the enrollment and the average student's preparation are in expectation. They rank the expected average student preparation, compute profit based on the expected enrollment, and then weight the two inputs.

There are three distinctions between public and private universities in this model. First, public

universities may put different weight on selectivity relative to profit. Second, public universities may have different costs, and third, they have the option of charging different in-state and out-of-state prices. Public universities typically have much lower in-state prices than out-of-state prices as a result of their interaction with state legislatures, which provide appropriations to support the low in-state rates. While I do not model the state legislature explicitly, public universities will internalize legislative preferences through a term capturing an additional value for enrolling an in-state student.

3.5.1 Private University Objective

Private universities choose prices, majors, and an admission threshold to maximize a Cobb-Douglass in selectivity and net revenue:

$$\max_{p_j, \mathbf{M_j}, \underline{s}_j} \tilde{s}(\underline{s}_j)^{\lambda_j} (\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j}))^{1-\lambda_j}, \tag{12}$$

where $\tilde{s}(\underline{s}_j)$ is the ranking (by admission index) of the average student. $\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})$ is expected revenue minus economic cost. The university-specific λ_j takes the form: $\lambda_j = \lambda_0^{priv} + \epsilon_j^s$ if j is a private university, so λ_0^{priv} is a parameter governing the average private university's preference for selectivity.

In the following paragraphs I define the elements of the university objective more precisely. As a preliminary, I define the enrollment of the university. Section 3.4 provides an equation for the enrollment share in each market ($\mathbb{P}(Enroll_{jl})$) as a function of the university's choice of prices, admission thresholds and offered majors. Multiplying by N_l (the market size) gives the enrollment for each institution from each market:

$$q_{jl} = N_l \mathbb{P}(Enroll_{jl}).$$

Summing q_{jl} over all markets gives the total first-year enrollment, q_j:

$$q_j = \sum_l N_l \mathbb{P}(Enroll_{jl}).$$

Revenue per student includes the net tuition p_j, financial aid paid by sources outside the university, and non-tuition revenues such as private gifts. Per-student financial aid from outside sources will be denoted with t_j. Other revenue sources such as private gifts, ψ_j, will be assumed to scale with enrollment. Then the

private institution's revenue can be written as:

$$\mathcal{R}_j(p_j) = (p_j + t_j + \psi_j)q_j$$

Cost is a function of majors offered and enrollment, with a flexible form that allows for non-linearity in both majors and enrollment:

$$\mathcal{C}(\mathbf{M_j}) = q_j(\alpha_0^{priv} + \boldsymbol{\alpha}_1^{priv}\mathbf{M_j} + \alpha_2^{priv}q_j + \boldsymbol{\alpha}_3^{priv}q_j\mathbf{M_j} + \boldsymbol{\alpha}_4^{priv}\mathbf{M_j}^2 + \omega_j).$$

I let $\omega_j = \omega_{0j} + \sum_{k=1}^{K} M_{jk}\omega_{kj}$. Thus marginal cost is heterogeneous, and may depend on the offered majors. This term may reflect cost differences due to differences in quality across universities.

The ranking of the average student by admission index, $\tilde{s}_j(\underline{s}_j)$, is constructed in two steps as illustrated in Figure 5. First, I find the model prediction for the average of the admission scores ($s_{ij} = s_i + \eta_{ij}$) among enrollees. This average depends on the admission threshold $\underline{s}_j$, but also on the demand system that generates the distribution of scores among enrollees. This is the $\bar{s}_j$ in the figure.

Then this average score is ranked against the distribution of all students' admission scores. The distribution of the admission scores is normal, so the CDF of the admission test score yields the percentile rank of the average student.

3.5.2 Public University Objective

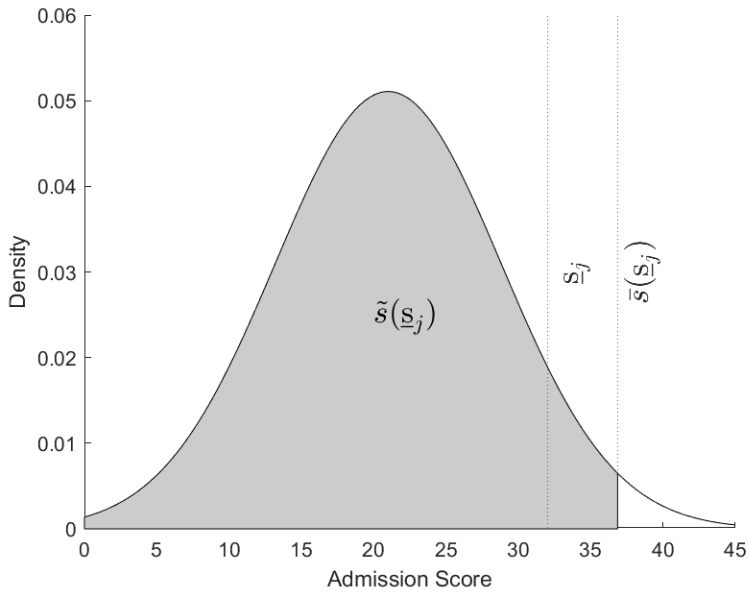
The objective for public universities is very similar to that of private universities in form, except that public universities separately choose in-state and out-of-state prices. The parameters of the model are also allowed to differ, so public universities may have different costs and preference for selectivity. The public university objective is:

$$\max_{p_j^{is}, p_j^{os}, \mathbf{M_j}, \underline{\mathbf{S}}_j} \tilde{s}(\underline{s}_j)^{\lambda_j}(\mathcal{R}(p_j^{is}, p_j^{os}) - \mathcal{C}(\mathbf{M_j}))^{1-\lambda_j}. \tag{13}$$

The average public university preference for selectivity is allowed to differ from that of private universities, so $\lambda_j = \lambda_0^{pub} + \epsilon_j^s$ for a public university.

For public universities, revenue depends upon both the in-state and out-of-state prices, p_j^{is} and p_j^{os} and in-state and out-of-state enrollments, q_j^{is} and q_j^{os}. Financial aid is assumed to be the same for in-state and

Figure 5: Calculating $\tilde{s}(\underline{s}_j)$



NOTE: This figure illustrates the calculation of the percentile rank of the average student, as described in the text.

out-of-state students, as data limitations prevent making a distinction. Appropriations are assumed to be a per-student rate for in-state students only, which creates a distinction between ψ_j^{is} and ψ_j^{os}.

$$R_j(p_j^{is}, p_j^{os}) = (p_j^{is} + t_j + \psi_j^{is})q_j^{is} + (p_j^{os} + t_j + \psi_j^{os})q_j^{os}.$$

The economic cost associated with each student is allowed to vary between in-state and out-of-state students for public universities, as they may have a mandate to serve in-state students. The cost function for public universities is as follows:

$$\mathcal{C}(\mathbf{M_j}) = q_j(\alpha_0^{pub} + \boldsymbol{\alpha}_1^{pub}\mathbf{M_j} + \alpha_2^{pub}q_j + \boldsymbol{\alpha}_3^{pub}\mathbf{M_j}q_j + \boldsymbol{\alpha}_4^{pub}\mathbf{M_j}^2 + \omega_j) + q_j^{is}(\alpha_5^{pub} + \omega_{0j}^{is}),$$

so that $\alpha_5^{pub} + \omega_{0j}^{is}$ reflects the extent to which public universities value in-state enrollment relative to out-of-state enrollment.

Given these objectives, public and private universities will simultaneously set majors offered, prices, and admission thresholds. The next section derives the first-order conditions to provide some economic intuition and to prepare for the identification and estimation of the model.

3.6 University First-Order Conditions

Universities set majors, prices, and admission thresholds simultaneously in order to maximize their objectives. Thus, observed choices are assumed to be solutions to the first-order conditions derived from the model. In this section I derive the first-order conditions, first for private and then public universities.

3.6.1 Private University First-Order Conditions

The pricing condition for a private university is:

$$\underbrace{\frac{\lambda_j}{(1-\lambda_j)}\frac{\mathcal{R}(p_j)-\mathcal{C}(\mathbf{M_j})}{\tilde{s}(\underline{\mathbf{s}}_j)}}_{MRS} = -\underbrace{\frac{\frac{\partial\mathcal{R}(p_j)}{\partial p_j}-\frac{\partial\mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial p_j}}{\frac{\partial\tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j}}}_{\substack{\text{Slope of Net Revenue/}\\ \text{Selectivity Frontier}}}.$$

The left hand side is the marginal rate of substitution between selectivity and net revenue. In a standard model of consumer behavior or of production, the marginal rate of substitution between two goods would be balanced with a price ratio derived from a budget constraint. Here, the "budget constraint" is a relationship between selectivity and net revenue, which I will refer to as the "Net Revenue-Selectivity Frontier." This is an implicit relationship between selectivity and net revenue that derives from the demand-side framework. As prices change, selectivity is affected through demand, with differential effects across the student test score distribution. Prices also affect net revenue directly and indirectly through changes in enrollment. This relationship traces out a feasible combination of selectivity and net revenue, which differs by university depending upon the university's characteristics and market conditions. The net revenue derivative with respect to price is:

$$\frac{\partial\mathcal{R}(p_j)}{\partial p_j}-\frac{\partial\mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial p_j} = (p_j+t_j+\psi_j)\frac{\partial q_j}{\partial p_j}+q_j-\frac{\partial q_j}{\partial p_j}\left(c_j(\mathbf{M_j})+q_j(\alpha_2^{priv}+\boldsymbol{\alpha}_3^{priv}\mathbf{M_j})\right),$$

where I let $c_j(\mathbf{M_j}) = \alpha_0^{priv}+\boldsymbol{\alpha}_1^{priv}\mathbf{M_j}+\alpha_2^{priv}q_j+\boldsymbol{\alpha}_3^{priv}\mathbf{M_j}q_j+\boldsymbol{\alpha}_4^{priv}\mathbf{M_j}^2+\omega_j^s$ represent the cost per student. The derivatives $\frac{\partial q_j}{\partial p_j}$ and $\frac{\partial\tilde{s}_j(\underline{\mathbf{s}}_j)}{\partial p_j}$ are computed numerically.

In addition to price, universities also choose an admission threshold and the K variables that measure the number and type of majors offered. All of the derivatives have a similar form from the Cobb-Douglas

objective function. The admission threshold first-order condition for a private university is:

$$\underbrace{\frac{\lambda_j}{(1-\lambda_j)}\frac{\mathcal{R}(p_j)-\mathcal{C}(\mathbf{M_j})}{\tilde{s}(\underline{\mathrm{s}}_j)}}_{MRS} = -\underbrace{\frac{\frac{\partial\mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathrm{s}}_j} - \frac{\partial\mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathrm{s}}_j}}{\frac{\partial\tilde{s}(\underline{\mathrm{s}}_j)}{\partial \underline{\mathrm{s}}_j}}}_{\substack{\text{Slope of Net Revenue/}\\ \text{Selectivity Frontier}}}.$$

When setting the admission threshold the university trades off the effect of the admission threshold on net revenue with the effect on selectivity. Selectivity is increasing in the admission threshold, while net revenue is decreasing in the admission threshold at the solution.

Figure 6 provides an illustration for intuition. The figure is generated using an example university and the estimated coefficients, but is meant only to be illustrative of the mechanics of the model. The solid green lines show the university's indifference curves between selectivity and net revenue. A university's objective is increasing as indifference curves shift towards the top-right. The optimal admission threshold is at the point of tangency between these indifference curves and the selectivity/net-revenue frontier, which gives the implied selectivity and net-revenue generated from the demand-side framework at each level of the admission threshold. In this graph, prices and majors are held constant.

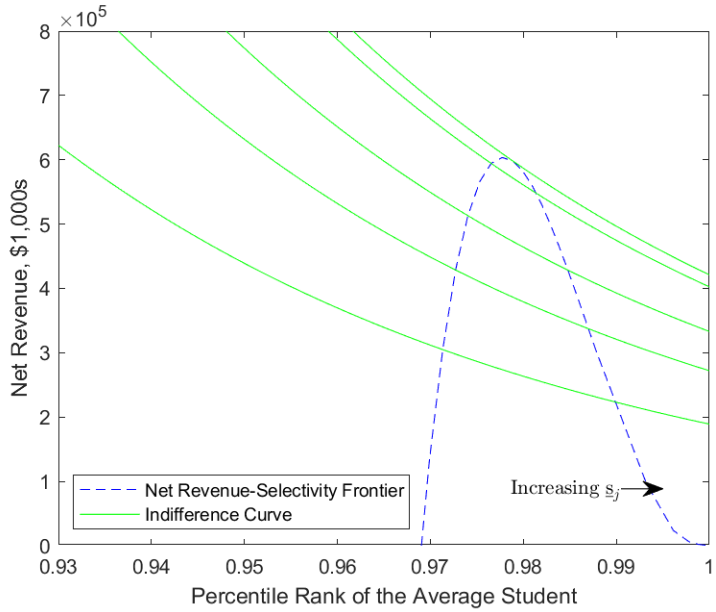
At an infinitely high admission threshold, the university admits no students and net revenue is zero. This is the bottom-right corner of the graph. As the admission threshold decreases, net revenue increases because the university admits more students, but selectivity decreases. At some point, the marginal cost per student increases to the point where an additional student actually reduces the net revenue and the selectivity/net-revenue frontier declines. The university's relative preference for selectivity will determine how close or far the optimal admission threshold is from the net-revenue maximizing point. Provided the preference for selectivity is positive, the optimal admission threshold is higher than the net-revenue maximizing point.

The derivative of revenue with respect to the admission threshold is negative, because increasing the admission threshold can only decrease the enrollment:

$$\frac{\partial\mathcal{R}_j(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathrm{s}}_j} = (p_j + t_j + \psi_j)\frac{\partial q_j}{\partial \underline{\mathrm{s}}_j}.$$

The derivative of cost with respect to the admission threshold is similar to the derivative with respect to

Figure 6: Optimization: Admission Threshold



NOTE: This figure illustrates the optimization over the admission threshold for an example university. The solid lines show the indifference curves between selectivity and net revenue. The dashed line shows the possible combinations of selectivity and net revenue as the admission threshold increases, keeping prices and majors offered fixed. The optimum is found at the point of tangency.

price, because the admission threshold and the price both affect cost only through the level of enrollment:

$$\frac{\partial \mathcal{C}_j(\mathbf{M_j})}{\partial \underline{s}_j} = \frac{\partial q_j}{\partial \underline{s}_j}(c_j(\mathbf{M_j}) + q_j(\alpha_2^{priv} + \boldsymbol{\alpha}_3^{priv}\mathbf{M_j})).$$

Finally, I find the first-order condition for a given k:

$$\underbrace{\frac{\lambda_j}{(1-\lambda_j)}\frac{\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})}{\tilde{s}(\underline{s}_j)}}_{MRS} = \underbrace{-\frac{\frac{\partial \mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial M_{jk}} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial M_{jk}}}{\frac{\partial \tilde{s}(\underline{s}_j)}{\partial M_{jk}}}}_{\substack{\text{Slope of Net Revenue/}\\ \text{Selectivity Frontier}}}.$$

As with the other conditions, the university trades sets the slope of the net revenue/selectivity frontier with respect to M_{jk} equal to the marginal rate of substitution between selectivity and net revenue. The derivatives of revenue with respect to the measures of the number and type of majors offered are:

$$\frac{\partial R_j(p_j)}{\partial \underline{s}_j} = (p_j + t_j + \psi_j)\frac{\partial q_j}{\partial M_{jk}},$$

so majors affect revenue only through enrollment. The derivatives of cost with respect to the M_{jk} include the cost of offering the majors and the effect on enrollment:

$$\frac{\partial \mathcal{C}_j(\mathbf{M_j})}{\partial M_{jk}} = \frac{\partial q_j}{\partial M_{jk}} c_j(\mathbf{M_j}) + q_j(\alpha_{1k}^{priv} + \alpha_{2k}\frac{\partial q_j}{\partial M_{jk}} + \alpha_{3k}^{priv} q_j + \frac{\partial q_j}{\partial M_{jk}}\alpha_{\mathbf{3}}\mathbf{M_j} + 2\alpha_{4k}^{priv} M_{jk} + \omega_{jk}).$$

Some algebra produces meaningful expressions for the markup (price minus marginal cost), the λ_j, and the supply of majors. First, for private universities, I solve the price and admission threshold first-order conditions to obtain an expression for marginal cost:

$$c_j(\mathbf{M_j}) + q_j(\alpha_2^{priv} + \boldsymbol{\alpha}_3^{priv}\mathbf{M_j}) = p_j + t_j + \psi_j + q_j \frac{\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}}{\left(\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}\frac{\partial q_j}{\partial p_j} - \frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial p_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}\right)}.$$

This expression shows that the markup (difference between marginal cost and price) is a function of both the price effect on selectivity and the price effect on the enrollment level. If the effect of price on selectivity is zero $\left(\frac{\tilde{s}(\underline{\mathbf{S}}_j)}{\partial p_j} = 0\right)$, then the markup is equivalent to that of a profit-maximizing firm in a standard model of oligopoly.

Next I take the ratio of the major supply first-order conditions to the pricing first-order conditions and obtain this expression for each k:

$$M_{jk} = \frac{(-\frac{\beta_{2k}}{\beta_1} - \alpha_1^{priv})}{\alpha_{4k}^{priv}} - \frac{\alpha_{3k}^{priv}}{\alpha_{4k}^{priv}} q_j - \frac{\omega_{jk}}{\alpha_{4k}^{priv}}. \tag{14}$$

So the level of each element of $\mathbf{M_j}$ is a linear function of enrollment and the universities' heterogeneous cost for each major.

Finally, an expression for λ_j can be obtained from the first-order condition for the admission threshold:

$$\frac{\frac{\frac{\partial \mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})}}{\left(\frac{\frac{\partial \mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})} - \frac{\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}}{\tilde{s}(\underline{\mathbf{S}}_j)}\right)} = \lambda_j. \tag{15}$$

3.6.2 Public University First-Order Conditions

Similarly to the private universities, a public university has first-order conditions for prices, admission thresholds, and each measure of the number and type of majors offered. One distinction between public and private universities is that the public universities have two pricing first-order conditions, one for the in-state price and one for the out-of-state price. For each price, the university trades off the effect on selectivity with the effect on net revenue.

Just as for private universities, the public universities' first-order conditions can be solved to produce a series of equations that are linear in the error terms. They are significantly more complicated because of the distinction between in-state and out-of-state rates, and are not significantly more instructive. I include these equations in Appendix D.

In this section, I presented a the model of the supply-side of higher education that features a trade-off between net revenue and selectivity. Universities choose prices, an admission threshold, and the number and type of majors offered. Universities have heterogeneous preferences for selectivity and heterogeneous costs, and public and private universities may differ in average preference for selectivity and average costs. Public universities are allowed to choose different in-state and out-of-state prices, and they internalize the government's preference for in-state students. In the next section, I describe how the parameters of the admission mechanism, enrollment utility and application cost, and the university objective are identified.

4 Identification

The parameters of the model include the variance of the student-university match value in the model of admissions, the enrollment utility parameters and the application cost parameters in the college choice problem, and the cost and preference parameters for universities. In the following subsections, I explain how each parameter is identified. The order of the presentation foreshadows the estimation strategy.

4.1 Identification: Admissions

The variance of the student-university match value, σ, governs how informative test scores are in the admission process. Recall from the model that the likelihood of admission to a university conditional on the student's test score s_i and the university's admission threshold $\underline{s}_j$ is $\Phi((s_i - \underline{s}_j)/\sigma)$, where $\Phi(\dot{})$ is the standard normal CDF. As the variance σ becomes large, students' test scores become less and less

informative about their admission probability. With individual application and admission data, the level of σ is identified by the variance of admission outcomes among students with the same difference between their test score and the university's threshold.

4.2 Identification: Enrollment Utility and Application Cost

The next set of parameters to identify are the parameters governing the enrollment utility and application cost, which can be found in equations 1 and 6. One identification concern is that observed university characteristics may be correlated with the unobserved university characteristics that enter the utility of enrollment. This will be the case for price and majors offered, as universities decide the prices and majors with knowledge of these unobserved characteristics. This is an endogeneity problem that is standard in the Industrial Organization literature (see Berry et al. (1995); Nevo (2000); Fan (2013)). Setting aside this concern for the moment, there remains the question of how each parameter can be separately identified using data on applications. I discuss this issue first and then discuss my solution to the endogeneity problem.

The parameters that govern the mean utility of enrollment (β_0, $\boldsymbol{\beta_1}$, $\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$) are separately identified by the covariation of the application and enrollment outcomes (application market penetration and enrollment share) and the university characteristics. All else equal, universities with a higher mean utility will receive applications from and ultimately enroll a greater proportion of high school graduates. Note that all of these parameters are interacted both with the corresponding university characteristic and the probability of admission, since they are within the expected value of an application portfolio. This means that, for example, while β_0 does govern the average utility of college enrollment, it also reflects the extent to which students value the probability of admission. As β_0 increases, students will be more likely to apply to every college, but the effect of β_0 varies with the admission probability: high β_0 implies that students will apply more often to "safe" options.

Applicant cost parameters capture the economic cost of each application. The constant in the application cost reflects the average level of applications overall, relative to the number of potential applicants who choose not to apply to any university. If this parameter is high, most students will send no college application, and if it is low, then students will be more likely to send at least one, which will generate a higher level of applications overall. The parameter on the log number of applications in the portfolio reflects the skewness of the distribution of applications within a market, apart from how the

skewness is driven by similarities in characteristics of universities in the same market. If this parameter is very large, then the second, third, and fourth applications are very costly relative to the first application, and most students will then send only one application. Because students will send only one application, they will be forced to choose between universities (based on their characteristics) and therefore the number of applications sent to the top university in the market will be very different from the number of applications sent to the second, and so on. On the other extreme, if the second, third and fourth applications were perfectly costless for all students, then all students who apply will send four applications, which increases the complementarity in applications and reduces the difference in application shares between institutions with different mean utilities.

Now we can return to the issue of the endogeneity of prices and majors. Without instruments, the estimated coefficients on price and majors offered in utility will be biased, because prices and majors are correlated with university characteristics that I do not observe (ξ_j). Thus, I need appropriate instruments for price and majors offered. In addition, I need instruments to identify the parameters on the number of applications within a portfolio, as the number of applications a student sends is a function of majors and prices at each institution. The following paragraphs describe the instruments and arguments for exogeneity and relevance.

For prices, the primary source of identification is an in-state indicator for public institutions. This is a valid instrument under the assumption that the only reason a student would choose to attend an in-state institution instead of an identical out-of-state institution is the price difference.

Instruments for majors offered are the number and type of graduate programs the university offers, measured the same way as for the undergraduate programs. These are relevant instruments (see Appendix B), and I argue that the relevance comes through correlation in the unobserved cost of providing each major. Exogeneity follows from the assumption that the decision to offer graduate programs is independent of demand for undergraduate education.

I use a state's average distance from the universities competing in the state as an instrument for the number of applications sent. Distance to each state is presumably exogenous to current demand shocks because universities largely chose their location many years prior to 2013, and location is relatively fixed. The instrument is correlated because students will generally apply to universities that are close, so students who live in remote states are unlikely to send as many applications as students who live in states that are home to many universities.

4.3 Identification of Supply Parameters

The supply-side parameters to be identified are the average public and private university preference for selectivity (λ_0^{priv} and λ_0^{pub}), and the marginal cost parameters α for both public and private universities. In this section I discuss the conditions under which the parameters are identified.

First, I focus on the first-order conditions for the private universities. The estimating equations derived from the the majors first-order conditions are given by equation 14. These K equations can be re-written as:

$$M_{jk} = \frac{(-\frac{\beta_{2k}}{\beta_1} - \alpha_1^{priv})}{2\alpha_{4k}^{priv}} - \frac{\alpha_{3k}^{priv}}{2\alpha_{4k}^{priv}} q_j - \frac{\omega_{jk}}{2\alpha_{4k}^{priv}},$$

where β_1 and β_{2k} are the estimated demand parameters.

From here, it is clear that the parameters α_1^{priv} and α_{3k}^{priv} can only be identified relative to α_{4k}^{priv} from this equation in isolation. Further, because universities are assumed to know their heterogeneous cost for each major, ω_{jk}, when they make their choice, the majors offered are a function of the heterogeneous cost. As enrollment, q_j, is also a function of majors offered—and therefore correlated with the heterogeneous cost—an instrument is required for enrollment. Exogenous demand-shifters are a natural choice. In practice, I use the average across markets of the exogenous portion of the enrollment utility.

Plugging this equation in to the pricing first-order condition yields the following estimating equation:

$$\alpha_0^{priv} - \sum_{k=1}^{K}(\frac{\beta_{2k}}{\beta_1} + \alpha_{4k}^{priv} M_{jk}) + 2\alpha_2^{priv} q_j + \boldsymbol{\alpha}_3^{priv} q_j \mathbf{M_j} + \omega_{0j} = p_j + t_j + \psi_j + q_j \frac{\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}}{\left(\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}\frac{\partial q_j}{\partial p_j} - \frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial p_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}\right)}$$

Prices are then a function of majors offered, total enrollment, and the heterogenous marginal cost per student. Because majors and enrollment are ultimately a function of prices, and therefore of the heterogeneous cost, ω_{0j}, instruments are required for each M_{jk} and the total enrollment. Again, I use the exogenous demand-shifters as instruments for the total enrollment. Just as with the demand-side, I use number and type of graduate programs as instruments for each of the k measures of majors offered. I claim that for each k, the corresponding measure of graduate programs is uncorrelated with the ω_{0j} portion of the heterogeneous cost per student.[18]

[18] While I don't model the choice of graduate programs, the relevance of the instrument is justified in a model where there is a heterogeneous cost per graduate program that is correlated with the undergraduate major-specific heterogeneous cost, ω_{jk}.

Because the α_3^{priv} can be separately identified from the other parameters in this equation, the α_4^{priv} and α_1^{priv} can be separated from α_3^{priv} in the majors first-order condition above when the two first-order conditions are combined.

The preference for selectivity, λ_0^{priv}, is estimated using the first-order condition for the admission threshold, reproduced here:

$$\frac{\frac{\frac{\partial \mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})}}{\left(\frac{\frac{\partial \mathcal{R}(p_j)}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial q_j}\frac{\partial q_j}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j) - \mathcal{C}(\mathbf{M_j})} - \frac{\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}}{\tilde{s}(\underline{\mathbf{S}}_j)}\right)} = \lambda_j.$$

I use a constant and the exogenous portion of mean utility as instruments in this equation.

The identification of the preference and cost parameters for public universities follows similarly, except that an additional equation allows identification of the cost difference between in-state and out-of-state students. In Appendix D, I describe an equation that yields an expression for the heterogeneous cost for in-state students in terms of data and known demand parameters. The only instrument required in this equation is a constant (in other words, I only assume that the mean of the error term in this equation is zero).

To summarize, the supply-side parameters are identified using a system of pricing, offered major supply, and selectivity equations. This system involves several cross-equation restrictions on parameters which help to identify the parameters in the major supply equation. Since prices, majors offered, selectivity and enrollment are endogenous, I instrument for these variables using the mean of the exogenous portion of enrollment utility. In addition, I use the number and type of graduate programs as instruments for undergraduate majors offered in the pricing equation.

5 Estimation

With the identification strategy outlined, I now turn to the estimation strategy and results for the admission, college choice, and supply-side parameters. In the first subsection, I describe how I simulate the individual choice sets. In subsections 5.2, 5.3 and 5.4 I describe the estimation of the admissions, demand, and supply-side parameters, following the order of the estimation routine.

5.1 Choice Sets

As presented in Section 3, each student may apply to any school within their individual-specific choice set, which is a subset of the population of universities that is considered by the student when the student makes their college application choice. In my setting, the distribution of the choice sets is both exogenous and unobserved, so I must make an assumption about these choice sets. There is a wide range of assumptions that would rationalize observed data. One potential assumption is that all students have full information and consider every university in the population. At the other extreme, I could assume every student always applies to every university in their choice set, so choice sets are indistinguishable from application portfolios (in this case, there would be effectively no choice at the application stage). All research on the application portfolio choice problem must make some assumption about the distribution of choice sets, but the assumptions vary. Howell (2010) utilizes an approach that allows estimation of a full-information model, and Arcidiacono (2005) uses a sampling approach to select choice sets from an exogenously determined distribution.

I use an assumption that students' choice sets are limited by their state, and by random information that generates differences in choice sets across students in the same state. This assumption is most similar to the approach used by Arcidiacono (2005). I use this assumption both because I think limited information is most consistent with the institutional setting, and because it helps resolve the computational difficulty of estimating the demand-side parameters.

The computation of the applicant penetration and enrollment share functions is challenging because it involves computing the expected value of application for all possible application portfolios in each market. Furthermore, for a single application portfolio, computing the expected value of an application portfolio requires computing the probability of admission and expected value of enrollment for all possible sets of admission outcomes given each application portfolio. A brute force method under the assumption that each student chooses an application portfolio of any size and is able to choose from any university in the country would render the problem impossible to solve. In this case, the potential number of application portfolios would be over 2^{800}, each with a number of possible application outcomes corresponding to the size of the application portfolio. The largest application portfolio would have over 2^{800} possible admission outcomes.

Substantial computational savings are generated by limiting each student to send a maximum of four applications. In the ELS:2002 data, which includes primarily students who would apply to college in 2004,

86% of college applicants apply to four or fewer of the four-year universities that are included in my analysis, so the assumption is not terribly restrictive.[19]

In Section 2.2, I described the geographic limitations on a university's market. This corresponds to a geographic limitation on each student's choice set, as students must choose from only the universities that compete in the student's state. After applying this limitation, the number of universities plausibly considered by students in each state varies, but the market with the largest number of universities has just over 150 universities competing for these students (see Appendix Figure 1). Combining this assumption with the limitation that students apply to no more than four universities still yields nearly 21 million possible portfolios in a state in which 150 universities compete.

I further limit the problem by making restrictions on the types of universities that are combined in a single application portfolio. Evidence from the individual data suggest that it is highly unlikely that students apply to, for example, three highly selective institutions (like Harvard, Yale and Stanford) and a non-selective university. To establish these results empirically, I examine patterns of selectivity in the ELS data using Barron's rankings from the 2004 *Profiles of American Colleges*.[20] I characterize application portfolios by the number of universities in each of six selectivity categories. I find that 95% of students' application portfolios fall in the top 74 combinations of selectivity categories, so I limit the possible portfolios to follow these patterns.

After applying these restrictions, I sample 1,000 individuals per market and generate their choice sets randomly from the universities that compete in their state. These choice sets correspond to the Υ_i in the model. In estimation, I assume that the choice sets include 20 universities that compete in the student's state.[21] There is no data to inform the likelihood of specific choice sets, but each option must be observed at least enough times that the applicant share data can be rationalized (i.e. a university that receives applications from 25% of all high school graduates must be in the consideration set for at least 25% of students). To achieve the desired properties, I weight the first seven options in the choice set by the observed share of all applications, and fill in the remaining 13 options with uniform sampling.

[19] The HSLS:09 data do not record all applications sent by each student, so it is not possible to get a corresponding number for students applying to college in 2013. The best information I have from that survey says that 78% of all applicants apply to four or fewer institutions among the entire population of postsecondary institutions (including 2-year institutions and other schools not included in my analysis).

[20] I use ELS here instead of the HSLS:09 because the HSLS:09 only records details on up to three applications per student.

[21] In Idaho, there are fewer than 20 universities competing in the state. In this case, no sampling is necessary; instead, all students see all options available in the state.

5.2 Admission Parameters

The first step in the estimation procedure is to obtain the parameters of the admission process. As outlined in Section 3.2, a student is admitted with probability: $\Phi((s_i - \underline{s}_j)/\sigma)$, where $\Phi(\cdot)$ is the standard normal CDF, s_i is the student's test score (on the ACT scale), and $\underline{s}_j$ is j's admission threshold.

The true value of the admissions thresholds are unobserved, but a close approximation may be the observed 25th percentile of the ACT score distribution within the university. This is a very public metric that is often used in college advising to determine where a student is likely to be admitted. In estimation I use these values as the admission thresholds.[22]

The variance of the match value, σ, is estimated from the HSLS individual data by running the probit regression:

$$Accept_{ij} = \Phi((s_i - \underline{s}_j)/\sigma), \tag{16}$$

where the $\underline{s}_j$ are taken from the data as discussed in the last paragraph. The results of this probit regression are shown in Table 2.

Table 2: Probit: Admissions

	coef	se
Student's Test Score - University's 25th Percentile ACT	0.150	(0.003)
Obs	12,360	
Log-Likelihood	-7,298	
Inferred Value of Sigma (1/coef)	6.653	

NOTE: This table displays the results from a probit regression of acceptance on the difference between each student's test score and the university's 25th percentile ACT score, as described in the text.

SOURCES: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, High School Longitudinal Study of 2009 (HSLS:09), Second Follow-Up. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), 2013 Institutional Characteristics File.

The estimated coefficient on the difference between the student's test score and the university's 25th

[22] Another approach would be to estimate the admission thresholds by inverting the enrollment equation (Equation 5.3) in the estimation routine to solve for the admission thresholds that set observe enrollment equal to expected enrollment. This adds substantial computational complexity, but it is feasible for future work. This method would allow public institutions' in-state and out-of-state admission thresholds to differ.

percentile ACT is 0.150, which implies a standard deviation of $1/0.15 = 6.653$ ACT points. This estimate suggests that factors other than the test score play an important role in the admission process.[23]

5.3 College Choice Parameters

Taking the variance parameter estimated in Section 5.2 as given, I then estimate the college choice model using the simulated method of moments. The moment conditions are derived from the equations for applicant and enrollment shares derived in Section . Given a set of instruments Z_{jl}, the moment conditions derived from the applicant share equation are $E[Z_{jl}\xi_{jl}(\theta)] = 0$, where θ are the demand parameters. I use the enrollment equation by setting $E[e_j(\theta)]$ equal to the observed enrollment, $Enroll_j$.

These moments are stacked in a Gx1 vector:

$$g(\theta') = \begin{pmatrix} E[Z_{jl}\xi_{jl}(\theta)] \\ E[e_j(\theta) - Enroll_j] \end{pmatrix}, \tag{17}$$

and then the method of moments objective function is can be written as:

$$\hat{\theta} = \underset{\theta}{argmin}\ g(\theta)'Wg(\theta). \tag{18}$$

W in this objective function is a block-diagonal weighting matrix. Because the parameters are just-identified with the chosen instruments, the choice of weighting matrix does not affect the estimates.

Computing the estimates is complicated by the fact that $\xi_{jl}(\theta)$ cannot be obtained by analytically inverting the applicant share equation. I used a nested routine that is similar to those used by Rust (1987) and Berry et al. (1995). The "outer loop" solves for the applicant cost parameters in θ. For each guess of θ considered in the outer loop, I find the mean enrollment utility ($\nu_j = \beta_0 + \beta_1 p_{jl} + \boldsymbol{\beta_2}\mathbf{M_j} + \boldsymbol{\beta_3}\mathbf{X_{jl}} + \xi_{jl}$) that sets the applicant share equation equal to the data. This is achieved by implementing a combination of the contraction mapping used in Berry et al. (1995) with Newton's method iterations, as in Rust (1987); Iskhakov et al. (2016).

The BLP contraction mapping in my case is:

$$\nu^{h+1} = \nu^h + \log(a(\nu^h, \theta_1)) - \log(A), \tag{19}$$

[23]The estimate is sensitive to the choice of admission threshold. In the future I plan to estimate the admission thresholds by inverting the enrollment share equation. This poses some computational challenges but is feasible for future work.

where the vector $a(\nu^h, \theta_1)$ contains the model applicant shares as a function of the mean enrollment utility and the non-linear demand parameters θ_1. A contains the applicant shares from the data, and h records the step number. Although my applicant share function is not a simple logit because of the admission and enrollment steps imbedded within it, the contraction mapping proposed by Berry et al. (1995) does converge to a solution globally, albeit slowly. Newton's method iterations, which are given by

$$\nu^{h+1} = \nu^h + [\nabla_{\nu^h} \log(a(\nu^h, \theta_1))]^{-1}(\log(a(\nu^h, \theta_1)) - \log(A)), \tag{20}$$

converge quite quickly, but converge locally. Therefore, I utilize BLP iterations until the model and data applicant shares are close enough for Newton's method iterations to converge. Once the ν_j are obtained, the linear parameters θ_1 and the residuals $\xi_{jl}(\theta)$ are found using linear IV-GMM.

The demand-side estimates are presented in Table 3, along with robust standard errors. For the variables that capture the number and type of major, I also show the marginal rate of substitution between the measure and price. The estimates show that students are willing to pay approximately \$136 per year for each additional major, holding the distribution across types fixed. Majors in Engineering, Math and Science are the most valuable. Students are willing to pay approximately \$890 per year for each one percentage point increase in the share in Engineering, Math and Science relative to the excluded category, which includes Health, Education, Criminal Justice, and Other. For Business and Communications, students are willing to pay approximately \$752 per year for each one percentage point increase. For majors in Arts, Social Sciences, and Humanities, students are willing to pay approximately \$630 per year for each additional percentage point increase relative to Health, Education, Criminal Justice and Other majors.

Distance to the university is also an important factor in a student's college choice, which is not surprising given results from Long (2004) and other studies. The effect of distance is convex—the point at which a greater distance becomes preferred is approximately 300 miles.

To give more intuition about the estimates, I present the total applications and enrollment under two counterfactuals, assuming there are no supply-side adjustments. First, I consider a drop in net price of \$1,000 at all schools. Second, I consider what would happen if all universities increased the number of majors in Engineering, Math and Science by 10% of the current level, holding other majors constant.

These simulations are presented in Table 4. A \$1,000 decrease in net tuition at all universities is predicted to increase the number of applications sent by 2.19%, and the number of students enrolled by

Table 3: Enrollment Utility and Application Cost Estimates

	coef	se
Net Price	-0.48***	0.05
# Majors	0.06***	0.02
Arts, Soc. Sci. and Humanities (share)	30.05	21.03
Business and Comm. (share)	35.83**	17.11
Eng., Math, and Sci. (share)	42.36**	19.18
Private	-3.39***	0.95
Distance	-13.83***	1.85
Distance Sq	4.5***	0.65
Appl. Cost Constant	7.64***	0.86
Log (1+ #Applications)	19.62***	0.30
Mandatory Testing Constant	-1.63	2.25
Mandatory Testing * Log (1+ #Applications)	1.96***	0.030

NOTE: The table displays the estimated enrollment utility and application cost parameters. The cost parameters are in the bottom panel. In addition to the variables displayed here, the regression includes a constant, the 6-year completion rate, an indicator for whether the university offers a graduate program, the average room and board charge, NCES Locale Codes, Barrons Ranks, and Carnegie Size and Setting classification codes.

Table 4: Demand-side Simulations

Simulation	Change in Applications	Change in Total Enrollment
$1,000 Decrease in Net Tuition	2.19%	1.74%
Increase Number of Engineering, Math and Science Majors by 10%	1.94%	1.57%

NOTE: The table displays the predicted changes in total applications and enrollment under two demand-side simulations, assuming no supply-side response.

1.74%. A 10% increase in the number of Engineering, Math and Science majors at all universities is predicted to increase applications by 1.94% and enrollment by 1.57%. When the number of Engineering, Math and Science major increases, it has an effect on the proportion of majors offered in the other categories, which is why these changes do not seem as large as the estimated effects relative to Health, Education, Criminal Justice and Other majors.

In the final demand-side exercise, I rank universities by the mean enrollment utility, loosely following an idea in Kapor (2016). Technically, I compute the ranking by removing all information and application frictions except for the regional market structure, and compute the share of students who would most prefer each option. The predicted ranking of universities for the top 10 is shown in Table 5. This simulation shows that students would generally attend a prestigious private university if they had the option. The effect of distance is reflected here in the relatively high ranking of centrally located universities, including Vanderbilt, Washington University in St. Louis, and Rice University. In other words, the reason we do not expect to see more universities from the Northeast in this list is because they are close to each other, and therefore not as valuable to students from around the country relative to a good university in a more central location.

Table 5: Ranking by Mean Utility

Institution Name	Rank	Avery et. al. (2013) Rank
HARVARD	1	1
VANDERBILT UNIV	2	36
STANFORD UNIV	3	5
UNIV OF CHICAGO	4	28
RICE UNIV	5	17
MASSACHUSETTS INST OF TECHNOLOGY	6	4
PRINCETON UNIV	7	6
YALE UNIV	8	3
WASHINGTON UNIV IN ST LOUIS	9	65
CALIFORNIA INST OF TECHNOLOGY	10	2

NOTE: The table displays the ranking of the top 10 universities from this model and the corresponding ranking from Avery et al. (2013).

The final column of the table compares my rankings to those proposed in Avery et al. (2013). In that paper, the authors use matriculation decisions to infer a ranking of universities, but the rankings are intentionally based on the preferences of high-performing students. The primary distinction between my method and theirs is that mine reflect average preferences among all high school graduates. My top 10

capture the top 6 from their list, but also includes centrally-located universities that are substantially lower in the Avery et al. (2013) rankings. It is most likely the case that high-performing students are less concerned about distance to college, which would generate the differences between the two rankings.

In this subsection I have presented the estimates from the college choice model and have shown that these estimates produce reasonable inferences about demand-side responses. In the next subsection, I take these parameter estimates as given and estimate the supply-side model.

5.4 Supply-Side Parameters

The first-order conditions presented in Section 3 provide six different equations for private universities and seven for public universities, corresponding to the choice of prices (one for private and two for public universities), the total number and percentage of majors (four variables), and the admission thresholds (one variable). I use these first-order conditions to estimate the preference and cost parameters given the demand parameters discussed in the previous section.

I estimate the parameters of the model using the simulated method of moments, with moment conditions based off of the heterogeneous selectivity preference and cost. The sources of heterogeneity are the heterogeneous component of the preference for selectivity ($\epsilon^s(\theta^s)$), and the heterogeneous marginal cost per student, ω_j. The heterogenous cost is decomposed into a constant and a heterogeneous cost for the number and each type of major. The moment conditions are $E[Z^s_{0j}\omega_{0j}(\theta^s)] = 0$, $E[Z^s_{kj}\omega_{kj}(\theta^s)]\ \forall k$, and $E[Z^s_{K+1,j}\epsilon^s(\theta^s)] = 0$, where Z^s is a set of instrumental variables for each equation, and θ^s are the supply-side parameters. Instruments are as discussed in Section 4.

The estimates from this regression are shown in Table 6. The selectivity preference parameter is high, at 0.87 for private universities and 0.92 for public universities. From this I conclude that a net-revenue maximizing framework does not characterize well the behavior of universities, at least in the context of this static model.

The marginal cost parameters are in many respects quite similar between public and private universities, although the shape of the cost function differs for some major categories. For public universities, the parameter on in-state students suggests that an in-state student is valued at nearly $10,500 dollars more than an out-of-state students. This difference reflects government preferences that are internalized by the public universities through legislative mandates and appropriations, to the extent that appropriations are tied to the share of in-state students.

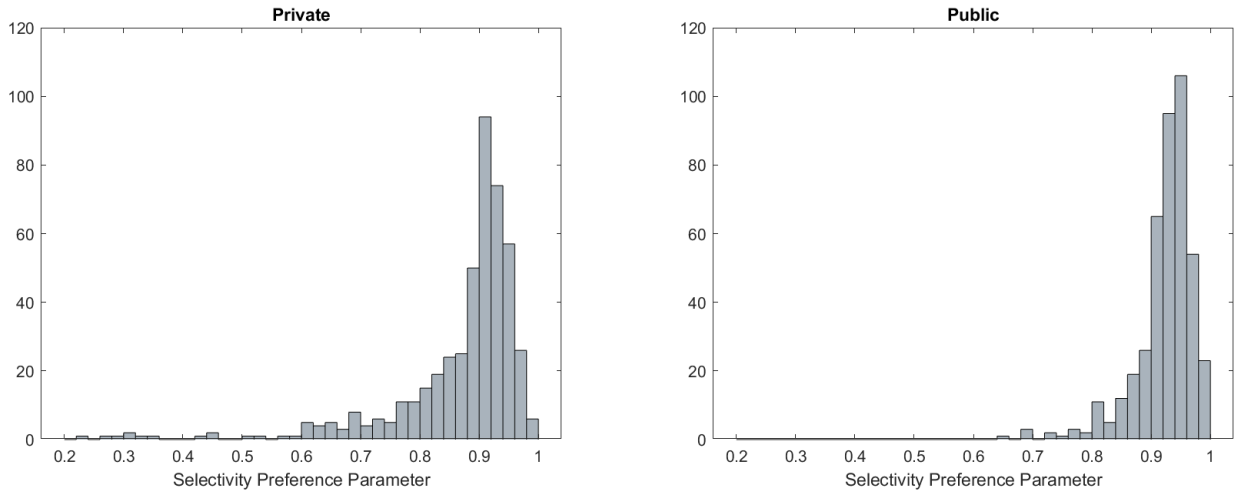
Table 6: Supply-Side Parameter Estimates

		(1)			(2)		
Parameter	Interpretation	Private		se	Public		se
λ	Average Preference for Selectivity	0.87	***	6.47e-03	0.92	***	4.16e-03
Marginal Cost Parameters							
α_0	Constant	65.61	***	6.22	-23.83	***	5.96
α_{11}	# Majors	-3.91	**	1.60	-0.09		0.12
α_{12}	Share Arts, Humanities and Soc. Sci.	62.98	***	0.07	45.12	***	4.92
α_{13}	Share Business and Comm.	68.17	***	1.13	66.51	***	3.04
α_{14}	Share Engineering	80.23	***	2.31	88.81	***	7.40e-02
α_2	q	-1.76e-02	***	5.51e-03	-3.30e-03	***	6.79e-04
α_{31}	q*# Majors	6.92e-04	***	1.25e-04	5.91e-05	***	1.11e-05
α_{32}	q*Share Arts, Humanities and Soc. Sci.	1.19e-04	***	1.88e-05	-3.38e-04		2.27e-04
α_{33}	q*Share Business and Comm.	1.06e-03	***	2.58e-04	4.15e-04	**	2.03e-04
α_{34}	q*Share Eng., Math, and Sci.	-5.56e-04	*	2.93e-04	4.18e-05	***	5.17e-06
α_{41}	# Majors Sq.	3.20e-02	*	1.65e-02	3.39e-04		1.21e-03
α_{42}	Share Arts, Humanities and Soc. Sci. Sq	2.83e-10		8.10e-02	25.67	***	7.00
α_{43}	Share Business and Comm. Sq	21.55	***	3.34	28.78	***	9.53
α_{44}	Share Eng., Math, and Sci. Sq	24.48	***	6.84	0.06		0.16
α_5	In-State Student				-10.49	***	0.02
Observations		466			428		

NOTE: The table displays the estimated supply-side parameters computed using a GMM objective function as described in the text.

The estimated cost parameters imply a distribution of university preference for selectivity, as captured by the heterogeneous Cobb-Douglas parameter λ_j. This distribution is shown in Figure 7. The distribution is tighter for public universities than for private universities — perhaps a reflection of the wide variation in selectivity and resources at private universities.

Figure 7: Distribution of Preference for Selectivity



NOTE: This figure shows the distribution of the estimated selectivity preference parameter among public and private universities.

6 Counterfactual Simulations

A series of counterfactual simulations illustrates the equilibrium adjustments to prices, selectivity, and majors offered when students are provided additional financial aid to attend public universities in a state. I use as an example the state of Wisconsin, but there is nothing in the model that restricts me to study Wisconsin alone. In the future, I can run similar counterfactuals for other states or even the entire nation.

In the counterfactuals, Wisconsin introduces a $1,000 grant for all Wisconsin residents to attend a public university in the state. The state also restricts public universities from responding with changes in prices, majors offered, or admission thresholds. Because the equilibrium calculations are time-intensive, I run these initial simulations only allowing adjustments by private universities that directly compete for Wisconsin students. The Nash equilibrium in each counterfactual is computed using iterated best-response.

These counterfactuals function as a proof-of-concept, and are primarily intended to illustrate the mechanics of the model. In future work, I plan to extend the counterfactuals in a number of ways, including allowing adjustments by public universities and the private universities that do not compete

directly for Wisconsin students. These additions will allow inference about policy responses in a more robust framework.

Table 7 shows the counterfactual changes in prices, applications, enrollment and selectivity at private universities in response to the grant offered at public universities. I break down the counterfactual into assumptions about the endogenous variables and about the competitive response. I first allow no response, then allow only prices to be flexible, then prices and admission thresholds, and finally, prices, admission thresholds, and the number of majors offered. Future versions of this paper will include a variation that allows the distribution of offered majors across types to be flexible. For each simulation, I break down the results further by first allowing each university to respond assuming no response on the part of other private universities, and then I allow a full Nash equilibrium in which each private university takes into account the responses of the other private universities.

The first row shows the change in applications, enrollment and the percentile rank of the average student when private universities do not respond. In this case, applications to private universities decrease by 1.12% and enrollment decreases by 2.87%. Although the admission threshold does not change in this simulation, the average student admission score goes down by 0.12%. This is because, on average, students with higher scores are more likely to switch to the public universities as a result of the grant. Since students with high scores are more likely to be admitted to a public university, they are more likely to have the option to attend a public university at the reduced rate.

If private universities are only able to change prices in response to the grant offered at public universities, the private universities reduce their prices, whether or not they take into account their private competitors' responses. When competitor's responses are taken into account the price reduction is slightly smaller. By reducing their prices, private universities are able to gain back most of the lost enrollment and some of the high-scoring students. However, the price reduction is not enough to recoup the selectivity loss due to the grant offered at public universities.

The conclusions change significantly when private universities are allowed to change both prices and admission thresholds. In this case, the prices and admission thresholds both increase, resulting in application and enrollment declines and a net increase in selectivity at private universities. The increases in selectivity that come from raising the admission threshold are valued more than the reduction in revenue through enrollment declines. Price increases on net have a positive effect on net revenue (in general universities are pricing below the net revenue-maximizing point). When competitor's responses are taken

into account, the price and admission threshold changes are muted slightly.

In the final simulation, I allow prices, admission thresholds, and the number of majors offered at private universities to respond. In this case, universities raise prices and admission thresholds as before, but now the universities can also adjust their set of offered majors to balance the cost of offered each major against the effect on selectivity and revenue. On average, I find that universities increase the number of majors offered. Again, allowing for private universities to respond to private competitors changes the equilibrium outcomes slightly. Prices and admission thresholds do not increase as much as before, while the number of majors increases by more than it did without the competitive response.

Allowing public universities to respond, increasing the size of the grant, or extending this counterfactual to include changes in the type of major are likely to produce different average effects. Another caveat that should be considered here is that my model does not include fixed costs and an exit condition. If universities must obtain a minimum revenue to avoid shutdown, that may substantially alter predictions for smaller universities.

The average effects obscure the heterogeneity in responses, especially with respect to university size and location. For small private universities in Wisconsin, a change to pubic universities in Wisconsin has a greater impact than it does to the University of Chicago, for example, as the University of Chicago competes for Wisconsin students but also draws most of its enrollment from outside of Wisconsin. Among the 27 private universities in this simulation, the correlation between the model-predicted enrollment and the percentage change in price is -0.11, so smaller universities increase their prices more than large ones do. Smaller universities also make larger increases in admission thresholds and in the number of offered majors.

Table 7: Private University Response to a $1,000 Scholarship for Wisconsin Public Universities

		Average Change Among Private Universities in:					
Endogenous Response	Competitive Assumption	Prices	Admission Thresholds	# Majors	Applications	Enrollment	Rank of Avg. Student
No Response	No Response				-1.1232%	-2.8682%	-0.1191%
Prices	No Competitor Response	-1.2428%			0.5136%	-0.3732%	-0.1105%
	Nash Equilibrium	-1.2419%			0.5132%	-0.3739%	-0.1106%
Prices and Admission Thresholds	No Competitor Response	19.9783%	2.1069%		-7.1314%	-10.1574%	1.1677%
	Nash Equilibrium	19.9706%	2.1051%		-7.1268%	-10.1509%	1.1675%
Prices, Admission Thresholds, and # Majors	No Competitor Response	20.2553%	2.1242%	1.1822%	-7.2429%	-10.3178%	1.1778%
	Nash Equilibrium	20.1625%	2.1092%	1.1911%	-7.1565%	-10.1926%	1.1699%

NOTE: The table displays counterfactual responses of private universities competing for Wisconsin students when in-state students are given $1,000 to attend public universities. Public universities are constrained not to respond.

7 Conclusions and Future Work

Despite notable advances in modeling this the higher education market, there is much more to be done. Economists have just scratched the surface in terms of understanding competition among universities and how this effects equilibrium responses to policy. This paper seeks to further our understanding through a model of oligopoly with differentiated universities. In the model, universities set prices, admissions selectivity, and offered majors in a strategic environment.

The model of college choice demonstrates that students value majors offered at approximately $136 on average, with substantial heterogeneity by type. Engineering, Math, and Science majors are valued the most, followed by Business and Communications, then Arts, Social Science, and Humanities.

On the supply-side, I show that costs differ by type of major. In a policy simulation, I consider the effect of increased financial aid offered to students who attend in-state public universities. As an example, I simulate a $1,000 grant offered to Wisconsin students to attend public universities, and I hold public universities' other characteristics fixed. Holding private universities' majors and selectivity fixed, the model predicts that on average, private universities' prices would decline by 1.2%.

The results demonstrate the importance of including prices, admissions selectivity, and offered majors as joint outcomes. If universities could only change their prices, then prices on average decline by 1.2%, but if they are allowed to change the admission threshold as well, prices and admission thresholds both increase, on average. Finally, if they also change the total number of majors offered the average effect is a price increase of 20%, an admission threshold increase of 2.1%, and an increase in the total number of majors offered of 1.2%.

The results should be interpreted with some caution, as there are a number of extensions and generalizations that may affect the estimates and counterfactual responses. In the following paragraphs I highlight a number of important directions for future work.

It is clear from summary data released by the College Board and ACT that test scores are correlated with student's intended major as indicated when they take the SAT or ACT. This motivates an interaction between the student's test score and their preference for certain types of majors.[24] This change would alter the current model from a vertical differentiation framework to a horizontal differentiation framework. It would also add an interesting interaction between the admission threshold and the expected application and

[24] A similar interaction is included in Arcidiacono (2005) and turns out to be important in that model.

enrollment. Instead of the admission threshold only affecting expected applications and enrollment through the admission probability, it would also affect the utility derived from the offered majors among likely enrollees. It is feasible to include such an interaction in the enrollment utility although it would impose some additional computational burden.[25]

Some other future improvements involve a more nuanced treatment of prices and the admission process. For example, given the joint distribution of race and test scores I can control for race in the admission score. Using data on the income distribution by state and average net price by income group, I can also simulate prices that vary with income. Universities would then choose the list price (tuition) and financial aid by income group would be determined exogenously.

Another area for future work deals with the admission thresholds. In the current estimation, admission thresholds are taken from the data. I assume that the observed 25th percentile ACT score among each university's enrollees is equivalent to the admission threshold in the model. Another way to obtain the admission thresholds would be to estimate them in the demand-side routine by inverting the enrollment equation for the admission threshold. Then the admission threshold would be the value that sets observed enrollment equal to the model expected enrollment for each university, given the other demand parameters. This method would be advantageous because it would allow for public universities to have different in-state and out-of-state admission criteria, and would not rely on a crude proxy for the admission thresholds. Conceptually this is not challenging, but practically it involves computing the solution to over 800 additional equations for each iteration of the GMM minimization routine. Furthermore, I would need to incorporate the estimation of the admission variance parameter in the demand estimation step. I plan to approach these problems using additional moments (based on the test scores of enrollees) and an estimation routine similar in spirit to the Mathematical Programming with Equilibrium Constraints (MPEC) approach proposed by Dube' et al. (2012).

While there is still work to be done, this research has made a significant step towards understanding the importance of competition and market structure in the U.S. higher education market. I demonstrate that there is substantial variation among universities in the number and type of majors that are offered, and that this variation has implications for college choice. Counterfactual simulations point to the importance of different assumptions about market structure and the nature of competitive response. Therefore, analysis of

[25] An even more thorough approach to modeling the effect of majors on demand would be to model the choice of college major as a fifth stage appended to the current framework. I see this as an important direction for future work.

higher education policy should thoughtfully address how market mechanisms affect the interpretation of results.

References

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

Avery, C. N., M. E. Glickman, and C. M. Hoxby (2013). A Revealed Preference Ranking of U.S. Colleges and Universities. *Quarterly Journal of Economics 128*(1), 425–467.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile Prices in Market Equilibrium. *Econometrica 63*(4), 841–890.

Berry, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics 25*(2), 242–262.

Bresnahan, T. F. ., S. Stern, and M. Trajtenberg (1997). Market Segmentation and the Sources of Rents from Innovation : Personal Computers in the Late 1980s. *The RAND Journal of Economics 28*(0).

Card, D. and A. B. Krueger (2005). Would the elimination of affirmative action affect highly qualified minority applicants? Evidence from California and Texas. *Industrial and Labor Relations Review 58*(3), 416–434.

Cook, E. E. and S. Turner (2019). Missed Exams and Lost Opportunities: Who Could Gain From Expanded College Admission Testing? *AERA Open 5*(2), 233285841985503.

Dube', J.-P., J. Fox, and C.-L. Su (2012). Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation. *Econometrica 80*(5), 2231–2267.

Epple, D., R. Romano, and H. Sieg (2006). Admission, Tuition, and Financial Aid Policies in the Market for Higher Education. *Econometrica 74*(4), 885–928.

Fan, Y. (2013). Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market. *American Economic Review 103*(5), 1598–1628.

Fu, C. (2014). Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market. *Journal of Political Economy 122*(2), 225–281.

Gentzkow, M. (2007). Valuing New Goods in a Model with Complementarity: Online Newspapers. *American Economic Review 97*(3), 713–744.

Gewertz, C. (2018). Which States Require Students to Take the SAT or ACT?

Howell, J. S. (2010). Assessing the Impact of Eliminating Affirmative Action in Higher Education. *Journal of Labor Economics 28*(1), 113–166.

Hoxby, C. and S. Turner (2019). The Right Way to Capture College "Opportunity". *Education Next 19*(2).

Hoxby, C. M. (1997). How the Changing Market Structure of U.S. Higher Education Explains College Tuition.

Hoxby, C. M. (2009). The Changing Selectivity of American Colleges. *Journal of Economic Perspectives 23*(4), 95–118.

Hurwitz, M., J. Smith, S. Niu, and J. Howell (2015). The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams ? *Educational Evaluation and Policy Analysis 37*(1), 138–159.

Hyman, J. (2017). ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice. *Education Finance and Policy 12*(3), 281–311.

Iskhakov, F., J. Lee, J. Rust, B. Schjerning, and K. Seo (2016). Comment on "Constrained Optimization Approaches to Estimation of Structural Models". *Econometrica 84*(1), 365–370.

Kapor, A. (2016). Distributional Effects of Race-Blind Affirmative Action. *Working Paper*.

Long, B. T. (2004). How have college decisions changed over time ? An application of the conditional logistic choice model. *121*, 271–296.

Nevo, A. (2000). A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand. *Journal of Economics & Management Strategy 9*(4), 513–548.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines : An Empirical Model of Harold Zurcher. *Econometrica 55*(5), 999–1033.

Smith, J., M. Pender, and J. Howell (2018). Competition among Colleges for Students across the Nation. *Southern Economic Journal 84*(3), 849–878.

Strauss, V. (2018, mar). A University of Wisconsin campus pushes plan to drop 13 majors – including English, history and philosophy.

Appendix A Estimating the Test Score Distribution

Data from the ACT and SAT tests provides the distribution of test scores among test-takers. Estimation of the model requires the distribution of scores among the entire population of high school graduates. To obtain the counterfactual score distribution as if all students took and admission test, I utilize variation by state over time in the existence of universal (or mandatory) college admission testing policies.

Since 2001, a growing number of states have adopted mandatory college admission testing policies, which require students to take an ACT or SAT test prior to high school graduation. There is a growing literature devoted to documenting the effects of these policies on testing participation and college enrollment (Hurwitz et al., 2015; Hyman, 2017; Cook and Turner, 2019). The most critical— though unsurprising—empirical point is that on average, students who choose not to take a test when it is voluntary have lower scores on average than students who would voluntarily take the test. Thus, I cannot assume that the score distribution among test-takers is, in general, the same as the distribution among all high school graduates.

In states with mandatory testing, I can use the distribution among test-takers because all high school graduates are required to take the test.[26] In the other states, I predict the average score and then assume the score distribution is normal with a standard deviation equal to the national standard deviation.

To predict the average score, I first combine ACT and SAT data on test scores by state. I utilize tables provided by ACT and College Board to convert SAT scores to ACT scores. I utilize ten years of data for this exercise (2006-2015), which gives me 510 observations for the 50 states plus D.C. over ten years. Then I estimate the following regression to predict average scores, where l is state and t is the year, and $Mand\_Test_{lt}$ is an indicator for whether the state l had a mandatory testing policy in year t:

$$avgscore_{lt} = \beta_1 Mand\_Test_{lt} + \delta_l + \nu_l t + \epsilon_{lt}.$$

The estimates from this regression show that a mandatory testing policy reduces the average score by 1.327 ACT points. I predict the average score under mandatory testing for all states. Then the distribution of scores is a normal distribution centered around the predicted values, with a standard deviation equal to the national standard deviation of ACT scores.

[26] Opt-out is limited. Typically participation rates are above 95% in these states.

Appendix B A Simplified Model of College Choice

The most basic question in this research is whether net tuition, majors offered, and admissions selectivity affect a student's choice of college. In this section, I answer this question in isolation by regressing measures of college choice (based on applications and enrollment) on variables affecting a student's choice of college. This presentation will introduce some variables and notation used in the structural model, and will preview the identification strategy.

B.1 Model

First, I define some notation, which is consistent with notation used in the structural model. Let $l = 1...L$ denote the high school graduate's state of residence. There are $J + 1$ universities $j = 0, 1, ..., J$, where $j = 0$ represents the outside option of four-year colleges not included here (such as for-profit universities), two-year colleges, other educational certificate programs, the labor market, or other options. Each university offers a number of majors within each of K broad categories. Let $\mathbf{M_j}$ be a $K \times 1$ vector where the first element, M_{j1} is the total number of majors university j offers, and the remaining $K - 1$ elements are the proportion of majors in each group (with Health, Education, and Other the excluded category). Universities also have a number of exogenous characteristics, $\mathbf{X_{jl}}$, such as whether the university is under private or public control and the distance to the population center of each state. Universities charge a net tuition, p_{jl}, which is tuition minus average grants and aid. For private universities, this price will be constant across l, while for public universities, the price varies by in-state status (whether l is the state where j is located).

Universities also have admissions criteria, which determine how difficult it is for a student to gain admission. The likelihood of admission is likely an important consideration when students choose where to apply — since applications are costly, a student will not apply to a university if he has very little chance of admission. The likelihood of admission also mechanically reduces enrollment, so it is important to control for this factor when explaining enrollment. I use the observed 25th percentile of ACT scores among entering freshmen at each university as an index of admission difficultly, and I call this $\underline{s}_j$. In the structural model, $\underline{s}_j$ takes on a more precise meaning as an admission threshold on a student-university match value.

The first dependent variable in this simplified model is constructed from applications—the natural logarithm of the ratio of university j's share of applications from state l to the share of students from state l

who sent no application ($j = 0$) (notated $ln(applicationshare_{jl}/applicationshare_{0l})$). This dependent variable can be motivated in two ways. First, one can think of this simply as a sensible alternative to using the number of applications to each university by state. Using shares is more appealing than just the number of applications because the shares account for the fact that some states graduate more students than others. Dividing by the share of students who do not apply to any university and taking logs generates a dependent variable that can take any real value, making it an appropriate dependent variable for a linear combination of explanatory variables.

A second way to motivate this dependent variable is to derive it from a logit demand system as in Berry (1994), with some ad hoc assumptions. First, we must assume that each application can be viewed as arising from a completely independent decision (as if each application corresponded to a separate student), where each application represents a single discrete choice. Second, we must assume admissions selectivity enters linearly into utility, as do prices and other characteristics. Then we can write the utility function as:

$$u_{ijl} = \beta_0 + \beta_1 p_{jl} + \boldsymbol{\beta_2}\mathbf{M_j} + \boldsymbol{\beta_3}\mathbf{X_{jl}} + \beta_4 \underline{s}_j + \xi_{jl} + \eta_{ijl}. \tag{21}$$

This utility function includes a student-university specific unobservable, η_{ijl}, which is distributed Type I Extreme Value, and a university-market specific unobservable ξ_{jl}.

Let $\nu_{jl} = u_{ijl} - \eta_{ijl}$ and normalize $\nu_{0l} = 0$. Then the application share can be written as:

$$applicationshare_{jl} = \frac{e^{\nu_{jl}}}{1 + \sum_{j'=1}^{J} e^{\nu_{jl'}}}. \tag{22}$$

Now the regression equation can be found by dividing the application shares by the share of students who do not apply to college:

$$ln(applicationshare_{jl}/applicationshare_{0l}) = \nu_{jl} = \beta_0 + \beta_1 p_{jl} + \boldsymbol{\beta_2}\mathbf{M_j} + \boldsymbol{\beta_3}\mathbf{X_{jl}} + \beta_4 \underline{s}_j + \xi_{jl}. \tag{23}$$

I also run a similar regression with a dependent variable based on enrollment on the left-hand side. In this case, the assumption that each student choose only one university to attend is not concerning. However, this model does not take into account how the application stage affects the enrollment share. The

regression equation in this case is:

$$ln(enrollmentshare_{jl}/enrollmentshare_{0l}) = \nu_{jl} = \beta_0 + \beta_1 p_{jl} + \boldsymbol{\beta_2}\mathbf{M_j} + \boldsymbol{\beta_3}\mathbf{X_{jl}} + \beta_4 \underline{s}_j + \xi_{jl}. \quad (24)$$

B.2 Identification and Estimation

The regression equations derived above can be estimated via OLS. However, if students and universities know the unobservable ξ_{jl}, and universities choose prices, majors, and admissions selectivity knowing that students will respond, then prices, majors, and admissions selectivity are correlated with ξ_{jl}. This is the classic endogeneity problem that is considered in Berry et al. (1995) and in related literature. Appropriate instruments can be used with a two-stage least squares estimation routine to obtain consistent estimates. The instruments I utilize in the regressions presented below are: 1) an indicator for public in-state status, 2) the average number of competitors faced by university j in its markets, and 3) the number and type of graduate programs the university offers, measured in the same way as the undergraduate majors.

The first-stage regressions are summarized in Table 8. As seen in the first column, the in-state public indicator is the main instrument responsible for the identification of the price coefficient. In other words, the price coefficient is identified primarily by comparing how in-state and out-of-state application or enrollment shares vary with the differences in in-state and out-of-state prices, for public institutions. The underlying assumption is that the only reason a student goes to an in-state public university instead of another university with the same observable characteristics (including the same location) is the difference in prices.

The second column in Table 8 shows that several of the instruments are correlated with the 25th percentile ACT score. Columns 3-7 demonstrate that graduate programs are positively correlated with undergraduate programs, both in total number and within type. Graduate programs are only valid instruments for undergraduate programs if they are also uncorrelated with the demand-side unobservable for undergraduates. I assume that the graduate and undergraduate program decisions are independent, but that common drivers on the cost side produce the observed correlations. A more thorough explanation is found in Section 4.

Table 8: First-Stage Results for 2SLS Estimation of Equation 23

	(1) Average Net Tuition	(2) 25th Percentile ACT	(3) Number of Majors	(4) Share Arts, Soc. Sci. and Humanities	(5) Share Business	(6) Share Eng., Math and Sci.
In-State Public Indicator	-12.46***	0.386***	-0.418	0.0151*	-0.00163	0.00411
	(0.47)	(0.11)	(1.76)	(0.01)	(0.00)	(0.01)
Mean Number of Competitors	0.0483***	-0.00856***	-0.0689***	0.000208*	8.83e-06	9.55e-05
	(0.01)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)
Number of Graduate Programs	0.00647	-0.00102	0.395***	-0.000456**	7.29e-05	-2.09e-06
	(0.01)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)
Grad Share Arts, Soc. Sci. and Humanities	2.373*	0.634*	-6.574*	0.168***	-0.0349**	0.0190
	(1.27)	(0.36)	(3.71)	(0.03)	(0.02)	(0.02)
Grad Share Business and Communications	3.463**	0.824**	-8.282	-0.0544	0.174***	0.0347
	(1.42)	(0.40)	(7.38)	(0.04)	(0.05)	(0.04)
Grad Share Eng., Math, and Sci.	-0.0532	3.021***	-20.24***	-0.282***	-0.0501**	0.494***
	(1.52)	(0.45)	(5.94)	(0.06)	(0.02)	(0.06)
Observations	3,324	3,324	3,324	3,324	3,324	3,324
SW F-Stat (1, 893)	128.84	32.29	127.96	82.71	65.55	58.26

NOTE: Standard errors are clustered at the university level. Exogenous university characteristics (the "Included Instruments") are not shown. They are: room and board charges, private institution indicator, distance and distance squared, whether or not the university offers a graduate program, the NCES Locale Codes, Carnegie Size and Setting classifications, Barron's ranks and whether the state has mandatory college admission testing.

B.3 Results

In Table 9 I display the results from the 2SLS estimation of equation 23 (column 1) and a corresponding equation for the enrollment share (column 2). The results are very similar across the two equations, except for the constant and the mandatory college admission testing indicator. In the description I will focus on the application share equation, and then briefly discuss the difference between the coefficient on universal testing in the two regressions. The first seven rows show the coefficients on the endogenous variables in the model. The coefficient on average net tuition is 0.144, which means that a $1,000 increase in net tuition will decrease the ratio of the application share of j to the share applying to no college by 14.4 percent. A 1-point increase in the 25th percentile ACT score has approximately the same effect on applications as a $2,292 increase in net tuition.

Both the number and type of majors affect applications. Adding a major while keeping the distribution across types the same has the same effect on applications as a $99 decrease in net tuition. The type of major is measure in proportions relative to Health, Education, Criminal Justice, and Other, so a 1 percentage point increase in majors in Arts, Social Science, and Humanities obtained by reducing the proportion of Education majors by 1 percentage point is equivalent to a $250 decrease in net tuition. The corresponding numbers for Business and Engineering, Math and Science categories are $420 and $411 dollars, respectively.

Other explanatory variables are included, such as room and board charges, an indicator for private control, distance and distance squared, whether the university has a graduate program, and the rate at which students complete the graduate program within 150% of the normal time-to-degree. All of these have intuitive signs. Controls for NCES locale codes (loosely, whether the school is in an urban, suburban or rural area) and Carnegie Classifications for the size and setting of the institution are also included.

Finally, an indicator for whether the state has mandatory college admission testing is included.[27] Mandatory testing is associated with substantial increases in score-report sending—each university would see changes in score reports received equivalent to those that would be observed after a $4,000 decrease in net tuition. However, it is not associated with any statistically significant effect on enrollment. This may occur because most of the new score reports are not associated with any applications, or because the new applications are from students who are not qualified for admission, or because universities respond to

[27] I used the list of states with mandatory testing policies by Gewertz (2018) confirming via internet research.

increased applications by increasing admission thresholds. Regardless, it is clear that universal admission testing substantially affects score report sending, while it is likely not associated with an increase in utility of attending each institution. Thus, in the full model, universal admission testing policies are assumed to affect application cost but not enrollment utility.

While this model shows that prices, selectivity and majors affect applications and enrollment, the model relied on some assumptions. The model of applications assumed that applications arise from an independent discrete choice where each student picked his most preferred option, but in the "real world", students often apply to more than one university because they are uncertain about admission outcomes. Observed applications are not necessarily first choices but may represent a second or third choices (or even lower down the list). Enrollment, on the other hand, does reflect a single discrete choice, but students must enroll only at institutions to which they apply and are admitted. The fact that the enrollment and application utility estimates differ very little reflects the strong correlation between applications and enrollment.

If students actually had to choose their most preferred option, we would likely observe a greater spread between the applicant (and enrollment) shares of the most- and least-preferred universities. For example, assume that there are two universities (1 and 2), and assume all students prefer university 1 to university 2, but prefer either university to attending no university. Assume also that there is some uncertainty in admissions. Then if applications are costless, all students will apply to both universities. If students were instead forced to choose between university 1 and 2 (e.g. by a high application cost), then all students would choose university 1 (assuming that the admissions selectivity doesn't change enough in equilibrium to alter students' rankings of the universities). In the first case, both universities have a 100% applicant share, and in the second case, university 1 has a 100% applicant share while university 2 has a 0% applicant share. Enrollment shares will change corresponding to equilibrium adjustments in admissions.

This example highlights the fact that differences in utility may not be captured correctly by a model that fails to account for the multiple-discrete choice nature of the application decision. While this is a concern in terms of interpreting the point estimates, the simple logit model does provide evidence that prices, majors offered and admission selectivity all affect applications and enrollment. Modeling the application, admission, and enrollment stages jointly gives even more—we learn separately about the application cost, admission uncertainty, and enrollment parameters, and we are able to analyze all three outcomes separately. Joining this college choice model with a supply-side model informs us about the university's preferences and likely policy responses. These advantages come at a cost, both in terms of the

assumptions imposed in the structural framework and in terms of computation. The benefits and costs are illuminated in the description of full model, identification and estimation strategy in Sections 3, 4, and 5.

Table 9: 2SLS Estimation of College Choice Logit (Equation 23)

	(1) Ln(Application Share Ratio)		(2) Ln(Enrollment Share Ratio)	
	coef.	se	coef.	se
Average Net Tuition	-0.144***	(0.0103)	-0.171***	(0.0103)
25th Percentile ACT	-0.330**	(0.138)	-0.347**	(0.143)
Number of Majors	0.0143***	(0.00492)	0.0148***	(0.00535)
Share Arts, Soc. Sci. and Humanities	3.600**	(1.655)	3.576**	(1.757)
Share Business	6.088***	(1.992)	6.453***	(2.109)
Share Eng., Math, and Sci.	5.914***	(1.797)	4.684**	(1.861)
Room and Board	-0.0262	(0.0206)	-0.0252	(0.0223)
Private Institution	-0.596***	(0.198)	-0.643***	(0.190)
Distance (1,000s mi.)	-1.808***	(0.251)	-1.957***	(0.266)
Distance Sq.	0.576***	(0.0910)	0.586***	(0.0976)
Any Graduate Program	0.363**	(0.146)	0.247	(0.165)
Completion Rt 150% Normal Time	7.458***	(1.962)	7.819***	(2.036)
Mandatory Testing	0.582***	(0.0581)	0.0197	(0.0482)
Constant	-1.579	(2.270)	-3.663	(2.405)
Observations	3,324		3,324	
R-squared	0.224		0.410	

NOTE: Standard errors are clustered at the university level. NCES Locale Codes, Carnegie Size and Setting classifications, and Barron's ranks are included but not shown.

Appendix C Technical Notes

In the demand estimation, analytic derivatives of the GMM objective function with respect to the application cost parameters are used to supply the solvers with the gradient. In addition, certain derivatives are required for calculation of standard errors. This Appendix shows the derivation of the required expressions.

C.1 Derivatives of Applicant Shares

In estimation, the applicant shares are computed by summing over simulated i:

$$a_j = \sum_i \sum_{\mathbf{Y} \in \Upsilon_i \ni Y_j = 1} \mathbb{P}(Apply_{\mathbf{Y}} | s_i, \Upsilon_i),$$

where:

$$\mathbb{P}(Apply_{\mathbf{Y}} | s_i, \Upsilon_i) = a_{i\mathbf{Y}} = \frac{e^{\sum_{\mathbf{D} \subseteq \mathbf{Y}} \left(\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) ln(1 + \sum_{j \text{ s.t. } D_j = 1} e^{\nu_j}) \right) - c(\mathbf{Y})}}{\sum_{\mathbf{Y} \in \Upsilon_i} e^{\sum_{\mathbf{D} \subseteq \mathbf{Y}} \left(\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) ln(1 + \sum_{j \text{ s.t. } D_j = 1} e^{\nu_j}) \right) - c(\mathbf{Y})}}.$$

The value of an application portfolio is as defined in Section 3, equation 6:

$$V(\mathbf{Y}|s_i) = \sum_{\mathbf{D} \subseteq \mathbf{Y}} (\mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) v(\mathbf{D})) - c(\mathbf{Y}) + \epsilon_{i\mathbf{Y}}, \tag{25}$$

Let ν_j represent the mean utility of university j, as defined in Subsection 3.1, and let θ_1 represent the parameters of the application cost. As in Nevo (2000), I find $\frac{\partial \nu}{\partial \theta_1}$ using the Implicit Function Theorem:

$$\frac{\partial \boldsymbol{\nu}}{\partial \theta_1} = - \begin{bmatrix} \frac{\partial a_1}{\partial \nu_1} & \cdots & \frac{\partial a_1}{\partial \nu_J} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_J}{\partial \nu_1} & \cdots & \frac{\partial a_J}{\partial \nu_J} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial a_1}{\partial \theta_{11}} & \cdots & \frac{\partial a_1}{\partial \theta_{1H}} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_J}{\partial \theta_{11}} & \cdots & \frac{\partial a_J}{\partial \theta_{1H}} \end{bmatrix}$$

The derivatives of the applicant shares with respect to the own-university mean values are:

$$
\begin{aligned}
\frac{\partial a_j}{\partial \nu_j} &= \frac{1}{ni}\sum_i \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \frac{\partial a_{i\mathbf{Y}}}{\partial \nu_j} \\
&= \frac{1}{ni}\sum_i \left[\sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \left[\frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_j} a_{i\mathbf{Y}} - a_{i\mathbf{Y}} \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_j} a_{i\mathbf{Y}} \right] \right] \\
&= \frac{1}{ni}\sum_i \left[(1 - \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} a_{i\mathbf{Y}}) \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_j} a_{i\mathbf{Y}} \right] \\
&= \frac{1}{ni}\sum_i \left[(1 - \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} a_{i\mathbf{Y}}) \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \left(\sum_{\mathbf{D}\in\mathbf{Y}\ni D_j=1} P(\mathbf{D}|s_i,\mathbf{Y}) \frac{e^{\nu_j}}{1+\sum_{g\in D_g=1} e^{\nu_g}} \right) a_{iY} \right]
\end{aligned}
$$

And the derivatives of j's applicant share with respect to other universities' mean values are:

$$
\begin{aligned}
\frac{\partial a_j}{\partial \nu_g} &= \frac{1}{ni}\sum_i \left[\sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1, Y_g=0} \frac{\partial a_{i\mathbf{Y}}}{\partial \nu_g} + \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1, Y_g=1} \frac{\partial a_{i\mathbf{Y}}}{\partial \nu_g} \right] \\
&= \frac{1}{ni}\sum_i \Bigg[- \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1, Y_g=0} a_{i\mathbf{Y}} \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_g=1} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_g} a_{i\mathbf{Y}} \\
&+ \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1, Y_g=1} \left[\frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_g} a_{i\mathbf{Y}} - a_{i\mathbf{Y}} \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_g=1} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_g} a_{i\mathbf{Y}} \right] \Bigg] \\
&= \frac{1}{ni}\sum_i \left[\sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1, Y_g=1} \left[\frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_g} a_{i\mathbf{Y}} \right] - \left[\sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} a_{i\mathbf{Y}} \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_g=1} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \nu_g} a_{i\mathbf{Y}} \right] \right]
\end{aligned}
$$

Next I find the derivative of the applicant share with respect to the nonlinear parameters θ_1:

$$
\begin{aligned}
\frac{\partial a_j}{\partial \theta_1} &= \frac{1}{ni}\sum_i \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \frac{\partial a_{i\mathbf{Y}}}{\partial \theta_1} \\
&= \frac{1}{ni}\sum_i \sum_{\mathbf{Y}\in\Upsilon_i \ni Y_j=1} \left[\frac{\partial V(\mathbf{Y}|s_i)}{\partial \theta_1} a_{i\mathbf{Y}} - a_{i\mathbf{Y}} \sum_{\mathbf{Y}\in\Upsilon_i} \frac{\partial V(\mathbf{Y}|s_i)}{\partial \theta_1} a_{i\mathbf{Y}} \right]
\end{aligned}
$$

This is the last derivative required to compute $\frac{\partial \nu}{\partial \theta_1}$ using the Implicit Function Theorem. This derivative is used in calculation of the standard errors for the application cost parameters, and to supply the gradient to the solver in the GMM estimation routine. In addition, the derivatives $\frac{\partial a_j}{\partial \nu_j}$ and $\frac{\partial a_j}{\partial \nu_g}$ $g \neq j$ are used to create

the gradient in the Newton's method iterations discussed in Subsection 5.3.

C.2 Derivatives of Enrollment

Derivatives of the enrollment share equation are also necessary to supply to the gradient of the solver, and to compute standard errors. I derive these here. In estimation, the total probability of enrollment in a state is simulated across individuals i, so the within-state enrollment share is given by:

$$e_j = \frac{1}{ni} \sum_i \mathbb{P}(Enroll_{ij}|s_i, \Upsilon_i),$$

where

$$\mathbb{P}(Enroll_{ij}|s_i, \Upsilon_i) = \sum_{\mathbf{Y} \in \Upsilon \ni Y_j = 1} \mathbb{P}(Apply_{\mathbf{Y}}|s_i, \Upsilon_i) \sum_{\mathbf{D} \in \mathbf{Y}} \mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) \mathbb{P}(Enroll_{ij}|\mathbf{D}).$$

The derivatives of the enrollment share wrt the application cost parameters θ_1 are:

$$\frac{\partial e_j}{\partial \omega} = \frac{1}{ni} \sum_{i \in l} \sum_{\mathbf{Y} \in \Upsilon \ni Y_j = 1} \frac{\partial \mathbb{P}(Apply_{\mathbf{Y}}|s_i, \Upsilon_i)}{\partial \theta_1} \sum_{\mathbf{D} \in \mathbf{Y}} \mathscr{P}(\mathbf{D}|s_i, \mathbf{Y}) \mathbb{P}(Enroll_{ij}|\mathbf{D})$$

Appendix D Solutions to the Public University FOCs

The pricing and selectivity FOCs can be solved together for $\alpha_5 + \omega_5$:

$$\frac{\left(\frac{\partial q_j^{is}}{\underline{\mathbf{s}}_j} + \frac{\partial q_j^{os}}{\partial \underline{\mathbf{s}}_j}\right)\left(\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{os}} q_j^{is} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} q_j^{os}\right) - \left[\frac{\partial q_j^{os}}{p_j^{os}} \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial \underline{\mathbf{s}}_j} q_j^{is} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial \underline{\mathbf{s}}_j} \frac{\partial q_j^{is}}{p_j^{is}} q_j^{os}\right]}{\left(\left[\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} \frac{\partial q_j^{is}}{\underline{\mathbf{s}}_j} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial \underline{\mathbf{s}}_j} \frac{\partial q_j^{is}}{p_j^{is}}\right] \frac{\partial q_j^{os}}{p_j^{os}} + \frac{\partial q_j^{os}}{\partial \underline{\mathbf{s}}_j} \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{os}} \frac{\partial q_j^{is}}{p_j^{is}}\right)} + p_j^{is} - p_j^{os}$$

$$= \alpha_5^{pub} + \omega_5$$

Derive an expression for the out-of-state per-student cost, $c_j^{os}(\mathbf{M_j})$, from the two pricing conditions:

$$\frac{\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{os}} \frac{\partial q_j^{is}}{p_j^{is}} (p_j^{is} - p_j^{os} - \alpha_5^{pub} - \omega_5) - (\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} q_j^{os} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{os}} q_j^{is})}{\left[\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} \frac{\partial q_j^{os}}{p_j^{os}} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{os}} \frac{\partial q_j^{is}}{p_j^{is}}\right]} = \Big((p_j^{os} + t_j) - (c_j^{os}(\mathbf{M_j}) + q_j(\alpha_2^{pub} + \boldsymbol{\alpha}_3^{pub}\mathbf{M_j}))\Big)$$

Majors FOCs:

$$\frac{1}{\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} q_j}\Bigg(\left[\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} \frac{\partial q_j^{is}}{M_{jk}} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial M_{jk}} \frac{\partial q_j^{is}}{p_j^{is}} + \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} \frac{\partial q_j^{os}}{M_{jk}}\right] ((p_j^{os} + t_j) - (c_j^{os}(\mathbf{M_j}) + q_j(\alpha_2^{pub} + \boldsymbol{\alpha}_3^{pub}\mathbf{M_j})))$$

$$- \left[\frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial p_j^{is}} \frac{\partial q_j^{is}}{M_{jk}} - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial M_{jk}} \frac{\partial q_j^{is}}{p_j^{is}}\right] (p_j^{os} - p_j^{is} + \alpha_5^{pub} + \omega_5) - \frac{\partial \tilde{s}(\underline{\mathbf{s}}_j)}{\partial M_{jk}} q_j^{is}\Bigg)$$

$$= \alpha_{1k}^{pub} + \alpha_{3k}^{pub} q_j + 2\alpha_{4k}^{pub} M_{jk} + \omega_{jk}$$

Selectivity FOC:

$$\frac{\frac{\frac{\partial \mathcal{R}(p_j^{is}, p_j^{os})}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j^{is}, p_j^{os}) - \mathcal{C}(\mathbf{M_j})}}{\left(\frac{\frac{\partial \mathcal{R}(p_j^{is}, p_j^{os})}{\partial \underline{\mathbf{S}}_j} - \frac{\partial \mathcal{C}(\mathbf{M_j})}{\partial \underline{\mathbf{S}}_j}}{\mathcal{R}(p_j^{is}, p_j^{os}) - \mathcal{C}(\mathbf{M_j})} - \frac{\frac{\partial \tilde{s}(\underline{\mathbf{S}}_j)}{\partial \underline{\mathbf{S}}_j}}{\tilde{s}(\underline{\mathbf{S}}_j)}\right)} = \lambda_0^{pub} + \epsilon_j^s$$

Appendix E Tables

Table 1: CIP Major Examples

Major Category	CIP Code and Description
Arts, Social Sci. and Humanities	54.0101 History, General.
Arts, Social Sci. and Humanities	54.0102 American History (United States).
Arts, Social Sci. and Humanities	54.0103 European History.
Arts, Social Sci. and Humanities	54.0104 History and Philosophy of Science and Technology.
Arts, Social Sci. and Humanities	54.0105 Public/Applied History.
Arts, Social Sci. and Humanities	54.0106 Asian History.
Arts, Social Sci. and Humanities	54.0107 Canadian History.
Arts, Social Sci. and Humanities	54.0199 History, Other.

NOTE: The table provides examples of majors that can be offered within social sciences and history, to demonstrate the level of detail in the major classification. This is not an exhaustive list of majors.

Table 2: Majors

Category	# Majors	Code	CIP Title
Arts, Soc. Sci. and Humanities	373	4	ARCHITECTURE AND RELATED SERVICES.
		5	AREA, ETHNIC, CULTURAL, GENDER, AND GROUP STUDIES.
		16	FOREIGN LANGUAGES, LITERATURES, AND LINGUISTICS.
		22	LEGAL PROFESSIONS AND STUDIES.
		23	ENGLISH LANGUAGE AND LITERATURE/LETTERS.
		24	LIBERAL ARTS AND SCIENCES, GENERAL STUDIES AND HUMANITIES.
		30	MULTI/INTERDISCIPLINARY STUDIES.
		38	PHILOSOPHY AND RELIGIOUS STUDIES.
		39	THEOLOGY AND RELIGIOUS VOCATIONS.
		42	PSYCHOLOGY.
		45	SOCIAL SCIENCES.
		50	VISUAL AND PERFORMING ARTS.
		54	HISTORY.
Business and Comm.	132	9	COMMUNICATION, JOURNALISM, AND RELATED PROGRAMS.
		10	COMMUNICATIONS TECHNOLOGIES/TECHNICIANS AND SUPPORT SERVICES.
		52	BUSINESS, MANAGEMENT, MARKETING, AND RELATED SUPPORT SERVICES.
Eng., Math, and Sci.	305	11	COMPUTER AND INFORMATION SCIENCES AND SUPPORT SERVICES.
		14	ENGINEERING.
		15	ENGINEERING TECHNOLOGIES AND ENGINEERING-RELATED FIELDS.
		26	BIOLOGICAL AND BIOMEDICAL SCIENCES.
		27	MATHEMATICS AND STATISTICS.
		40	PHYSICAL SCIENCES.
		41	SCIENCE TECHNOLOGIES/TECHNICIANS.
Other	723	1	AGRICULTURE, AGRICULTURE OPERATIONS, AND RELATED SCIENCES.
		3	NATURAL RESOURCES AND CONSERVATION.
		12	PERSONAL AND CULINARY SERVICES.
		13	EDUCATION.
		19	FAMILY AND CONSUMER SCIENCES/HUMAN SCIENCES.
		21	TECHNOLOGY EDUCATION/INDUSTRIAL ARTS.
		25	LIBRARY SCIENCE.
		28	MILITARY SCIENCE, LEADERSHIP AND OPERATIONAL ART.
		29	MILITARY TECHNOLOGIES AND APPLIED SCIENCES.
		31	PARKS, RECREATION, LEISURE, AND FITNESS STUDIES.
		33	CITIZENSHIP ACTIVITIES.
		34	HEALTH-RELATED KNOWLEDGE AND SKILLS.
		36	LEISURE AND RECREATIONAL ACTIVITIES.
		43	HOMELAND SECURITY, LAW ENFORCEMENT, FIREFIGHTING AND RELATED PROTECTIVE SERVICES.
		44	PUBLIC ADMINISTRATION AND SOCIAL SERVICE PROFESSIONS.
		46	CONSTRUCTION TRADES.
		47	MECHANIC AND REPAIR TECHNOLOGIES/TECHNICIANS.
		48	PRECISION PRODUCTION.
		49	TRANSPORTATION AND MATERIALS MOVING.
		51	HEALTH PROFESSIONS AND RELATED PROGRAMS.

NOTE: The table provides the total number of majors (six-digit CIP codes) in each of the four categories that I use in my analysis. I also list the two-digit CIP codes that were grouped to create each category.

Table 3: Data Example: Score Reports Sent by CT High School Graduates

State of Residence	University	University State	Score Reports Sent	% of HS Grads
CT	UNIV OF CONNECTICUT	CT	12,362	27.86
CT	CENTRAL CONNECTICUT STATE UNIV	CT	5,156	11.62
CT	EASTERN CONNECTICUT STATE UNIV	CT	4,315	9.73
CT	SOUTHERN CONNECTICUT STATE UNIV	CT	4,198	9.46
CT	QUINNIPIAC UNIV	CT	3,631	8.18
CT	UNIV OF RHODE ISLAND	RI	3,081	6.94
CT	NORTHEASTERN UNIV	MA	3,034	6.84
CT	WESTERN CONNECTICUT STATE UNIV	CT	2,889	6.51
CT	UNIV OF MASSACHUSETTS AMHERST	MA	2,542	5.73
CT	BOSTON UNIV	MA	2,455	5.53
CT	UNIV OF HARTFORD	CT	2,348	5.29
CT	UNIV OF VERMONT	VT	2,087	4.70
CT	UNIV OF NEW HAVEN	CT	1,928	4.35
CT	FORDHAM UNIV	NY	1,793	4.04
CT	BOSTON COLL	MA	1,518	3.42

NOTE: This table shows the number of score reports sent and the percent of high school graduates sending reports to the top 15 universities among high school graduates from Connecticut in 2013.

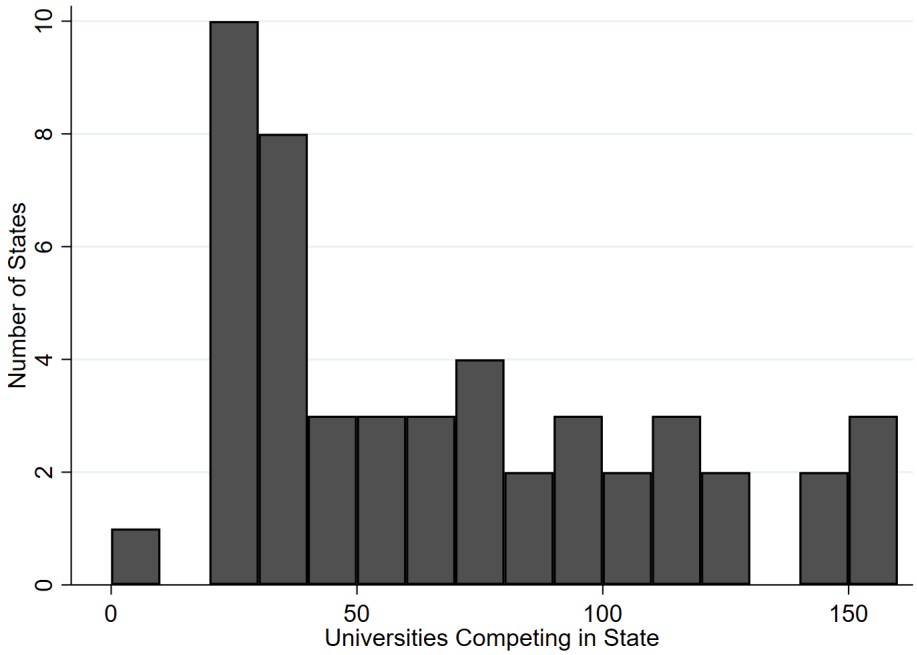
Table 4: Summary of Application and Enrollment Data

	Application Penetration	Enrollment Share
Mean	0.02171	0.00440
Min	0.00150	0.00005
10th Percentile	0.00212	0.00025
25th Percentile	0.00372	0.00041
50th Percentile	0.00811	0.00091
75th Percentile	0.01847	0.00278
90th Percentile	0.04726	0.00943
Max	0.42951	0.15634

NOTE: This table shows the mean, min, max and percentiles of the enrollment and application variables used to estimate the college choice model. The application penetration for an institution is the proportion of high school graduates from a state market who sent a score report to the institution. The enrollment share is the proportion of high school graduates who enrolled at the institution.

Appendix F Figures

Figure 1: Number of Universities Competing in Each State Market



NOTE: This graph shows the distribution of the number of universities competing in each market.