College Admissions and the (Mis)Allocation of Talent*

Ricardo Marto[†]
University of Pennsylvania

Yaacov Wittman[‡]
University of Pennsylvania

Job Market Paper

October, 2021

Click here for the latest version

Abstract

Empirical studies have found that high-achieving, low-income students are less likely to apply to selective colleges despite the generous financial aid typically offered. To reconcile this seeming puzzle, we build a model of the college market featuring tuition discrimination and a decentralized admissions system. Students, who differ in their financial resources and innate ability, apply to a subset of colleges and are uncertain about their prospective admissions and financial aid. Colleges observe only a noisy signal of student ability, and compete by choosing admissions standards and tuition schedules. We find that differences in application rates are due to student expectations over admissions and financial-aid, which are consistent with college policies in equilibrium. We address the puzzle by finding that low-income students receive high financial-aid at selective colleges because only the highest-ability among them apply, making their signals highly informative. If signals became less informative (e.g. colleges stopped using the SAT), all high-ability students would be worse off and only low-ability, high-income students would modestly benefit. Finally, we find overall welfare gains from increasing Federal need-based financial-aid, which would greatly benefit low-income, high-ability students by alleviating credit constraints.

Keywords: College Market, Tuition Discrimination, Admissions, Sorting, Credit Constraints
JEL Classifications: E21, I23, D58, D83

---

*We thank Damien Capelle, Jesús Fernández-Villaverde, Jeremy Greenwood, Dirk Krueger, and Víctor Ríos-Rull for their insightful comments and suggestions.

[†]Department of Economics, University of Pennsylvania. Email address: rmarto@sas.upenn.edu.
[‡]Department of Economics, University of Pennsylvania. Email address: ywittman@sas.upenn.edu.

1 Introduction

Investment in college education is important for human capital accumulation, earnings growth, and can be a powerful tool for social mobility. Despite the benefits of completing college, however, there are well documented gaps in college outcomes across the parental income distribution. Students born to parents from the bottom quartile of family income are much less likely to complete college than students from the top of the distribution (Bailey and Dynarski 2011) and are less likely to be represented at more selective colleges (Chetty et al. 2020). While these gaps can partly be explained by differences in levels of preparedness, gaps remain even for high-achieving students.

One explanation for these income gaps is that they arise from differences in application patterns between low-income and high-income students. Empirical studies such as Hoxby and Avery (2014) and Dillon and Smith (2017) find that there is substantial undermatching at the application stage: low-income students tend to apply to schools that they appear overqualified for relative to their higher-income peers. That is, low-income students appear to be under-represented in the college market not because they are excluded from the colleges directly but because they do not apply in the first place. This finding is troubling because it suggests that student misallocation may exist not only at the extensive margin of college enrollment, but also at the intensive margin of college choice. Moreover, as Hoxby and Avery (2014) note, the gap in application rates seems puzzling since many colleges (especially the most selective ones) provide substantial amounts of need-based financial-aid to low-income students.

Why then, even after controlling for test scores, do we observe different application patterns for students across the income distribution? Our paper examines the hypothesis that colleges effectively limit the share of low-income students through their admissions and financial-aid policies. It is then rational for low-income students to apply at lower rates since they may not expect to receive sufficient financial-aid. To study this hypothesis, we build and estimate a novel model of the college market featuring an application and admission system similar to the one used by U.S. colleges. Our model recognizes that while colleges would like to enroll all high-ability students, it is expensive for them to offer generous financial-aid to their low-income applicants: they face a tradeoff between admitting high-ability, low-income students and admitting lower-ability, high-income students who are willing to pay higher tuition. This trade-off will cause colleges to offer generous financial-aid only to the highest-ability students among their low-income applicant pool. Thus, low-income students will find it optimal to apply at lower rates because they have a lower chance of receiving sufficient financial-aid.

Our paper proceeds in two parts. First, we empirically study student application and enrollment patterns using the High School Longitudinal Study (HSLS). We learn that low-income students are indeed less likely to apply to highly selective colleges even when they have high test scores, confirming the findings of the previous literature. Moreover, we find that applicants face substantial risk in not receiving enough financial-aid after being admitted: almost a quarter of students indicated that they did not attend their preferred college due to costs. Second, we build an equilibrium model of the college market featuring a noisy application and admissions system and estimate it using the HSLS. We show that our model is able to account quantitatively for the application and enrollment patterns observed in the data, with realistic tuition schedules that vary based on students' parental income and test-scores. We use the model to study the role of the admissions system in shaping the allocation of students in the college market, and study the effect of higher education policies in our environment. We find that selective colleges offer high financial-aid to low-income students because only the highest-ability among them apply, making the colleges confident that their low-income applicants are more likely to be higher ability. Additionally, we find that making applications less informative would lower merit-based financial-aid and reduce admissions standards, hurting all high-ability students and modestly benefiting low-ability, high-income students.

In the empirical part of the paper, we use detailed student-level data from the HSLS to provide descriptive evidence on college application and enrollment patterns, and how they vary by parental income and student test scores. We show that both parental income and test-scores are important predictors in determining students' application behavior, even when controlling for other demographic and economic characteristics. This is true both at the extensive margin, where we see if students applied at all to any 4-year colleges, and at the intensive margin, where we see if students included any highly-selective colleges in their application portfolios. For example, descriptive analysis of our data suggests that for students with median household income and SAT scores in the 90th percentile, a one standard deviation increase in parental income is associated with a 10% increase in the probability of applying to a selective college. Moreover, we use our data to show that there is substantial uncertainty not only in admissions, but also in financial-aid. We find that almost a quarter of all students admitted to both selective and non-selective colleges do not attend their top choice because of high tuition costs. This risk of not receiving sufficient financial-aid even after being admitted is more likely to occur to low-income students, even if their SAT scores are within the top 10% of the distribution.

Motivated by these facts, we then develop an equilibrium model of the college market featuring student heterogeneity, college competition, and a decentralized admissions system. Colleges differ by the value they add to students in the labor market. Students, who differ by their innate ability and parental transfers, choose to apply to a subset of colleges or not apply at all. Applications are costly, and if a student decides to apply they then send a noisy signal of their ability to the colleges. Each college has its own threshold for the minimum acceptable signal necessary for admissions, and each college charges a tuition level that varies based on a student's financial resources and signal. Thus admissions *and* financial-aid are risky for the student: they make their application decisions based on an expectation over the possible realizations of the signal. Finally, once the signal is realized and students have offers from the colleges they were accepted to, they decide which college to enroll in (if any).

On supply side of the market, there are a discrete number of colleges who differ in their technology, endowment income, and costs. The objective of the colleges is to maximize the value they add to their students, which depends on the level of instructional spending per student and the average ability of the student body. To capture the uncertainty of the admissions process, we assume that colleges are unable to observe the true ability of the students in their pool of applicants but instead observe only the noisy signal mentioned above. Thus, the admissions system is modeled as a signal extraction problem for the colleges. To maximize their objective, colleges choose their minimum acceptable signal (admissions standard) and tuition schedules. When colleges compete, they take as given the admissions standards and tuition schedules of the other colleges. Thus, their pricing decisions are limited by the students' outside options if their applicant pool is likely to contain students who applied to multiple colleges.

Combining the signal extraction problem of the college with their ability to price discriminate is new to our model, and introduces an important mechanism that influences the degree of price discrimination due to the signal. Colleges can fully observe a student's parental transfer, and while they do not observe ability directly, we assume they know the joint distribution of income and ability among their applicants (they take student application choices as given in equilibrium). Thus, if only the highest-ability low-income students choose to apply, signals from low-income students will be more informative to the college because they are more likely to have come from high-ability students. Colleges will then offer high financial-aid to the low-income students they enroll because they are confident that such students are high-ability. This mechanism helps explain why we jointly observe low application rates and high financial-aid for low-income students.

We estimate our model using various student and college level data sources available from the National Center for Education Statistics (NCES) which provide detailed student-level data on applications, enrollment, and financial-aid, and college-level data on revenue, expenses, and enrollment. Empirically, we find that the model is able to account for the different application and enrollment rates observed across the parental income distribution. It is also able to capture the variation in tuition both across and within colleges, with low-income students receiving significant levels of financial-aid at selective colleges. In estimating the parameters of our model, we find that non-pecuniary costs of education are substantially lower than what is typically found in the literature. Much of the literature has argued that such 'psychic' costs of schooling are necessary to explain why many students do not pursue a college education despite the financial returns from doing so. By fully accounting for the application and admissions process, our model is able to rationalize that many students choose not to participate in the college market not beacuse they have a strong preference against it, but because they understand that they are unlikely to be admitted or receive sufficient financial-aid.

Next, we use the model to study the effect of the application and admissions system on student allocation in the college market. We first examine the effect of low-income student application decisions in order to account for the importance of signal informativeness in our baseline equilibrium. We solve for the equilibrium of a counterfactual economy where low-income students apply as much as their higher income peers. This introduces low-ability students to the low-income applicant pool, which makes the signals of the high-ability, low-income students less informative. Selective colleges will then reduce the financial-aid for all low-income students since the applicant pool has worsened. We find that this effective increase in tuition is large, and lowers low-income student enrollment at the selective colleges by about a quarter. Overall, this finding shows that high-ability, low-income students benefit greatly from more informative signals due the lower application rates of their lower-ability peers. Thus, interventions that encourage low-income students to apply may be harmful for their overall enrollment if the interventions are not targeted by ability.

We also study the effects of application signals becoming *less* informative, motivated by many colleges' decision to pause their use of standardized tests in the admissions process during the Covid-

19 pandemic. To do this, we recalculate a counterfactual equilibrium where we increase the noise associated with student applications, making it more difficult for colleges to infer the students' true ability. We find that in the new equilibrium, colleges to endogenously reduce their admissions standards and merit-based financial-aid. Students at the top of the ability distribution are then made worse off from the change since they now have a lower chance of being admitted to the colleges. High-ability, low-income students are particularly harmed because they also receive less financial-aid. The only students who have modest gains from the less informative signals are the low-ability, high-income students who now find it easier to match with the colleges. Low-ability, low-income students are largely unaffected because the gains they experience from increased admissions are offset by the losses they experience from lower financial-aid. Overall, the gains that accrue to the low-ability students are not enough to offset the losses to the high-ability students who benefit the most from attending college, leading to overall welfare losses.

Finally, we study the effects of a large expansion in the Pell Grant program, which would increase the amount of federal grant funding to low-income students, and make middle-income students eligible for federal aid. We find that the policy change is most beneficial to high-ability, lower-income students who were before less likely to enroll due to credit constraints. Conditional on ability, the increase in grants flattens the student enrollment across the parental income distribution. This considerably reduces the concentration of the income distribution in selective colleges, but gaps still remain since income and ability are correlated. In terms of welfare, we find that higher-income students lose from the change due to the higher tax rates and increased competition with lower-income students. The degree to which high-income students lose, however, depends on their ability: those with high ability lose less since it is easier for them to compete with the newly unconstrained low-income students. Overall, we find that the policy has a net-positive effect on welfare. The value of a college education is higher for the high-ability, low-income students who benefit from the policy compared to the lower-ability, high-income students who lose from it.

Related literature. This paper builds on three different strands of the literature. The first is the large empirical literature documenting inequality in higher education and the role of applications and the admissions system. The second relates to the structural literature on the college market and the admissions problem. The third relates to the literature which studies the distributional effects of education policies.

Our paper is complementary to the empirical literature on the college market and its outcomes. Recent work by Chetty et al. (2020) document a large degree of income segregation within and across U.S. colleges. Consistent with our empirical facts and model, they find that students from wealthier backgrounds are disproportionately represented at more selective schools. They also find that more selective colleges give higher returns to education, and estimate the causal effect of an individual college in earnings to be around 80%. Differences in application behavior related to income differences or ability have also been documented by Hoxby and Avery (2014), Dillon and Smith (2017), Delaney and Devereux (2020). Dynarski et al. (2018) further study the role of expectations about tuition and admission at the application stage. They conduct an experiment where low-income high-achieving high school students are informed that they will be offered free tuition if admitted at the University of Michigan. They find large increases application and enrollment, consistent with our findings of the importance of tuition in driving these decisions. Our empirical analysis on tuition discrimination also relates to Fillmore (2020), who studies the effect of different FAFSA information disclosure policies on tuition levels.

Our model builds on the work of Epple et al. (2006, 2017), who consider quality-maximizing colleges that price discriminate among students to study the effect financial aid policies. Gordon and Hedlund (2016) build on that framework to study the rise in college tuition, showing that demand forces helps explain most of this increase. Similar to these papers, we assume that colleges compete monopolistically and choose a tuition schedule to maximize their value-added, which depends on the composition of their students. Importantly, our model also draws on Chade et al. (2014), who introduce matching frictions in college admissions problem, and allow students to make multiple college applications. Our model also complements Fu (2014), which jointly models tuition and admissions, but ours adds the important margin of heterogeneity in parental income and credit constraints. Several recent papers have also studied the college market and its interaction with inequality and intergenerational mobility. Cai and Heathcote (2018) study the role of income inequality in explaining the rise in tuition using a novel model that gives rise to an endogenous distribution of colleges. Using a similar framework, Capelle (2019) studies the role of the college market in shaping intergenerational mobility for heterogeneous students. Our paper includes realistic features of the application and admissions problem, and studies their effect on the college market.

Finally, this paper is also related to the literature on the macroeconomic effects of education policies. Several papers have modeled and quantified the effect of policies on school choice, inequality, or labor market returns (Fernandez and Rogerson (1996), Bénabou (2002), Lochner and Monge-Naranjo (2011), Ionescu and Simpson (2016), Krueger and Ludwig (2016), Kotera and Seshadri (2017), Caucutt and Lochner (2017), Abbott et al. (2019), Colas et al. (2021)). In particular, our paper complements the analysis of Abbott et al. (2019), who study the effect of financial aid policies and intergenerational transfers on welfare, Ionescu and Simpson (2016) and Lucca et al. (2018), who examine policy changes in student loan limits on college enrollment and tuition, as well as Krueger and Ludwig (2016), who analyze the optimal mix of tax and education subsidies and their impact on human capital accumulation.

Outline. The remainder of the paper is organized as follows. Section 2 presents evidence on application and enrollment patterns among students transitioning from high school to college, Section 3 describes our model of the college market, Section 4 presents our calibration and estimation procedure, Section 5 and 6 discuss our results, Section 7 provides counterfactual policy analysis, and Section 8 concludes.

2 Empirical Evidence

We empirically study the college application and enrollment behavior of U.S. high-school students by using detailed micro-level data combined with institutional college level data. We first describe the datasets we use and then document stylized facts on student application and enrollment patterns.

2.1 Data

We use the High School Longitudinal Study of 2009 (HSLS) to study high-school students' outcomes in the college market. Published by the National Center for Education Statistics (NCES), the HSLS consists of a nationally representative sample of more than 23,000 ninth graders from 944 high schools who are followed throughout their secondary and postsecondary education. The students and their parents are first interviewed in 2009, then again in 2013 once the students have graduated from high-school, and then once more in 2016. The dataset includes rich information about students' test scores, college application and enrollment behavior, as well as demographic and economic characteristics. Our analysis relies on the restricted-use version of the HSLS, which also provides student SAT and ACT scores, lists of colleges applied to and enrolled in, and more detailed information about household economic variables.

Our goal is to study how the students' college application and enrollment decisions vary based on their parental income and college preparedness. Household income is provided directly by the parents in the HSLS survey, where they are asked for their households' income from all sources in 2011. For college preparedness, we use the students' high-school GPA and their SAT or ACT scores[1]. Each student's GPA is reported directly by their high-school, and is honors-weighted in a procedure by the NCES used to make the GPA comparable across different high-schools. The SAT is reported directly by the students' colleges, and is therefore unavailable for students who did not attend college or did not take the test. For these students, we simply impute the test scores using the mean SAT score of other HSLS students from the same parental income group, level of parental education, race, and GPA decile.

Since we are interested in the composition of student application portfolios, we follow Chetty et al. (2020) by using the Barron's selectivity index (Profile of American Colleges, 2015) to categorize colleges into different groups. We restrict our focus to all four-year non-profit colleges and use two groups. The first group, which we call "Highly-selective" colleges, corresponds to Barron's Tier 1 and 2 colleges and universities. This group makes up 192 schools. For reference, a list of all colleges and universities in this group is provided in Appendix B.1. All other four-year non-profit colleges and universities are counted in the second group, which we refer to as "Non-selective". In our HSLS sample, we find that the Highly-selective colleges account for about 16% of all four-year, non-profit enrollment (representing about 7% of students in the overall sample).

The HSLS also includes detailed information about each student's application behavior, which we use to categorize the students' application portfolios. In the follow up survey after completing high-school, students were asked to provide the college they were currently attending, and also to list two other colleges they applied to and were seriously considering. Additionally, students were asked to provide the *total* number of applications they sent. The answers to these questions then help us determine the strength of each student's application portfolio. While the data include only the three most relevant colleges in the portfolio, most students indicated that they had applied to three or fewer

[1]The NCES converts the ACT score into an equivalent SAT score for students who only took the ACT instead of the SAT. Henceforth, whenever we mention the SAT, we refer to either the SAT or ACT score.

schools, suggesting that the HSLS gives us a good picture of the total application portfolios.

If we do not observe any non-profit four-year applications among the three colleges the students applied to, we consider the student as not having applied. If we observe at least one non-profit four-year application but none of the three schools are Highly-selective, we consider the student as only applying to Non-selective colleges. If we see at least one Highly-selective and one Non-selective application, we consider the student has having applied to both. Finally, if we see only Highly-selective applications, we count these as having applied to both only if the number of *total* applications is greater than the number of *listed* applications (that is, we assume such students also applied to a Non-selective school as a safety).

2.2 College Attendance in the HSLS

We first use the HSLS to document how different students apply and sort into different types of colleges. The left panel of Figure 1 below shows how enrollment rates vary across different levels of parental income and SAT score quartiles[2]. We see that conditional on test score quartile, student outcomes are correlated with parental income both at the extensive margin of college attendance, and the intensive margin of college selectivity. For example, even among students in the top test score quartile, those with parental income below the median ($55,000) were more than twice as likely not to attend 4-year non-profit colleges as students in the top parental income group. The data also reveal a similar pattern for college selectivity: students from high-income families are significantly more likely to attend Highly-selective colleges. For students in the top quartile of SAT scores, we again see that those in the highest income group are about twice as likely to attend a Highly-selective college compared to students in the bottom group.

Next, we turn to the application behavior of the students in the sample, which is presented in the right panel of Figure 1. This plot shows the fraction of students from each parental income and SAT group who do not apply at all, as well as the fractions who apply to the Highly-selective colleges, the Non-selective colleges, or both. Looking at applications as opposed to enrollment from the left panel of Figure 1 we see that the patterns are very similar, suggesting that the gaps in college enrollment are due to differences in application rates between low- and high-income students. Thus in order to explain differences in enrollment, it is necessary to understand student application decisions.

2.3 Analysis of College Application Behavior

Motivated by the importance of college application behavior in explaining the differences in college enrollment, this subsection studies whether parental income and test scores have predictive power in

[2]We present the same figures using the students' GPA instead of SAT scores in Figure A.1. Note that we do not group parental income into quartiles because the surveyed parents in our data reported their income by selecting from a group of ranges (i.e. less than $15,000, $15,000 to $35,000, $35,000 to $55,000, etc.). The percentages for each grouping are too coarse to fit neatly into quartiles or quintiles. For reference, the parental income distribution from our HSLS sample is provided in Figure B.5.

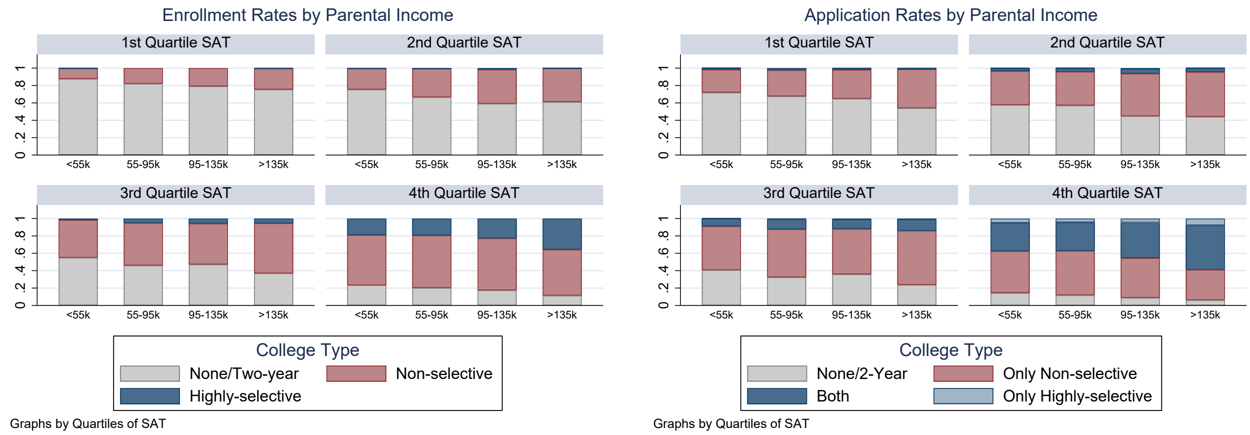


Figure 1: College application and enrollment by parental income and student SAT/ACT score. Left panel: Student enrollment. Right panel: Student applications.

determining application behavior while controlling for other factors. We run two basic logit models using our HSLS student-level data.

In the first logit model we consider, we study the factors correlated with the extensive margin of the choice to apply, where the dependent variable is an indicator for whether or not the student decided to apply to college at all. Table A.1 shows the results of the estimation, where only the coefficients for parental income and test scores are shown. The table shows that regardless of whether we measure student preparedness by GPA or SAT score (which are standardized in the estimation), we see that both test scores and parental income are strongly correlated with a student's decision to apply to college. This relationship remains after for controlling for other related variables including race, sex, parental education level, and high-school level characteristics. To aid in interpreting the table, we plot the predicted application probabilities implied by the logit estimates for different parental income levels and SAT scores in Figure A.2. We see that while the relationship between parental income and applications declines for higher test-scores, there is still a strong effect for average students with median SAT scores.

We next turn to the intensive margin by studying how parental income and test scores are correlated with a student's application portfolio choice. Restricting our sample only to students who apply at all, the dependant variable in our next set of logit models is an indicator for whether or not the student included a Highly-selective college in their application portfolio. This includes students who applied to both Highly-selective and Non-selective colleges, and those who applied only to Highly-selective colleges as discussed in the previous section. Table A.2 shows the estimation results from these logit models, which again are estimated separately for GPA and SAT scores using our HSLS sample.

Table A.2 reveals similar patterns for a student's decision to include at least one Highly-selective college in their portfolio. We see that test scores are strongly correlated with the decision, and parental income by itself is insignificant while the interaction between test scores and income is highly significant. These relationships remain even after controlling for various student characteristics. We plot the predicted probability of applying to a Highly-selective college for different SAT scores and parental income levels in Figure 2. Since income by itself is insignificant, we see that there is little difference across the income distribution in application rates to Highly-selective colleges for students with relatively low test scores. However, since the interaction between SAT and income is significant, we see a strong correlation between parental income and the decision to apply to a Highly-selective college for students at the top of the SAT score distribution. For example, an increase in family income of about $80,000 is associated with roughly a 10% increase in the probability that an average student with a 90th percentile SAT will apply to a Highly-selective college. Overall, these estimates show that parental income is an important factor in determining not only whether or not a student will apply, but also which types of colleges they decide to target in their applications.

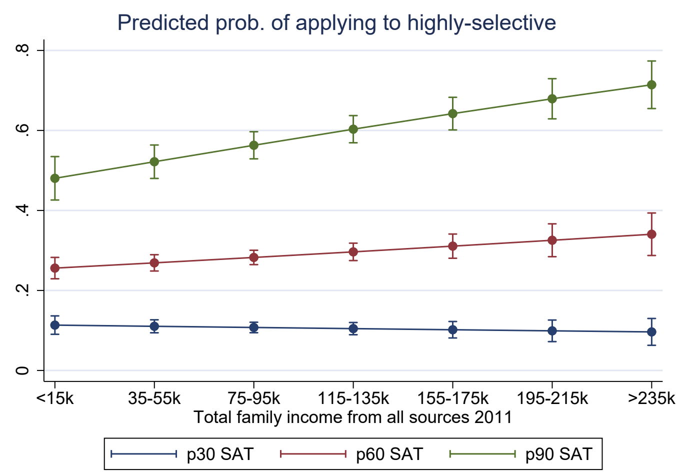


Figure 2: Predicted relationship between SAT, parental income, and the decision to apply to a Highly-selective college implied by our estimated logit model. See Table A.2 for estimation results. See Figure A.3 for the same plot using GPA.

2.4 Enrollment Conditional on Admission

Finally, we provide evidence that financial-aid is risky for the students who apply. Even if they are admitted to their top choice among the colleges they applied to, students may choose to enroll in a different college if they do not receive sufficient financial-aid. Using the HSLS, we study how students sort into colleges conditional on being accepted and examine the role of costs in guiding their decision. Since students reported whether or not they were admitted to each college they listed from their application portfolio, we are able to see how students decided to enroll conditional on being admitted. Table 1 reports the enrollment choices for students who were admitted to both Highly-selective and Non-selective colleges, and those who were admitted to only one type. We see that for students accepted to both types of colleges, 39% choose to go to a Non-selective college despite being admitted to a Highly-selective college, and for students with the option of attending a Non-selective college, 16% choose not to attend.

Importantly, the HSLS also allows us to understand the extent to which these ex-post decisions are driven by tuition costs rather than preferences. In the survey, students were asked to identify the college in their application portfolio that they would choose to attend if not for costs. Using these survey responses, we show in the bottom row of Table 1 that a significant fraction of students would have chosen a different college they were admitted to if not for costs. Note that this fraction is largest among students who were accepted to both types of colleges, suggesting that many students choose not to attend Highly-selective colleges even after being admitted because they do not receive sufficient financial aid.

	Accepted to...		
Enrollment choice	Both	Non-selective Only	Highly-selective Only
No College/ 2-Year	3.1%	16.2%	3.0%
Non-selective	38.9%	83.8%	–
Highly-selective	58.0%	–	97.0%
% Not attending preferred college because of costs	23.5%	21.6%	3.5%

Table 1: Enrollment patterns among admitted students in the High School Longitudinal Study.

Finally, we show that low-income students are more likely not to attend their preferred college due to costs. We restrict our HSLS sample to students who were accepted to both Highly-selective and Non-selective colleges and estimate a probit model where the dependent variable is an indicator equal to one if the student was not attending their preferred college due to costs and zero otherwise. Our independent variables include family income, race, sex, parental education, and high-school characteristics. We illustrate our estimation results in Figure A.4, where we see that parental income is strongly correlated with the likelihood that a student does not attend their preferred college due to costs. This relationship is still evident even when we restrict our sample to students in the top of SAT score distribution, though our estimates become less precise since there are fewer high scoring low-income students. Only when the SAT threshold is set to 1400 (the top 2% of test scores in our sample) do we find that the relationship with parental income disappears. Overall, these results suggest that low-income students face uncertainty not only in terms of whether or not they will be admitted to the colleges they apply to, but also whether or not they will receive sufficient financial aid.

To summarize, we find that there is a strong correlation between college attendance and parental income at both the extensive margin of overall enrollment, and the intensive margin of college selectivity. We find that these gaps are largely due to differences at the application stage, where low-income students are less likely to apply at all, and less likely to apply to Highly-selective colleges when they choose to apply. Finally, our analysis suggests that students face uncertainty over financial-aid, with many low-income students reporting that they do not attend their preferred college due to costs even after they are admitted. Motivated by the empirical findings of this section, we next turn to our model of college application and enrollment. We use the model to rationalize the facts observed in the data, and study the effect of the admissions system in shaping student outcomes in the college market.

3 Model

Overview. The economy is populated by a unit measure of heterogeneous individuals, two colleges of different types, and a government. Individuals live for two periods: young and old. Young students start life with parental transfer y and ability level ℓ, and decide whether they want to work or invest in their human capital by attending college. If they decide to attend college, they must choose a subset of colleges to apply to. Admissions, however, are risky since colleges can only observe a noisy signal σ of the student's true ability ℓ. We assume that a students' parental transfer is fully observable to colleges[3]. Students with a high enough realization of σ receive an offer of admission and a college-specific tuition level that depends on y and σ. Once the uncertainty is resolved and students know their admissions and financial-aid decisions, they choose which college to enroll in. At any stage, students may choose the outside option of working instead of going to college (for simplicity, we do not separately consider two-year colleges).

Colleges maximize the value-added they provide to their students on the labor market, denoted Γ_s. Value-added is taken as given by the students but determined endogenously by the colleges. A college's value-added will depend on its average instructional spending and the average ability of its student body. Each college competes by setting different admission standards and tuition schedules. Note that there is only one college per type, so competition occurs across types rather than within types. When colleges make an offer of admission, they take into consideration that students may have received other offers and decide not to enroll. This option value for students makes the choice of tuition depend not only on college-specific characteristics, but also on the pricing and admissions policy of the other college. Both colleges differ exogenously in their endowment income, efficiency, costs, and tuition caps. We refer to the elite college with the high endowment as College 1, and the other as College 2. Finally, the government taxes the working population to subsidize colleges and pay for grant programs that support low-income students.

Model timing. The timing of events in the first period is as follows:

1. Individuals choose either to apply to college or go straight to the labor market. Those who apply must choose an application portfolio which includes either or both colleges.

2. Colleges receive applications and choose which students to accept by setting their admission standards and tuition schedules.

3. Students make their attendance decision given the admission and financial-aid offers received.

4. Individuals make their consumption and savings decisions.

[3] In order to receive federal grants or loans students must complete the Free Application for Federal Student Aid (FAFSA), which states students' parental income and financial assets. The FAFSA is fully observable by the colleges that the student applies to, and we find that 77% of students in the HSLS complete it.

3.1 Student Application and Enrollment Decisions

We proceed in chronological order by first introducing the problem of an applicant, then the decision problem involving the acceptance and rejection of offers, and finally the problem of a matched student. The problem of a worker comes last.

Student's value from applying. A student who decides to apply must choose either to send one application to College $s \in \{1,2\}$, or apply to both. If a student applies they draw a realization of σ, which is a noisy signal of their ability ℓ and is unknown to the student at the time of their application decision. The signals are drawn from a continuous conditional density $g(\sigma|\ell)$ with cdf $G(\sigma|\ell)$ and support $[0,\infty]$. For simplicity, only one signal is drawn regardless of the application portfolio, so that both colleges observe the same signal[4]. Taking the admissions standard $\underline{\sigma}_s$ as given, a student is admitted to College s if $\sigma \geq \underline{\sigma}_s$. In the rest of the section, we focus (without loss of generality) on the case where College 1 is at least as selective as College 2 (i.e. $\underline{\sigma}_1 \geq \underline{\sigma}_2$)[5].

Applications are costly, and include a financial cost as well as a 'psychic' disutility cost meant to capture the effort of completing the applications [6]. We denote the disutility by $\phi_a,\ a \in \{1,2,B\}$, so that it depends on whether the student applies to College 1, 2, or both. For simplicity, we assume that the financial application cost does not depend on the application portfolio, and denote it by ψ. We include both types of costs since financial application costs alone are not enough to account for the observed application rates[7].

Let $V^{AB}(y,\ell)$ be the expected value from applying to both colleges, and $V^{As}(y,\ell)$ the expected value from applying to College s only, $s \in \{1,2\}$. The expected value of applying to both colleges is given by

$$
\begin{aligned}
V^{AB}(y,\ell) \quad = \quad & \int_{\underline{\sigma}_1}^{\infty} V^{OB}(y,\ell,T^1(y,\sigma),T^2(y,\sigma))g(\sigma|\ell)d\sigma + \int_{\underline{\sigma}_2}^{\underline{\sigma}_1} V^{O2}(y,\ell,T^2(y,\sigma))g(\sigma|\ell)d\sigma + \\
& G(\underline{\sigma}_2|\ell)V^W(y,\ell,1) - \phi_B, \qquad (1)
\end{aligned}
$$

where $V^{OB}(y,\ell,T^1(y,\sigma),T^2(y,\sigma))$ is the value of a student with both offers of admission in hand, each priced at $T^1(y,\sigma)$ and $T^2(y,\sigma)$ for College 1 and College 2 respectively. $V^{O2}(y,\ell,T^2(y,\sigma))$ is the value of a student with only College 2's offer of admission who was rejected by College 1, and $V^W(y,\ell,1)$ is the value of a worker (the "1" in the value function denotes that the student will still pay the financial application cost).

[4]This assumption, while made for tractability, is also supported empirically. We find that in the HSLS, 97% of students who were admitted to a highly-selective college (as defined in Section 2.1) were also admitted by a less-selective college (conditional on having applied to both).

[5]While it is possible for College 2 to have a higher admissions standard in equilibrium, we confirm that this is not the case in our baseline estimation and subsequent analysis.

[6]This cost includes the effort and lost time associated with writing essays and filling application forms, preparing for and taking the SAT (perhaps multiple times), or the time spent researching which colleges are worth applying to.

[7]In our estimation, we calculate the financial costs directly from the data so we can separately identify the non-pecuniary disutility costs.

The expected value from applying only to College s is given by

$$V^{As}(y,\ell) \quad = \quad \int_{\underline{\sigma}_s}^{\infty} V^{Os}(y,\ell,T^s(y,\sigma))g(\sigma|\ell)d\sigma + G(\underline{\sigma}_s|\ell)V^W(y,\ell,1) - \phi_s, \tag{2}$$

where $V^{Os}(y,\ell,T^s(y,\sigma))$ is the value of a student who only applied to College s and was admitted at price $T^s(y,\sigma)$.

Optimal application. The optimal application decision for a student with characteristics (y,ℓ) solves the following simple discrete choice problem:

$$\max\left\{V^{A1}(y,\ell), V^{A2}(y,\ell), V^{AB}(y,\ell), V^W(y,\ell,0)\right\}. \tag{3}$$

Student's value with offers of admission. After the signal is realized and students know their admissions and financial-aid status, they decide which offer, if any, to accept. We assume that admitted students draw idiosyncratic preference shocks ϵ^s over their available alternatives. These shocks are mean zero Type I extreme value shocks with scale parameter $\lambda_c > 0$. Including these shocks simplifies the analytical tractability of the model because they allow for closed form solutions of the students' enrollment functions, which are taken as given by the colleges. The shocks can be interpreted as allowing the students to change their mind about their college preference before enrolling for idiosyncratic reasons (e.g. campus visits, or seeing how their friends enroll).

The value of a student with both offers of admission in hand is the expected value from choosing between accepting College 1's offer, College 2's offer, or working,

$$\begin{aligned} V^{OB}(y,\ell,T^1(y,\sigma),T^2(y,\sigma)) \quad = \quad & \int \max\left\{V^{C1}(y,\ell,T^1(y,\sigma)) + \epsilon^1, V^{C2}(y,\ell,T^2(y,\sigma)) + \epsilon^2,\right. \\ & \left. V^W(y,\ell,1) + \epsilon^3\right\} dG_\epsilon, \end{aligned} \tag{4}$$

where $V^{Cs}(y,\ell,T^s(y,\sigma))$ is the value of attending College s and paying tuition $T^s(y,\sigma)$. The value of a student with only one offer of admission in hand is the expected value from choosing the maximum value between accepting College s's offer and working, i.e.

$$V^{Os}(y,\ell,T^s(y,\sigma)) \quad = \quad \int \max\left\{V^{Cs}(y,\ell,T^s(y,\sigma)) + \epsilon^s, V^W(y,\ell,1) + \epsilon^3\right\} dG_\epsilon, \tag{5}$$

for $s = \{1,2\}$. Given the extreme value shocks, the value functions simplify to

$$V^{OB}(y,\ell,T^1(y,\sigma),T^2(y,\sigma)) \quad = \quad \frac{1}{\lambda_c}\log\left(e^{\lambda_c V^{C1}(y,\ell,T^1(y,\sigma))} + e^{\lambda_c V^{C2}(y,\ell,T^2(y,\sigma))} + e^{\lambda_c V^W(y,\ell,1)}\right) \tag{6}$$

$$V^{Os}(y,\ell,T^s(y,\sigma)) \quad = \quad \frac{1}{\lambda_c}\log\left(e^{\lambda_c V^{Cs}(y,\ell,T^s(y,\sigma))} + e^{\lambda_c V^W(y,\ell,1)}\right) \tag{7}$$

Optimal acceptance. With the Type I extreme value shocks over the students' options conditional on offers of admission, we can solve for the probability that a student with characteristics (y,ℓ), signal

σ, and application strategy $i \in \{A1, A2, AB\}$ accepts the offer of College s given tuition levels and admissions thresholds. Given admissions thresholds $\underline{\sigma}_1, \underline{\sigma}_2$, the probabilities are given by

$$q_i^s(y,\ell,\sigma,T^1,T^2) \quad = \quad \begin{cases} \frac{\exp\{\lambda_c V^{Cs}(y,\ell,T^s)\}}{\sum\limits_j \exp\{\lambda_c V^{Cj}(y,\ell,T^j)\}+\exp\{\lambda_c V^W(y,\ell,1)\}} & \text{if } i = AB,\ \sigma \geq \underline{\sigma}_1 \\ \frac{1}{1+\exp\{\lambda_c[V^W(y,\ell,1)-V^{Cs}(y,\ell,T^s)]\}} & \text{if } i = As,\ \ \sigma \geq \underline{\sigma}_s, \\ & \text{or } i = AB,\ \ \underline{\sigma}_1 \geq \sigma \geq \underline{\sigma}_2,\ s = 2 \end{cases} \tag{8}$$

Note that the acceptance probability depends on both tuition levels only if a student chose to apply to both colleges and was admitted to both. Students who applied only to one college, or applied to both and were rejected by College 1 will therefore not have an outside option of attending a different college.

Student's value from attending college. Once students have accepted an offer of admission from a particular college, they face a two period consumption-savings problem. The first period corresponds to four years of college and the next period corresponds to the rest of their life. In the college period, the student must finance their consumption, tuition payment, and application costs using parental transfers, grants, and student debt. Students may borrow up to a limit denoted by $\underline{a}_s$[8]. Grants depend on the student's income and are separated into grants from outside sources, denoted $Gr(y)$, and grants that are paid for by the government, denoted $P(y)$. Outside grants are exogenous in the model and are meant to capture private scholarships or limited public scholarships. Grants from the government are paid for by taxes in the model, and are meant to capture the federal Pell Grant program.

After graduating college and moving to the next period, the student enters the labor market and receives a wage w_o net of taxes τ per unit of human capital acquired in college. The resulting human capital is given by $\Gamma_s \ell^{\alpha_s}$, which depends on the college's value added, Γ_s, and parameter α_s that governs the returns to ability. The student also pays her loans back priced at the interest rate R, and discounts the future with discount factor β. Finally, the student has preferences over consumption given by $u(c)$, and incurs an additional utility cost (or benefit if positive) for attending College s, captured by $\nu^s(\ell)$. Including this extra 'psychic' cost component is motivated by the literature which has found that pecuniary returns can account for only a part of observed college enrollment patterns. We allow this term to vary by ability as we expect the non-pecuniary costs of completing college to be lower for high-ability students.

[8]This borrowing limit is motivated by the existing limits to federal student loans imposed by the Department of Education. Consistent with the federal limit, the modeled borrowing limit does not depend on the student's earnings potential.

The value of an individual enrolled in College s, paying tuition T^s, is then given by

$$\begin{aligned} V^{Cs}(y,\ell,T^s) &= \max_{c,c',a'} u(c) - \nu^s(\ell) + \beta u(c') \qquad (9) \\ \text{s.t.} \qquad c + a' &= y - T^s - \psi + Gr(y) + P(y) \\ c' &= R\,a' + (1-\tau)w_o\,\Gamma_s \ell^{\alpha_s} \\ a' &\geq \underline{a}_s \end{aligned}$$

Denote by $a^{s'}(y,\ell,T^s)$ the student's borrowing or savings policy function if studying at College s. The student's optimal consumption-savings decision is then governed by the usual Euler equation

$$u_1(c^s) \;\; \geq \;\; \beta R\, u_1(c^{s'}), \qquad (10)$$

where $c^s(y,\ell,T^s)$ is the student's consumption. Note that there is a maximum tuition level the student can afford, which is given by

$$T_{max}(y) = y - \underline{a}_s - \psi + Gr(y) + P(y). \qquad (11)$$

Therefore, any tuition level $T^s(y,\sigma) \geq T_{max}(y)$ is automatically rejected by a student with parental transfers y, regardless of the value drawn for σ.

Value from working. Individuals end up as workers either by choosing not to apply, being rejected, or by choosing not to attend college conditional on an offer of admission. The income of an individual working without a college degree is $(1-\tau)w\ell_w^\alpha$, where $\alpha_w < \alpha_s$ reflects that higher ability individuals have a higher return from attending college. $V^W(y,\ell,n)$ is the value of an individual with parental transfers y and ability ℓ who submitted $n \in \{0,1\}$ college applications:

$$\begin{aligned} V^W(y,\ell,n) &= \max_{c,c',a'} u(c) + \beta u(c') \qquad (12) \\ \text{s.t.} \qquad c + a' &= y + (1-\tau)w_y\,\ell^{\alpha_w} - \psi \mathbf{1}_{\{n>0\}} \\ c' &= R\,a' + (1-\tau)w_o\,\ell^{\alpha_w}. \end{aligned}$$

3.2 Colleges

There are two colleges who compete for students by setting their tuition schedules and admission standard taking as given the tuition schedules, admission standard, and value-added of the other college. The choices of one college affects the other through the student enrollment function defined in (8). An equilibrium in the college market will then be a fixed point in the set of possible college policies: the optimal choice of each college will be consistent with the optimal choice of the other college.

The objective of each college is to maximize its value-added, which is given by

$$\Gamma_s \equiv \xi_s Q(I_\mu, L_\mu). \qquad (13)$$

ξ_s is the efficiency with which quality is transformed into value added and $Q(I_\mu, L_\mu)$ denotes the college's quality, which depends on the average amount of instructional spending per student I_μ and the average ability of the student body L_μ. The dependence of college quality on L_μ accounts for peer effects, where students benefit more from the college when its student body has a higher average ability. We also assume that Q is strictly increasing and differentiable, with $Q_{L_\mu} > 0$ for $L_\mu > 0$, $Q_{I_\mu} > 0$ for $I_\mu > 0$.

In setting their tuition schedules, colleges may offer different prices to students of different types. We assume that y is perfectly observable by the colleges, but that ℓ is unobservable. Instead, colleges can only observe the signal σ of ability described in Section 3.1. Colleges thus face a signal extraction problem: they set their tuition and admissions policy to influence the distribution of their student body, which is determined through Bayesian updating. We assume that colleges know the distribution of initial student characteristics $\mu : \mathbb{R}_+ \times \mathbb{R}_+ \to [0,1]$, the conditional signal distribution function, and the application decisions for students of each type (which are determined in equilibrium according to 3). Taking these distributions as given, colleges take into account how their choices for tuition and the admissions standard will affect the final distribution of the student body.

To calculate the distribution of characteristics in each college, it is helpful to first define the total probability a student with characteristics (y, ℓ) will enroll in College s. Let $p_i(y, \ell)$ denote the probability that such a student chooses application strategy $i \in \{AB, A1, A2\}$. The total probability is then:

$$\begin{aligned}\tilde{q}^s(y, \ell, \sigma, T^s, T^{-s}(y, \sigma)) = {} & q^s_{AB}(y, \ell, \sigma, T^s, T^{-s}(y, \sigma)) p_{AB}(y, \ell) \\ & + q^s_{As}(y, \ell, \sigma, T^s) p_{As}(y, \ell),\end{aligned} \tag{14}$$

where $q^s_i(y, \ell, \sigma, T^s, T^{-s})$ is given in Equation (8). Note that $\tilde{q}^s$ depends on the tuition schedule $T^{-s}(y, \sigma)$ of the other college if the student was accepted to both colleges, revealing the nature of competition between both colleges.

Having defined the total probability of enrollment for a student of type (y, ℓ), we can easily calculate the total enrollment, average ability of the student body, and the total tuition revenue for each college. Integrating over all student types and acceptable signals, the total enrollment and average student ability in College s are given by

$$\kappa = \int_{\underline{\sigma}_s}^{\infty} \int \tilde{q}^s(y, \ell, \sigma, T^s(y, \sigma), T^{-s}(y, \sigma))\, g(\sigma|\ell)\, d\mu(y, \ell)\, d\sigma \tag{15}$$

$$L_\mu = \frac{1}{\kappa} \int_{\underline{\sigma}_s}^{\infty} \int \ell\, \tilde{q}^s(y, \ell, \sigma, T^s(y, \sigma), T^{-s}(y, \sigma))\, g(\sigma|\ell)\, d\mu(y, \ell)\, d\sigma. \tag{16}$$

Similarly, total tuition revenue is given by

$$\mathcal{T}^s = \int_{\underline{\sigma}_s}^{\infty} \int T^s(y, \sigma)\, \tilde{q}^s(y, \ell, \sigma, T^s(y, \sigma), T^{-s}(y, \sigma))\, g(\sigma|\ell)\, d\mu(y, \ell)\, d\sigma. \tag{17}$$

Colleges balance their budget. Their revenue is derived from their own exogenous endowment income E^s, government subsidies or appropriations Tr^s (which may depend on the fraction of students enrolled, κ), and the total amount of tuition paid by students, $\mathcal{T}^s$. In addition to total instructional spending, $\kappa^s I_\mu$, the college faces operating expenses $C^s(\kappa)$, increasing in the college's enrollment level. For instance, these operating expenses could relate to administrative or maintenance costs that do not increase the value-added to students in the labor market, but are covered by the tuition students pay. The budget constraint of College s is thus given by

$$\kappa I_\mu + C^s(\kappa) \quad = \quad E^s(\kappa) + Tr^s(\kappa) + \mathcal{T}^s. \tag{18}$$

College's problem. College s then solves the following problem:

$$\max_{\underline{\sigma}_s, T^s(y,\sigma), \kappa, I_\mu, L_\mu} \quad \xi_s Q(I_\mu, L_\mu) \tag{19}$$

$$\text{s.t.} \qquad T^s(y,\sigma) \leq \bar{T}^s,\ \underline{\sigma}_s \geq 0 \quad \text{and} \quad (15),(16),(17),(18).$$

Note that the problem of College s depends on the tuition and admissions standard chosen by the other college due to the dependence of $\tilde{q}^s$ on the policies of *both* colleges.

Finally, note that we have introduced an upper bound on tuition for each college. We introduce these tuition caps for two reasons. The first is to improve the empirical properties of the model. A well known feature of the higher education market is that colleges post a sticker price tuition level, which a certain fraction of the student body pays[9]. This means that colleges are limited in how much they are able to extract from students with the highest willingness to pay.

The second reason is that this constraint on tuition also induces a non-trivial choice for $\underline{\sigma}_s$. Without limits on tuition, colleges can simply charge the low-signal students enough to compensate for the decrease in average student ability they cause. Thus colleges will not need to use an admissions threshold and will simply set $\underline{\sigma}_s = 0$. When tuition is bounded, however, students with low signals who are willing to pay $\bar{T}^s$ will not be able to compensate the college for lowering its average ability. The college would therefore find it optimal to screen these students out by raising its admissions threshold $\underline{\sigma}_s$.

Optimal tuition. An interior solution for the optimal tuition level applied to students with observable

[9]We rule out the possibility of admissions being influenced by donations for students at the very top of the income distribution.

characteristics (y, σ) and positive probability of being accepted $(\sigma \geq \underline{\sigma}_s)$ is given by

$$T^s(y,\sigma) = \underbrace{I_\mu + C^{s'}(\kappa) - E^{s'}(\kappa) - Tr^{s'}(\kappa)}_{\text{Marginal resource cost}} + \underbrace{\frac{\mathbb{E}\big[\tilde{q}^s\big(y,\ell,T(y,\sigma)\big)|y,\sigma\big]}{\mathbb{E}\big[\frac{-\partial\tilde{q}^s(y,\ell,T(y,\sigma))}{\partial T^s(y,\sigma)}|y,\sigma\big]}}_{\text{(Posterior) markup}}$$

$$- \underbrace{\frac{Q_{L_\mu}}{Q_{I_\mu}}\left(\mathbb{E}[\ell|y,\sigma] - L_\mu\right)}_{\text{(Posterior) ability discount}} \qquad \forall\, y, \sigma \geq \underline{\sigma}_s. \tag{20}$$

The equation is derived in Appendix D.1. It is found by combining the first order conditions for the college-level aggregates κ, I_μ, L_μ, with the first order condition for the tuition charged to a student with income y and signal $\sigma \geq \underline{\sigma}_s$.

The optimal tuition level can be broken down into three components. First, tuition must cover the marginal resource cost incurred by the college for enrolling an additional student. Note that this cost is common to all students and does not depend on y or σ. Second, since the college has market power, the tuition level also takes into account the student's willingness to enroll in the college conditional on being admitted. This introduces the familiar markup over marginal cost, which will be increasing in y since colleges will capture the higher willingness to pay among higher income students. Note that since the enrollment probability depends on the unobservable ℓ, colleges will also need to consider the distribution of possible ℓ values informed by the observable σ and y.

Third, since the college values a higher average ability among its students, there is a discount for students who have a higher (posterior) average ability than the overall average ability of the student body (and a penalty for students with a lower average ability). Since colleges cannot observe ℓ directly, they use the posterior mean of the distribution of ℓ that is obtained after observing the student's signal σ and parental transfer y. The posterior distribution of ℓ is given by:

$$f(\ell|y,\sigma) = \frac{\frac{\partial\tilde{q}^s}{\partial T^s(y,\sigma)} g(\sigma|\ell)\mu(y,\ell)}{\int \frac{\partial\tilde{q}^s}{\partial T^s(y,\sigma)} g(\sigma|\ell)\mu(y,d\ell)}. \tag{21}$$

Since higher ability levels yield higher average signals through the signal density, the (posterior) ability discount will be increasing in σ. Importantly, the distribution also depends on parental transfers through $\tilde{q}^s$, as defined in (14). This allows the equilibrium application choices of the students to affect the tuition they are offered. For example, suppose that at a given income level only high-ability students apply. Since colleges take the student application decisions as given, they recognize that applicants from such an income level are more likely to be higher ability and will therefore offer them higher levels of financial-aid. This mechanism, which is new to our model, helps explain why low-income students receive high levels of financial-aid, since only the highest ability among them apply.

Optimal admission standard. Lastly, the optimal admissions standard (given the policies of the other College) satisfies the following inequality for each College s:

$$\frac{\int T^s(y,\underline{\sigma}_s)\tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)}{\int \tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)} \geq I_\mu + C'(\kappa) - E'(\kappa) - Tr'(\kappa) - \frac{Q_{L_\mu}}{Q_{I_\mu}}\left(\frac{\int \ell\ \tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)}{\int \tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)} - L_\mu\right), \tag{22}$$

where $\tilde{q}^s(\underline{\sigma}_s) = \tilde{q}^s(y,\ell,\underline{\sigma}_s,T(y,\underline{\sigma}_s))$. The first order condition in (22) holds with equality when $\underline{\sigma}_s > 0$, and its derivation is provided in Appendix D.2. Equation (22) can be interpreted to mean that the average tuition revenue received from the lowest-signal students must be enough to compensate the College for the resource cost they impose, and the change they bring to the average ability.

Note, however, that the College's tuition policy in Equation (20) guarantees that it will be compensated for admitting the lowest-signal students since tuition depends on both σ *and* y. Thus, without any constraints on tuition, a College will not choose an interior value for $\underline{\sigma}_s$. Proposition 1 below formalizes this by demonstrating that when there are no caps on tuition, (22) will never hold with equality since the College will be able to charge the lowest-signal students enough to compensate for the change they bring to the average ability of the student body.

Proposition 1. *Suppose $\bar{T}^s$ is large enough so that the constraint $T^s(y,\sigma) \leq \bar{T}^s$ never binds. Then $\underline{\sigma}_s = 0$. That is, if there are no tuition caps, the college does not exclude students using an admissions threshold.*

Proof. See Appendix D.3 □

When there is a binding tuition cap, however, students with very low-signals will not be able to compensate the College for lowering its average ability. The College will then have to raise its admissions standard to exclude these students. This can be seen by examining (22). A low enough tuition cap will lower the left hand side by decreasing tuition revenue. Once the left hand side is sufficiently small, the College will need to increase $\underline{\sigma}_s$ to lower the right hand side and achieve equality. Increasing $\underline{\sigma}_s$ lowers the right hand side because it raises the average ability of the College's lowest-signal students, $\frac{\int \ell\ \tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)}{\int \tilde{q}(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)d\mu(y,\ell)}$.

3.3 Government

The government taxes labor income and uses the revenue to finance Pell Grants and college subsidies. The tax base is composed of all workers who did not attend college (including those who applied and those who did not) in both periods, and college educated workers in the second period. The intertemporal government budget constraint must hold according to

$$\sum_s \left[Tr^s(\kappa^s) + \int_{\underline{\sigma}_s}^{\infty} \int P(y) \tilde{q}^s g(\sigma|\ell) d\mu(y,\ell) d\sigma \right] =$$

$$\tau\, w \left[(1+R^{-1}) \int \ell^{\alpha_w}\, \tilde{q}^w(y,\ell)\, d\mu(y,\ell)\; + R^{-1} \sum_s \Gamma_s \int_{\underline{\sigma}_s}^{\infty} \int \ell^{\alpha_s} \tilde{q}^s g(\sigma|\ell)\, d\mu(y,\ell) \right], \tag{23}$$

where $\tilde{q}^w$ is the total probability of not attending college:

$$\tilde{q}^w(y,\ell) = 1 - \sum_s \int_{\underline{\sigma}_s}^{\infty} \tilde{q}^s(y,\ell,\sigma,T(y,\sigma)) g(\sigma|\ell) d\sigma. \tag{24}$$

3.4 Equilibrium

An equilibrium in the college market consists of value functions for applicants V^{Aj}, students with offers of admission V^{Oj}, enrolled students V^{Cj}, and workers V^W for $j = \{1,2,B\}$, policy functions $\{d, a^{s'}, c^s, a^{w'}, c^w\}$, probabilities $\tilde{q}^s, \tilde{q}^w$, admissions standards $\underline{\sigma}_s$ and tuition schedules $T^s : R_+ \times R_+ \to (-\infty, \bar{T}^s]$, college choices $\{\kappa^s, I^s_\mu, L^s_\mu, \Gamma_s\}$ for $s = \{1,2\}$, and tax rate τ, such that

1. Given $\{\tau, \Gamma_s\}$, $\{V^{Cs}, a^{s'}, c^s\}$ solves (9) for $s \in \{1,2\}$, and $\{V^W, a^{w'}, c^w\}$ solves (12).
2. Given $\{V^{Cs}, V^W\}$ and T^s for $s \in \{1,2\}$, V^{Oj} satisfy (6)-(7) for $j \in \{1,2,B\}$.
3. Given V^{Oj} and $\{T^s, \underline{\sigma}_s\}$ for $s \in \{1,2\}$, V^{Aj} solve the problem of the applicant in (1)-(2) for $j \in \{1,2,B\}$.
4. $d : R_+ \times R_+ \to \{W, AB, A1, A2\}$ is the application choice which solves (3).
5. $\tilde{q}^s, \tilde{q}^w$ are, respectively, the total probabilities of ending up at College s and ending up as a worker:

$$\begin{aligned} \tilde{q}^s(y,\ell,\sigma,T(y,\sigma)) &= q^s_{AB}(y,\ell,\sigma,T^s(y,\sigma)) \mathbf{1}_{\{d(y,\ell)=AB\}} \\ &\quad + q^s_{As}(y,\ell,\sigma,T(y,\sigma)) \mathbf{1}_{\{d(y,\ell)=As\}} \\ \tilde{q}^w(y,\ell) &= 1 - \sum_s \int_{\underline{\sigma}_s}^{\infty} \tilde{q}^s(y,\ell,\sigma,T(y,\sigma)) g(\sigma|\ell) d\sigma. \end{aligned}$$

   where $q^s_i, i \in \{AB, As\}$ is defined in (8).

6. Given $\tilde{q}^s$ for $s \in \{1,2\}$, $\{\kappa^s, I^s_\mu, L^s_\mu, \Gamma_s, T^s, \underline{\sigma}_s\}_{s\in\{1,2\}}$ is a solution to the the College game presented in Section 3.2. That is, for each College $s \in \{1,2\}$, given the tuition schedule, admissions threshold and value added of the other college, $\{\kappa^s, I^s_\mu, L^s_\mu, T^s, \underline{\sigma}_s\}$ is a solution to (19), and

$$\Gamma_s = \xi_s Q(I^s_\mu, L^s_\mu).$$

7. The government balances its budget according to equation (23).

3.4.1 Equilibrium Selection

The presence of peer effects introduces the potential for multiple equilibria. We follow Epple et al. (2006) and Epple et al. (2017) in focusing on an equilibrium where the ranking of college quality corresponds to the ranking of the endowment size. In our baseline equilibrium and subsequent analyses, College 1 will then have the higher value-added and attract the higher-ability students. This is consistent with its higher endowment which allows for higher amounts of spending per student. It is beyond the scope of our paper to show that an equilibrium under this refinement is unique. However, we note the our numerical procedure used to calculate the equilibrium converges consistently for different initial guesses, and is robust to small changes in the parameter space.

Appendix G provides details on how the equilibrium of the college market is solved numerically.

4 Estimation

We are able to choose a subset of the model parameters directly from the data and the literature. The remaining parameters are estimated using the simulated method of moments. Table 2 summarizes the parameters chosen outside the model, while Table 4 summarizes the resulting estimated parameters.

4.1 Data

We rely on three main datasets to estimate our model: the High School Longitudinal Study of 2009 (HSLS), the 2012 cohort of Beginning Postsecondary Students Longitudinal Study (BPS), and the Integrated Postsecondary Education Data System (IPEDS).

The HSLS consists of a nationally representative sample of high-school students who are tracked over time as they transition from high-school to college. A detailed description of this dataset is provided in Section 2.1. We use the HSLS to calculate key moments regarding student application and enrollment rates which we then match in our estimation procedure. We also use the HSLS to estimate the parameters governing the distribution of students characteristics.

The BPS is a longitudinal dataset provided by the National Center for Education Statistics (NCES) which follows a representative cohort of students over time starting with their first year of undergraduate studies. The data include details of all financial-aid received by the students who began college in the 2011-2012 academic year (close to our HSLS cohort who began college in the fall of 2013). This rich information about enrollment, tuition, and financial-aid supplements the HSLS which has only self reported financial-aid data that is not broken down by source (private institutional aid vs. government grants). We use the BPS to set the level of non-institutional grants available to students, and to determine how grants and financial-aid vary with SAT score and parental income. We also use it to determine the college tuition caps.

IPEDS is a public college-level database managed by the NCES. It is established from a series of mandatory surveys of all U.S. colleges who participate in federal student financial-aid programs. IPEDS is helpful for estimating the college side of our model because it contains rich information on college-level financial variables. The data contain details about college revenue from all sources including tuition, government appropriation, and endowment income. The data also report and categorize all college costs and expenses. We provide details about how we construct our sample of colleges from IPEDS in Appendix B.3.

4.2 Students' attributes

We start by describing our choices for the parameters that govern the student's problem. A list of these parameters and their values are given in Table 2.

Table 2: Exogenous Parameters

Variable	**Description**	**Value**	**Source**
Returns to education			
α_s	Returns to ability (college)	0.78	Abbott et al. (2019)
α_w	Returns to ability (no college)	0.55	Abbott et al. (2019)
Aggregate prices and borrowing constraint			
$\underline{a}_s$	Student borrowing limit	-0.775	USED
R	Gross interest rate	1.0386	USED
w	Wage	2.7	CPS
Budget Parameters			
ψ	Financial application cost	0.00375	IPEDS
g_0	Intercept of grant function	0.43	BPS
g_1	Slope of grant function	0.19	BPS
P_{max}	Pell Grant maximum	0.5645	USED
College Parameters			
o_0^s	Fixed costs	[0.15, 0.25]	IPEDS
o_2^s	Quadratic costs	[5.77, 0.43]	IPEDS
E^s	Endowment income	[0.12, 0.06]	IPEDS
Tr_1^s	Gov transfers per student	[1.00, 0.62]	IPEDS
$\bar{T}^s$	Tuition Cap	[2.50, 1.20]	BPS

Preferences. Individuals have logarithmic preferences over consumption each period:

$$u(c) = \log(c)$$

In our two period model, we consider the first period to account for 4 years (time in college), and the second period to account for 60 years (time spent working). In order to pick the appropriate values for $\{\beta, R, w\}$, we show in Appendix E how a life-cycle model with $T+1$ periods maps into our two period model, where we take $T = 15$. In the life-cycle model, we set $\beta = (1/R)^4$. Mapped into our two-period model, this gives us the values $\tilde{\beta} = 5.48$ that correspond to the problem presented in (9).

In order to simplify the computation of the equilibrium, we add a set of Type I extreme value shocks to the discrete choice problem in (3). With the shocks, the problem becomes

$$\max \left\{ V^{A1}(y,\ell) + \epsilon_A^1, V^{A2}(y,\ell) + \epsilon_A^2, V^{AB}(y,\ell) + \epsilon_A^3, V^W(y,\ell,0) + \epsilon_A^4 \right\}. \tag{25}$$

where we let $\epsilon_A \sim Gumbel(\frac{1}{\lambda_a})$. The shocks simplify the equilibrium computation since they smooth the application probabilities taken as given by the Colleges when setting tuition. The shocks allow these probabilities to respond continuously to changes in the tuition schedule in each iteration, which simplifies the convergence of the solution algorithm described in Appendix G. Since the shocks are used only to simplify the computation, we pick λ_a to be large in order to minimize their variance. We set $\lambda_a = 40$. Note that these shocks are included in addition to the shocks that occur after the application stage, when the students make their enrollment decisions. We include the scale parameter for the second-stage shocks in our method of moments estimation, described in more detail below.

Finally, we assume that the college specific non-pecuniary costs of completing college are given by the linear function

$$\nu^s(\ell) = \nu_0^s - \nu_1^s \ell, \tag{26}$$

where the parameters $\{\nu_0^1, \nu_1^1, \nu_0^2, \nu_1^2\}$ are included in the joint method of moments estimation below.

Aggregate Prices. We choose $w = 2.7$, so that the model produces the average wage calculated from the Current Population Survey (CPS). We set $R = (1.0386)^4$ since annual borrowing rates for undergraduate students was 3.86% in 2013-2014 academic year. This rate was set by the US Department of Education (USED) through the Federal Student Loan Program. Again, note that we need to adjust these life-cycle level values to fit into our two period model using Appendix E.

Distribution of characteristics. We take the distributions of student characteristics, i.e. (y, ℓ), from the HSLS. Since the model deals directly with the level of parental transfers the students receive, we abstract from a theory of parental transfers and therefore do not use parental income from the data directly. Instead we use the student's Expected Family Contribution (EFC), which is determined according to rules set by the Department of Education using the FAFSA filled out by the students and their parents. EFC is a measure of the amount of resources a student reasonably has available to them in order to attend college (before financial-aid), and so maps well into our notion parental transfers used in the model.

In the HSLS, we are able to observe the EFC for all students who completed the FAFSA and attended college, and we calculate it for those who did not using the household income reported by their parents in the survey. Our direct calculation uses the EFC formula established by the Department of Education, and is described in detail in Appendix C.1. We provide the details of how we use the HSLS to fit the distribution of student characteristics (y, ℓ) in Appendix C.2. Finally, note that since students in the HSLS began college in the 2013-2014 academic year, we use 2013 dollars as our numeraire and re-scale by \$40,000. Note that the first period in the model corresponds to four years. Thus, for example, $y = 1$ in the model corresponds to an EFC of \$10,000 per year over four years.

Signal distribution. We assume that signals follow a normal distribution conditional on ability, with mean ℓ, and variance σ_g^2. We also assume that the distribution is truncated below at 0, so that the lower bound of the support of the signals is finite. We include the variance σ_g^2 as part of the joint parameter estimation below.

Financial application costs. The application process involves two types of costs: a non-pecuniary cost and a financial cost. We find in the HSLS that students send about 3 applications on average, and from IPEDS that the average application cost is \$50. This total cost corresponds to a value of $\psi = 0.00375$ in the model. Note that for simplicity, the financial cost of applying is the same regardless of whether students send one or two applications. The marginal cost of sending an additional application in the model is instead captured by the non-pecuniary costs provided that $\phi_B > \max\{\phi_1, \phi_2\}$. Since the financial cost includes only the direct resource cost of applying to college, all other indirect costs are captured by the psychic costs.

Returns to ability and education. We take the parameters governing the labor market returns to ability α_s, α_w directly from Abbott et al. (2019). They find that for college graduates, the ability gradient is 0.797 for males and 0.766 for females. For high school graduates they find the gradients to be 0.517 and 0.601 for males and females respectively. Hence we simply set $\alpha_s = 0.78$ for graduates of both colleges and set $\alpha_w = 0.55$. The labor market returns for attending the more selective college are then captured only by differences in Γ_s.

Borrowing constraints. According to the Federal Student Loan Program, the aggregate limit for dependent students who are attending an undergraduate degree is \$31,000 in federal loans. We therefore set the student borrowing limit to $\underline{a}_s = -0.775$, which corresponds to \$31,000 over four years.

Grants and Aid. We consider two types of grants available to students who attend college and allow these to vary by the student's level of parental transfer. The first are grants that are exogenous in the model, denoted $Gr(y)$, which stand in for unmodeled state grants or private grants and scholarships. We assume they take the following form:

$$Gr(y) = \max\{g_0 - g_1 y, \underline{g}\}. \tag{27}$$

Using the BPS, we pick $\underline{g} = 0.15$ and estimate $g_0 = 0.43, g_1 = 0.19$. The details are provided in Appendix C.3.

Pell Grant amounts are set according to the Department of Education. In order to qualify, students must demonstrate sufficient financial need as measured by the difference between their EFC and the Net Cost of Attendance (tuition plus room and board minus institutional financial aid). The Pell Grant makes up this difference, up to a maximum level. In the model, we have

$$P(y) = \max\{P_{max} - y, 0\}, \tag{28}$$

so that Pell grants effectively give all low-income students a minimum level of transfers. In 2013 the maximum Pell Grant amount was \$5,645 per year, so we set $P_{max} = 0.5645$. We later study the effects of changes to this parameter in the policy section of the paper.

4.3 Colleges' attributes

College Types. To separate both types of colleges in the model, we rely on Barron's 2015 rankings of U.S. colleges. This ranking is commonly used in the literature as a measure of college quality. We consider College 1 to be Barron's Tier 1 and 2 schools ("elite" and "highly selective"), and College 2 to be all other four-year colleges and universities (excluding for-profit colleges). Table 3 summarizes key empirical differences across these college types using our sample of colleges from IPEDS.

College Type	**1**	**2**
Number of Colleges	182	1,621
Undergraduate Enrollment	17%	83%
Fraction Public	53%	75%
Average SAT Score	1297	1056
Rejection Rate	57%	32%
Instructional Expenditures per Student (US\$)	22,468	8,846
Average Tuition and Fees (US\$)	26,327	12,077
Net Cost for Bottom 20% Income (US\$)	8,750	9,646
Median earnings 10 Years After Entry (US\$)	58,706	41,906
Endowment Assets per Student (US\$)	141,838	12,332

Table 3: Empirical Differences Across College Types. Statistics are enrollment-weighted averages. Source: IPEDS, College Scorecard.

Table 3 shows that the differences between each college type are consistent with the model's predictions. The top colleges enroll a smaller share of the total student population, have higher expenditure per student, higher SAT scores, higher average tuition, while still offering low tuition to students at the bottom of the income distribution. The Barron's ranking thus leads to a natural partition between colleges that we use for our quantitative analysis. Note that while the top colleges are more likely to be private, as a group their overall enrollment mostly consists of students in public colleges. Therefore, unmodeled institutional differences between public and private schools are unlikely to play a big role in explaining differences between the two types since both groups have a similar composition of students within each type.

Technology. Colleges' quality is defined by

$$Q(I_\mu, L_\mu) \quad = \quad I_\mu^{1-\rho_L} L_\mu^{\rho_L},$$

where ρ_L is included as part of our internal estimation below.

Cost function and government transfers. We follow Epple et al. (2017) and assume college operating costs are a quadratic polynomial in their total enrollment,

$$C^s(\kappa) = o_0^s + o_1^s \kappa + o_2^s \kappa^2. \tag{29}$$

We set the linear cost $o_1^s = 0$, and estimate the fixed cost and quadratic cost terms using our sample of colleges from the IPEDS data. We find $(o_2^1, o_2^2) = (5.77, 0.44)$, and $(o_0^1, o_0^2) = (0.15, 0.25)$. The details of our estimation are provided in Appendix C.4.

Government spending on public universities (which account for the majority of enrollment in both types of colleges) accounts for a large share of college funding. We assume that the government subsidizes colleges on a per-student basis:

$$Tr^s(\kappa) = Tr_1^s \kappa.$$

Similar to the cost function estimation, we estimate this relationship using our IPEDS sample. We find $(Tr_1^1, Tr_1^2) = (1.0, 0.62)$. See Appendix C.4 for details.

Endowment income. We assume that private endowment income at the colleges has a component that depends on enrollment (private donations that are restricted to be used on scholarships per student enrolled), and a fixed component:

$$E^s(\kappa) = E_0^s + E_1^s \kappa. \tag{30}$$

We make this distinction for endowment income because the enrollment levels for each college are determined largely by the net fixed costs: $o_0^s - E_0^s$[10]. Enrollment will then be determined by the fraction of the endowment used to offset the fixed costs, and the fraction used to offset per-student costs. This is important for College 1 where the endowment is large, so we estimate E_1^1 to match the observed enrollment in College 1 (discussed in the estimation procedure below). For College 2, we find that we are unable to match the observed enrollment given our estimate of o_0^2 even if we assume that $E_0^2 = 0$. We thus leave the enrollment at College 2 as an un-matched target and simply assume that $E_0^2 = 0$.

Finally, to find the remaining endowment parameters, we use our IPEDS sample to calculate the total level of private endowment income received by each college by adding up all private revenue the college received over our 2013-2016 sample period. This includes unrestricted revenue the college may use from gifts, investment return from their endowment, or contributions from affiliates. We find the *total* endowment incomes to be $E^1 = 0.12, E^2 = 0.06$. Thus for College 1, we set $E_0^1 = 0.12 - E_1^1 \kappa^1$ (using our SMM estimate for E_1^1), and for College 2 we have $E_1^2 = 0.06/\kappa^2$ (where κ^1, κ^2 are the enrollment levels observed in the HSLS).

Tuition caps. To find the tuition caps $\bar{T}^s$ we rely on the BPS data, which include details about all financial-aid received by each student. We note that in the data, the average level of tuition minus college-specific financial-aid becomes flat in Expected Family Contribution (EFC) for high-EFC

[10] Since colleges only care about *per* student variables, they have an incentive to keep their enrollment levels as small as the net fixed costs permit.

students. This fact, illustrated in Figure C.6 reflects that high-income students generally do not receive financial-aid, as institutional grants are more likely to be given to low-income students. In the model, the caps on tuition bind for students with the highest parental transfers. We therefore pick $\bar{T}^1 = 2.5$, $\bar{T}^2 = 1.2$, which correspond to the average upper bound on tuition paid by high-income students within each college in the data (see Figure C.6).

Value added. Colleges' efficiency parameter ξ_s is chosen to match the college wage premium in the data. Abbott et al. (2019) estimate an average college wage premium of 0.6. Since 17% of students are enrolled in selective colleges, we have that the average wage premium is such that

$$0.17 \log \Gamma_1 + 0.83 \log \Gamma_2 = 0.6.$$

To find the dispersion between Γ_1, Γ_2, we rely on estimates by Chetty et al. (2020) who report that 80% of the difference in median log earnings 10 years after college can be explained by differences in colleges' selectivity. Using the numbers from Table 3, we have

$$\log \Gamma_1 - \log \Gamma_2 = 0.113.$$

This gives us $\Gamma_1 = 2.01, \Gamma_2 = 1.79$. We fix these corresponding values for Γ_s in the baseline estimation below, and find ξ_s so such that the resulting value for I_μ^s, L_μ^s are consistent with the corresponding Γ_s values.

4.4 Method of moments estimation

We estimate the remaining parameters $\Theta = \left\{\phi_1, \phi_2, \phi_B, \lambda_c, \nu_0^1, \nu_0^2, \nu_1^1, \nu_1^2, E_1^1, \rho_L, \sigma_g^2\right\}$ jointly by minimizing an unweighted quadratic (percent) distance criterion function between data moments and simulated model moments. The results of the estimation are presented in Table 4, where we provide descriptions of the data moments used.

Application disutility costs. To help identify the application disutility costs, ϕ_1, ϕ_2, ϕ_B, we match the fraction of students applying to each college type only, and the fraction of those applying to both. In the HSLS, we observe that 2% apply only to College 1, 42% apply only to College 2, and 14% apply to both. Note that we find the estimated College 1 application cost to be lower than the College 2 application cost. This reflects that in the model it is very risky only to apply to College 1, so a relatively low application cost is necessary match the correct fraction of students who would choose only to apply there.

To assess the magnitude of these costs, we calculate the equivalent consumption a student would forgo in order to remove the disutility cost. In Table 5 we report the average consumption equivalent values as a percentage of average life-cycle consumption for all students. We see that our estimates are considerably higher than those found in the literature, but note that this is largely due to high marginal utility of consumption among high-income students and those who choose not to enroll. If we restrict only to lower-income college enrollees who have lower consumption due to credit constraints, we see that their application disutility costs are substantially lower.

Variable	Description	Value	S.E.
Preference			
ϕ_s	Application cost (College s)	0.100	0.004
		0.185	0.025
ϕ_B	Application cost (both)	0.320	0.015
ν_0^s	College psych cost (intercept)	2.410	0.175
		1.700	0.401
ν_1^s	College psych cost (slope)	-1.467	0.319
		-0.758	0.383
λ_c	Extreme value scale parameter	1.855	0.007
Technology			
E_1^1	Coll 1 endowment	1.371	0.008
ρ_L	Quality param	0.854	0.002
σ_g^2	Variance of signal	0.110	0.008
	Moment	**Model**	**Data**
	% Applying Coll 1	2.00	2.00
	% Applying Coll 2	46.0	42.0
	% Applying to both	15.0	14.0
	College 1 rel app rate (SAT quintile 5/4)	3.43	3.33
	College 2 rel app rate (SAT quintile 4/2)	1.63	1.55
	Coll 2 attnd rate (1 app)	0.79	0.83
	Coll 1 attnd rate (accepted both)	0.55	0.58
	Coll 2 attnd rate (2 app)	0.94	0.90
	% Enrolled Coll 1	7.00	7.00
	Δ Tuition wrt SAT	-0.08	-0.08
	Coll 1 accpt rate	0.69	0.70

Table 4: Jointly estimated parameters. All moments were calculated using the HSLS, except the change in tuition with respect to SAT, which is calculated using the BPS. "College 1 rel app rate (SAT quintile 5/4)" refers to the relative application rates to selective colleges between students in the 5th vs 4th SAT quintile. For example, students in the 5th quintile are about 3.33 times more likely to apply to a College 1 than students in the 4th SAT quintile. "College 2 rel app rate (SAT quintile 4/2)" is defined similary. "Coll 2 attnd rate (2 app)" refers to the College 2 attendance rates for students who applied to both types and were only admitted to College 2. "Δ Tuition wrt SAT" is determined by regressing net-tuition on SAT in our BPS data (controlling for EFC and college-fixed effects).

	Overall average	Enrollee average (Bottom 50% income)	Fu (2014)
Apply College 1	0.78%	0.52%	0.30%
Apply College 2	1.19%	0.58%	0.34%
Apply Both	1.87%	0.89%	0.45%

Table 5: Average application disutility costs among the student population expressed as a percentage of average life-cycle consumption (calculated in the model to be about $900,000). The application costs calculated in Fu (2014) are provided for reference to the literature. The costs from Fu (2014) are calculated assuming two applications for College 1, three for College 2, and six for both (which we observed among average applicants in the HSLS).

College preference parameters. We introduced a 'psychic' disutility cost of college in (9), which we allow to vary by college and assume depends linearly on ability:

$$\nu^s(\ell) = \nu_0^s - \nu_1^s \ell.$$

Matching student enrollment decisions conditional on being accepted helps us identify the intercept parameters, while variation in student applications across the SAT distribution help us identify the slopes. In Table 4 we find that while the intercept ν_0^s term is higher in College 1, the slope term is more negative. This reveals substantial differences in preferences for each college across the ability distribution. College 1 will be less costly than College 2 for high-ability students while the reverse is true for low-ability students.

To get a sense of the magnitude of these psychic costs of schooling, we again calculate consumption equivalent amounts for each student and report them as a percentage of average life-cycle consumption in Table 6. We note that the application costs are smaller than those estimated in the literature (even when adding application costs). While the gap is smaller among enrollees, we see that our psychic costs are considerably smaller than the estimates from Abbott et al. (2019) for the overall population. The reason is that our model accounts for the fact that students need to be admitted to college in order to attend, and the probability of not getting in is sufficiently low for many students. Thus our model can help rationalize the relatively high psychic costs of schooling found in the literature: many students choose not to enroll not because of preferences, but because they are unlikely to be admitted and therefore do not apply in the first place. A relatively small psychic cost is then sufficient to exclude many students from the college market.

Finally, λ_c is the scale parameter for the extreme value shock introduced in (4)-(5) that is realized after a student receives offers from either college. Unlike the shocks that were added to the application choice, these shocks are meaningful because they encourage students to apply more aggressively by increasing the option value of having another college to choose from (compare 6 with 7). Therefore, if λ_c is low then a large fraction of students applying only to College 2 will be unlikely to enroll since their applications were motivated more by the increased option value rather than the value-added from attending College 2. Thus, variation in College 2 attendance conditional on applying only there and being accepted created by variation in λ_c helps us identify the parameter.

	Avg psychic cost, $\nu^s(\ell)$	Abbott et al. (2019)
College 1	1.23%	10.7%
College 2	2.76%	
Among enrollees	1.86%	4.6%

Table 6: Average psychic costs expressed as a percentage of average life-cycle consumption (calculated in the model to be about \$900,000). The costs calculated in Abbott et al. (2019) are provided for reference to the literature.

College quality and signal strength. The ability parameter ρ_L in the colleges' quality function and the variance of the signal σ_g^2 are chosen to help the model match the average responsiveness of tuition with respect to ability, and the average acceptance rate to College 1 conditional on applying. As discussed in Section 3, the (posterior) ability discount depends on the strength of the signal, so that the average responsiveness of tuition to the signal is sensitive to the choice of the variance. In the BPS, we estimate this effect by regressing tuition minus institutional grants on SAT score, controlling for EFC and college fixed effects. Similarly, matching the average conditional acceptance rate for College 1 helps us identify ρ_L because we find that a higher value for this parameter increases the size of the ability discount, which increases the attractiveness of applying to College 1. All else equal, this will tend to reduce acceptance rate due to the increase in applications with little change to the enrollment.

Endowment and Fixed Costs. Lastly, the fixed cost each college faces is an important variable for determining its enrollment level. As mentioned above, colleges care about *per student* variables, so they will remain as small as their fixed costs allows them to be. Thus, variation in the fixed component of a college's endowment, E_0 in (30), will have an effect on its overall enrollment level. We find that this is particularly important for College 1, who has a high total endowment income relative to its estimated fixed cost. We then estimate E_1^1 to match the College 1 enrollment level observed in the data, and back out the corresponding value for E_0^1 from the observed total endowment income calculated from IPEDS. For College 2, we find that the empirical enrollment level is too large to be explained only by its endowment. We thus leave its enrollment level to be an un-targeted moment.

5 Results

5.1 Model Fit

College Level Statistics. Table 7 below compares college level statistics produced by the model to ones we calculate in the data. We find that the model somewhat underestimates the average EFC within each college, though it captures the fact that EFC in College 1 is about twice as high as in College 2. In order to compare average student ability in the model and in the data, we report the average SAT and average application signal within each college (since they are both noisy signals of true ability). To put the two measures in similar units, we report the standardized values (where we standardize the signal according to the distribution of applicants). As we see in the second row of the table, the model does a good job of allocating relatively high signal students into College 1, and students with average signals into College 2.

	College 1		**College 2**	
	Model	*Data*	*Model*	*Data*
% Enrolled	7.15	7.00	32.09	36.0
Avg Student EFC	2.22	3.04	1.14	1.55
Avg Applicant Signal/SAT	1.45	1.12	0.29	0.05
Instr Spending per Student	2.1	2.25	0.98	0.89
Avg Net Tuition	1.6	1.41	0.61	0.50
Income Distribution				
Q1 Income	9.8%	9.6%	22.1%	23.4%
Q2 Income	11.6%	12.1%	20.5%	16.4%
Q3 Income	26.4%	23.1%	30.6%	28.1%
Q4 Income	52.2%	55.2%	26.8%	32.2%

Table 7: College Level Statistics (all untargeted except College 1 enrollment). Income refers to parental transfers in the model, and Expected Family Contribution (EFC) in the data. The average signal (in the model) and SAT score (in the data) among applicants are standardized for comparison.

Finally, the model is able to capture the large difference in instructional spending per student across each college. The amount spent per student in College 1 is large due to its higher tuition levels and large endowment. The model is also able to capture the fact that spending per student is significantly higher than average tuition revenue per student, which reflects how government subsidies and grants are able to cover remaining college expenses. Note, however, that average net tuition at each college is slightly higher in the model compared to the data.

Student Distribution Within Colleges. Importantly, the model is able to deliver student income distributions within each college that are consistent with the data. Table 7 shows the distribution of parental transfers (EFC in the data) within in each college. We see that in both colleges the EFC distribution is well accounted for, especially in College 1 where the model very closely captures the correct share of low-income students. Normally in this type of college-market model, colleges have a strong incentive to enroll mostly students from the top of the income distribution since they pay higher tuition, and are generally higher ability since income and ability are correlated. Epple et al. (2017), for example, over-predict the share of high-income students, and argue that unmodeled social objectives like affirmative action may help explain the gap in low-income student enrollment.

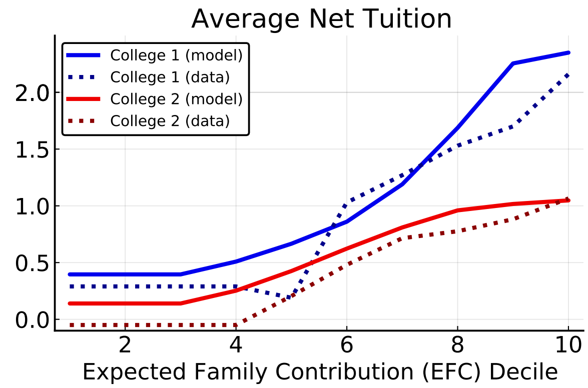


Figure 3: Average net-tuition (sticker-price tuition minus all grants) for students by Expected Family Contribution decile. Source: BPS.

In our model, the application and admissions system helps us capture the correct share of low- and high- income students at the more selective college without making any additional assumptions about social objectives. This happens for two reasons. The first is that since colleges cannot observe ability perfectly, they will limit the size of the ability-based tuition discounts they would otherwise offer. Thus, in order to match the sensitivity of tuition to SAT observed in the data, we require a relatively large value for ρ_L, the parameter which governs college's willingness to substitute average instructional spending for a higher average ability. This provides the colleges with a smaller motive to raise revenue and instead enroll higher ability students, many of which are lower-income and would otherwise be excluded. The second reason is due to selection effects arising from the application and admissions system. Since low-income students are less likely to apply to the more selective college, the ones who do apply are very high-ability in equilibrium. Thus the selective college can be confident that enrolling low-income students will help increase its average ability, and it will give large tuition discounts to the low-income students it enrolls. This selection effect due to the application choices of low-income students is explored further in the next section.

Tuition Variation. In Figure 3, we plot the average net-tuition in the model and in the data for students grouped by EFC decile within each college. In the data, net-tuition is defined as sticker-price tuition minus all grants available to the student. In the model, we then set net-tuition for a student with parental transfers y and signal σ to be $T(y, \sigma) - P(y) - Gr(y)$. Due to the presence of the tuition caps we have imposed, the model is generally able to capture the tuition for students at the very top of the EFC distribution, especially in College 2. The model is generally successful at predicting both the level and change in net-tuition across the parental income distribution, but somewhat overestimates the tuition for low-income students. One reason net-tuition for low-income students is high relative to the data is that nearly all students in the model borrow the full amount of student loans available to them, allowing the colleges to charge relatively high tuition. Student borrow the full amount because they have high second period earnings and there is no income risk, so they try to equalize consumption across both period.

Student Application and Enrollment Patterns. The model is able to account for the low application and enrollment rates observed for low-income students despite the presence of substantial

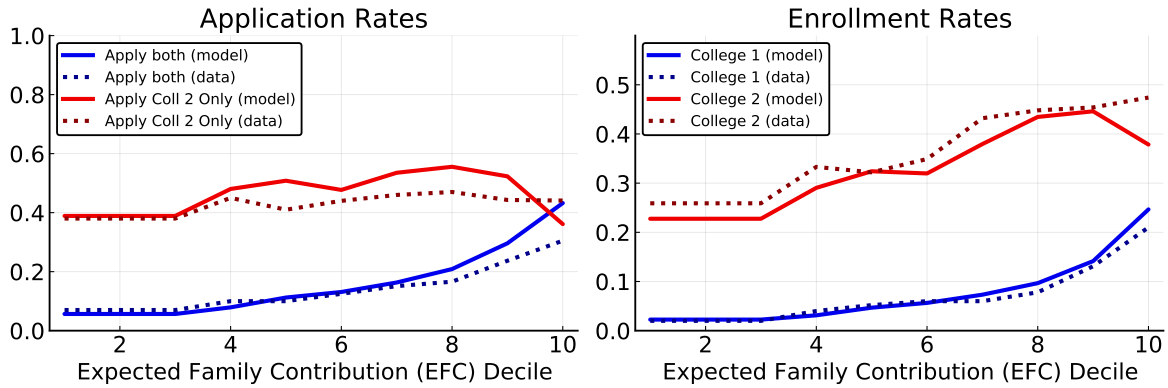


Figure 4: The left panel displays the average application rates for students by Expected Family Contribution (EFC) decile, and the right panel displays the average enrollment rates to each college. In the left panel, we include only applications to College 2 or applications to both since the fraction of students only applying to College 1 is small both in the model and the data. Source: HSLS.

financial-aid described above. Figure 4 shows the model predicted average application and enrollment rates alongside the ones observed in the HSLS data for students grouped by Expected Family Contribution (EFC) decile. We see that the model correctly predicts that applications to the more selective college are increasing in EFC (we do not report the fraction only applying to College 1 since it is small). Note that the model overestimates the application rates for students at the very top of the EFC distribution.

The right panel in Figure 4 shows the model predicted enrollment rates and the ones observed in our data for students grouped by EFC decile. We find that the model does a good job of capturing the increasing patterns in enrollment for students at both colleges, though the model underestimates the enrollment at the top for College 2. Note that for each plot the overall means of the application rates and College 1 enrollment rates are targeted in the baseline estimation, but the variation in applications by EFC and the enrollment rates in College 2 are not targeted.

5.2 Effect of Applications on Low-Income Student Enrollment

This section illustrates the importance of the application and admissions process in accounting for low-income student enrollment. As discussed before, an important mechanism in the model that accounts for merit-based financial-aid at the colleges is the informativeness of the students' signals. If, for example, only the highest ability among low-income students apply, then their signals will be highly informative about their ability and they will receive high levels of financial-aid since the colleges will be confident that they are high ability.

To illustrate this mechanism, Figure 5 shows what happens to high-ability student enrollment when all low-income students apply in the same way as their higher-income peers. Specifically, we adjust the application rates from the baseline equilibrium by requiring that low-income students apply to both colleges at the same rate as the highest-income students. We then recompute the equilibrium, holding fixed these new application pools (thus this is a partial-equilibrium analysis since student application behavior remains fixed). Note, however, that we still allow tuition, admissions policies, and student enrollment behavior to adjust. This exercise helps isolate the effect of student application portfolios on the overall allocation of students.

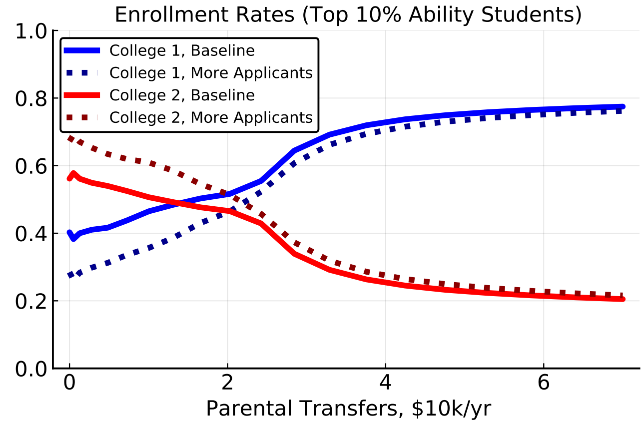


Figure 5: Average enrollment rates for students in the top 10% of the ability distribution by parental transfers. The plot shows the effect on enrollment of switching to an equilibrium where low-income students apply to both Colleges at the same rates as high-income students, regardless of their ability.

As we see in Figure 5, worsening the application pool of low-income students reduces the enrollment of high-ability, low-income students in College 1 by about a quarter. We see that with the substantial increase in low-ability, low-income applicants, the signals of all low-income students are now much more likely to have come from low-ability students. This will cause the college to lower the financial-aid it offers the low-income students, causing them to enroll at much lower rates. Note that since College 1 is now unable to enroll as many high-ability students, its value-added will also decline due to the lower average ability in its student body. Finally, we see that this effect is present not only for low-income students, but also for middle-income students who also enroll in College 2 instead. A summary of the overall changes to the college market in the new equilibrium is provided in Table F.5.

Overall, this section highlights an important insight of the model: the composition of college application pools affects the informativeness of the signals and hence the colleges' tuition levels. This helps resolve the "puzzle" of low application rates from low-income students despite the presence of high financial-aid. According to the model, low-income students receive high financial aid precisely *because* they are less likely to apply. Thus, policies aimed at increasing applications that are not targeted by ability will be harmful to low-income, high-ability students who would otherwise benefit from having their signal be more informative.

6 The Role of Application Informativeness

In this section we study the general equilibrium effects of changes to the signal variance. In our first counter-factual exercise we examine the effect of increasing the signal variance, motivated by concerns over the use of the SAT (assuming its use increases application informativeness). In our second exercise, we study the effect of switching to perfectly informative signals by removing admissions uncertainty for students, and allowing colleges to fully observe student ability.

6.1 Ending the SAT

What would happen if the application signals were to become less informative? This question is motivated by the fact that many colleges waived SAT or ACT requirements during the Covid-19 pandemic. Moreover, the use of standardized tests for college admissions has recently come under scrutiny as the University of California system has begun phasing out their reliance on the SAT/ACT for admissions. If we consider the use of the SAT or ACT as part of the technology that increases the informativeness of the signals, we can model the removal of these tests by making the signals less informative. To study the effects of less informative signals, we increase the variance of the signal by a factor of six from the baseline estimation and recompute the equilibrium. We choose this value for the signal increase because we find that it minimizes average welfare in comparison to the baseline equilibrium[11]. Thus, through the lens of our model, this experiment can be interpreted as an upper bound for the losses caused from removing the SAT.

Table F.6 shows the effect of increasing the signal variance on college-level variables. In the new equilibrium the higher variance leads to poorer sorting based on ability, which causes the average student ability to drop in both colleges. Less informative signals also reduce the marginal cost of admitting lower signal students, since their signals are now more likely to have come from higher ability applicants. Thus we see increases in the fraction of students enrolled, and decreases in the admissions standards. While there is little change to the income distribution within each college, the ability distribution has become more diffuse, reflecting that lower ability students can now more easily sort into the colleges by getting sufficiently high signals by chance. Finally, the colleges endogenously increase their spending per-student, but it is not enough to offset the decreases in average ability, leading to lower value-added at both colleges.

To see which students are affected by the decrease in the informativeness of the signals, Figure 6 plots the percent of consumption in each period different types of students would be willing to forgo in order to be born in the new equilibrium. Note that for reference, we have also plotted the changes to student enrollment resulting from the new equilibrium in Figure F.8. As expected, the high ability students are the ones who are most hurt from switching to the new equilibrium since the higher variance decreases their likelihood of being admitted to the colleges at the expense of the lower ability students. Importantly this effect is strongest for the low-income, highest-ability students because College 1 now offers lower levels of financial-aid to the highest signal students.

The students who benefit from this change are the high-income, low-ability students, who can now more easily be admitted to the colleges. As can be seen from the solid line in Figure 6 and from Figure F.8, students at very bottom of the ability distribution with high parental transfers benefit from less informative signals because they can now more easily enroll in College 2. Similarly, the

[11] Figure F.7 in the appendix plots the change in welfare for a range of possible increases in the signal variance.

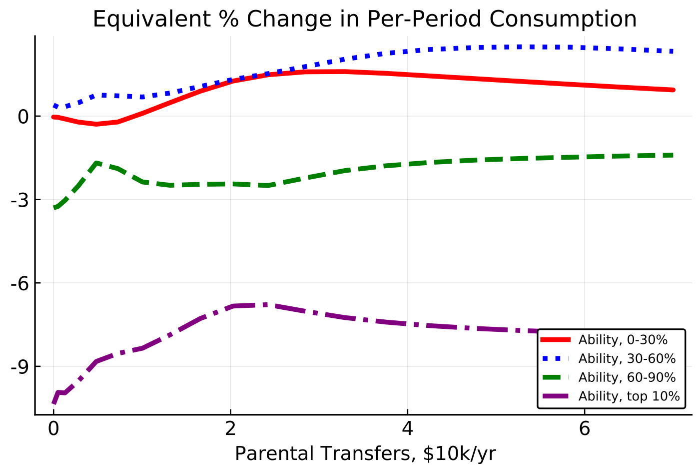


Figure 6: Percent of lifetime consumption students of different parental-income and ability groups are willing to give up in order to switch to the equilibrium with less informative signals.

highest-income students in the 30-60% ability range also benefit from being able to more easily sort into College 1, while the lower-income students in the 30-60% ability range benefit from more easily sorting into College 2. Finally, we find very little change in welfare for low-ability, low-income students because gains they experience from the increased chance of being admitted are offset by the decreases in the financial-aid they can expect to receive.

6.2 Perfectly Informative Signals

To further study the effects of the information frictions due to the application and admissions system, we now set the signal to be perfectly informative so that the colleges may observe the student's true ability (i.e. $g(\sigma|\ell) = 1$ for $\sigma = \ell$). In this exercise, colleges set their tuition and admissions standards based on ability instead of the signal, and students know ex-ante whether or not they will be admitted and exactly how much financial aid they will receive. We then compare this perfect information equilibrium to the baseline equilibrium.

The effect of the new equilibrium on each college is shown in Table F.7. If signals are perfectly informative, the marginal cost of enrolling a high-ability student decreases substantially because colleges can perfectly tell them apart. This will lower tuition for high-ability students and raise the average ability within the student body, thereby increasing the marginal cost for relatively lower-ability students. As the marginal cost of enrolling high-income, low-ability student rises, colleges will increase their admissions threshold to exclude them since the tuition cap is not high enough to justify admitting them. The higher average ability makes it more costly to admit lower ability students, leading to an overall decrease in enrollment. Additionally, the decreases in marginal cost for the high-ability enrollees leads to greater levels of financial-aid which lowers tuition revenue. This drop in revenue lowers the average instructional spending per students, but the increase in average student ability causes the peer-effect component of college quality to be high enough to increase the overall value-added at both colleges.

The effect of perfectly observable signals on student enrollment at each college is shown in Figure 7 below. We see that sorting based on ability is vastly improved in the new equilibrium, with students at the bottom of the ability distribution substantially reducing their enrollment. There is a large drop in students from the 30-60% ability group enrolling into College 2, replaced by a large increase in students from the 60-90% range and low-income students from the top 10% group. College 1 is now able to enroll students almost exclusively from the top 10% of the ability distribution, mostly to the benefit of high-income students who no longer have to compete with lower-ability applicants.

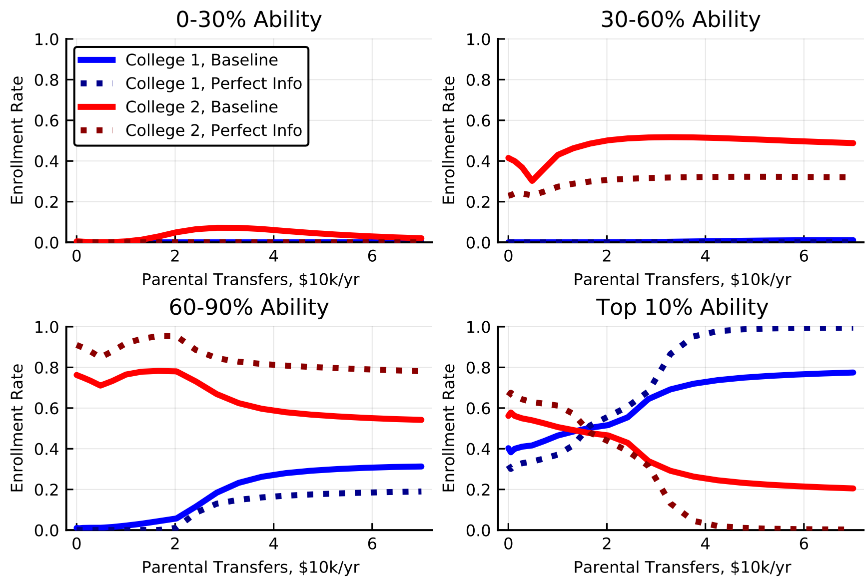


Figure 7: Average enrollment rates for students by parental transfers in the baseline equilibrium and perfect information equilibrium.

Perhaps surprisingly, perfect information actually *reduces* the enrollment of high-ability, low-income students in College 1. This seems puzzling at first because tuition decreases at College 1 for high-ability students since they can now be perfectly sorted. The reason for the decrease in College 1 enrollment is that College 2 now lowers its tuition enough to attract these high-ability students enough to switch relative to the baseline. College 2 could not offer such high levels of financial-aid before because their pool of low-income applicants was much weaker since it included many relatively low-ability applicants. With perfect information, high-ability low-income students cannot be mistaken for lower-ability applicants.

To understand the strength of college competition under perfect information, Figure 8 shows College 1 enrollment for students in the top 10% of the ability distribution in a partial equilibrium world where uncertainty disappears for students, but tuition and admissions policies remain the same (the dotted line). In this scenario, the high-ability, low-income students apply and enroll in College 1 at higher rates since they know they will be admitted at low tuition levels. However, when colleges adjust their policies in response to signals becoming perfectly informative, the low-income students prefer to attend College 2 where tuition is even lower.

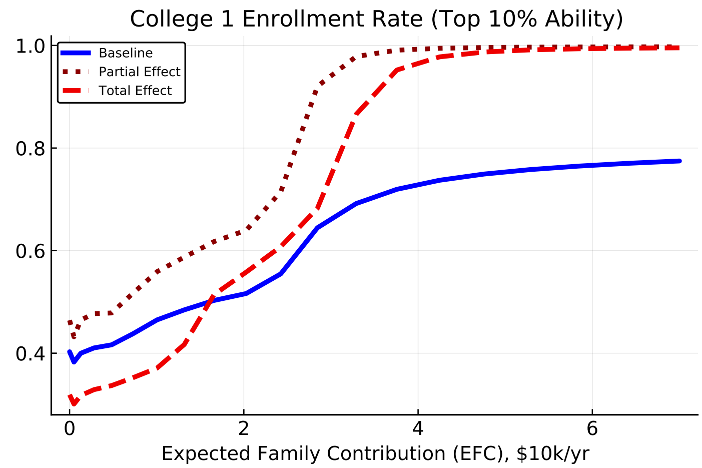


Figure 8: Average College 1 enrollment rate by EFC for students in the top 10% of the ability distribution. The 'Partial Effect' shows how enrollment changes when students have perfect information, but tuition and admissions are the same as in the baseline. The 'Total Effect' shows the perfect-information equilibrium where colleges adjust tuition and admissions.

7 Policy Experiment: Increasing Pell Grants

Increasing the maximum level of Pell Grants was part of President Biden's proposal for higher education policy on the campaign trail. It is now being discussed more concretely as part of the recently released "American Families Plan", which calls for the maximum to be raised to $7,895 per year. In this section, we study the effects of increasing the Pell Grant maximum to $25,000 per year. We choose this level, which is equal to the tuition cap at College 1, because we find that it is the welfare maximizing value for the Pell Grant maximum. Note that students receive Pell Grant aid if their Expected Family Contribution (or parental transfer in our model) is lower than the set Pell Grant maximum. The amount they receive is then equal to the difference between the maximum and their EFC (see 28). Note that in our experiment, the increase in Pell Grants is paid for by tax increases, as implied by the government budget in Equation (23).

The effects on colleges of increasing the Pell Grant maximum can be seen in Table F.8, and the effects on student enrollment can be seen in Figure F.9. In Table F.8 we see that there is a large effect on the income distribution within College 1, with fewer students coming from the top income quartile. By increasing the funding available for lower-income students, College 1 will now charge them higher tuition as they enroll at higher rates, which will make the income distribution more diffuse and lead to an overall increase in tuition revenue and instructional spending per student. With costs no longer an issue for lower-income, high-ability students, they now enroll at higher rates which increases the average ability of the student body. Note however, that this increase is quantitatively modest, suggesting that the newly enrolled high-ability students from lower income groups are not much higher ability on average than the high-income students they replace.

Figure F.9 illustrates the effect of the higher Pell Grants on student sorting, where enrollment is plotted against parental transfers for students at different parts of the ability distribution. We see that for College 1, the enrollment profiles across parental income become flat conditional on ability. This is important for high-income students in the 60-90% group, who now enroll at College 1 at much lower rates. Interestingly, we also see that there is a small increase in College 2 enrollment for low-income students in the lowest ability group. As shown in Table F.8, this has the effect of lowering the average ability in College 2 enough to lower its value-added.

Finally, we examine the welfare effects of the policy change. Figure 9 plots the percent change in period consumption students of different income and ability groups would be willing to forgo in order to switch to the equilibrium with higher Pell Grants. As expected, we see that the students who benefit from the policy are the relatively low-income, college-going students. They benefit directly from the increased consumption while in college (which also alleviates the effect of the credit constraint), and from more easily being able to sort into the colleges. Higher-income students are made worse off from the policy due to the higher taxes, and due to higher competition in enrollment with lower-income students who can now attend more easily. Note that for students at the very bottom of the ability distribution, the only change is the loss experienced from higher taxes since they enroll at very low rates in either equilibrium. Overall, we find that average welfare change of the policy in terms of per-period consumption equivalent units is 1.97%.

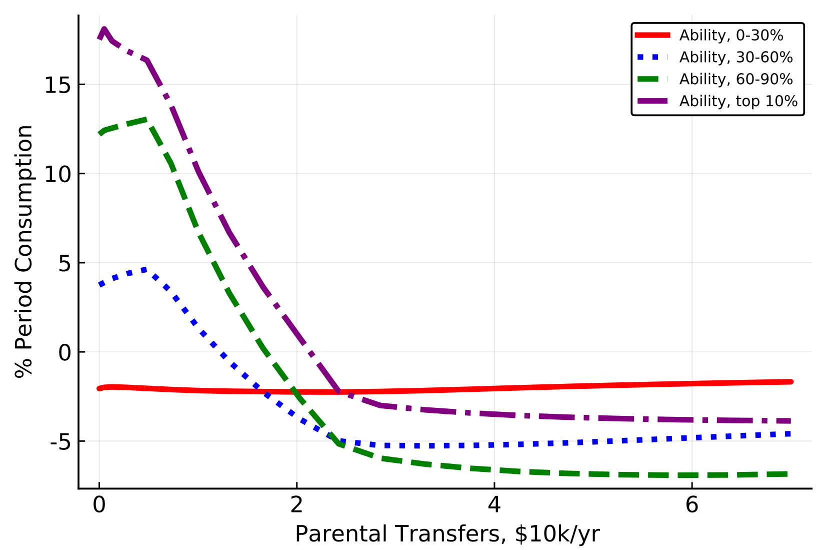


Figure 9: Percent of lifetime consumption students of different EFC and ability are willing to give up to switch to the equilibrium with the higher Pell Grant maximum.

8 Conclusion

This paper studies the role of the admissions system in shaping the allocation of students in the college market. Using micro-level data on high-school students transitioning to college, we showed that parental income is correlated with college applications and enrollment. We found that higher-income students are more likely to apply to college not only at the extensive margin, but also at the intensive margin by applying to more selective colleges. Moreover, we provided suggestive evidence that applicants face risk not only in the college admissions decision, but also in the financial-aid decision as many students reported that they were unable to attend their preferred college due to costs.

Motivated by our empirical findings, we then built an equilibrium model of the college market with student heterogeneity and a non-trivial application and admissions system. We found that our model is able to jointly reconcile the income differences in application and enrollment rates on the student side of the market, and high levels financial-aid available to low-income students on the college side of the market. Through the lens of our model, we learned that lower-income students apply to selective colleges at lower rates because of expectations that they will not receive sufficient financial-aid. Since a college education and student ability are complementary, and since higher-ability students expect higher average application signals, we found that only the highest-ability among the low-income students apply to selective colleges. This makes the selective colleges confident that their low-income applicants are of high ability, justifying the high levels of financial-aid we observe.

In focusing on the role of applications, we have abstracted from many important issues in the college market. By assuming there are only two colleges, we were unable to distinguish between public and private colleges. This is an important distinction since the funding for state schools depends on state governments, and these schools are able to offer substantially lower levels of tuition to in-state students. Adding two more colleges to allow for this distinction would complicate the model, but would allow for more realistic competition on the college side of the market. We have also abstracted from the source of the parental transfers. In reality, a parent's willingness to invest in their children's education is conditional on the student's decisions, and is an important margin of adjustment in response to policy. Finally, we have assumed that students are guaranteed to graduate and can perfectly forecast their post-college earnings. In reality students may face substantial drop-out and post-college earnings risk, which may be an important factor in determining a student's willingness to pursue a college education. We leave these considerations for future research.

References

B. Abbott, G. Gallipoli, C. Meghir, and G. L. Violante. Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624, 2019.

M. J. Bailey and S. M. Dynarski. *Inequality in Postsecondary Education*, pages 117–132. Russell Sage Foundation, 2011.

R. Bénabou. Tax and education policy in a heterogeneous-agent economy: What levels of redistribution maximize growth and efficiency? *Econometrica*, 70(2):481–517, 2002.

Z. Cai and J. Heathcote. College tuition and income inequality. Working paper, July 2018.

D. Capelle. The great gatsby goes to college: Tuition, inequality and intergenerational mobility in the us. *Working Paper*, 2019.

E. Caucutt and L. Lochner. Early and late human capital investments, borrowing constraints, and the family. *Working Paper*, 2017.

H. Chade, G. Lewis, and L. Smith. Student portfolios and the college admissions problem. *Review of Economic Studies*, 81:971–1002, 2014.

R. Chetty, J. N. Friedman, E. Saez, N. Turner, and D. Yagan. Income Segregation and Intergenerational Mobility Across Colleges in the United States*. *The Quarterly Journal of Economics*, 02 2020.

M. Colas, S. Findeisen, and D. Sachs. Optimal need-based financial aid. *Journal of Political Economy*, 129(2):492–533, 2021.

J. M. Delaney and P. J. Devereux. Choosing differently? college application behavior and the persistence of educational advantage. 2020.

E. Dillon and J. Smith. Determinants of the match between student ability and college quality. *Journal of Labor Economics*, 35:45–66, 2017.

S. Dynarski, C. Libassi, K. Michelmore, and S. Owen. Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students. Working Paper 25349, National Bureau of Economic Research, 2018.

D. Epple, R. Romano, and H. Sieg. Admission, tuition, and financialaid policies in the market for higher education. *Econometrica*, 74:885–928, 2006.

D. Epple, R. Romano, S. Sarpca, and H. Sieg. A general equilibrium analysis of state and private colleges and access to higher education in the u.s. *Journal of Public Economics*, 155:164–178, 2017.

R. Fernandez and R. Rogerson. Income distribution, communities and the quality of public education. *The Quaterly Journal of Economics*, 111(1), 1996.

I. Fillmore. Price discrimination and public policy in the u.s. college market. *Working Paper*, 2020.

C. Fu. Equilibrium tuition, applications, admissions and enrollment in the college market. *Journal of Political Economy*, 122(2):225–281, 2014.

G. Gordon and A. Hedlund. Accounting for the rise in college tuition. Working Paper 21967, National Bureau of Economic Research, February 2016.

C. M. Hoxby and C. Avery. The missing "one-offs": The hidden supply of high-achieving, low income students. *Brookings Papers on Economic Activity*, 2014.

F. Ionescu and N. Simpson. Default risk and private student loans: Implications for higher education policies. *Journal of Economic Dynamics and Control*, 64(1):119–147, 2016.

T. Kotera and A. Seshadri. Educational policy and intergenerational policy. *Review of Economic Dynamics*, 25:187–207, 2017.

D. Krueger and A. Ludwig. On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. *Journal of Monetary Economics*, 77(C):72–98, 2016.

L. J. Lochner and A. Monge-Naranjo. The nature of credit constraints and human capital. *American Economic Review*, 101(6):2487–2529, October 2011.

D. O. Lucca, T. Nadauld, and K. Shen. Credit Supply and the Rise in College Tuition: Evidence from the Expansion in Federal Student Aid Programs. *The Review of Financial Studies*, 32(2):423–466, 2018.

Appendix

A Additional Empirical Results

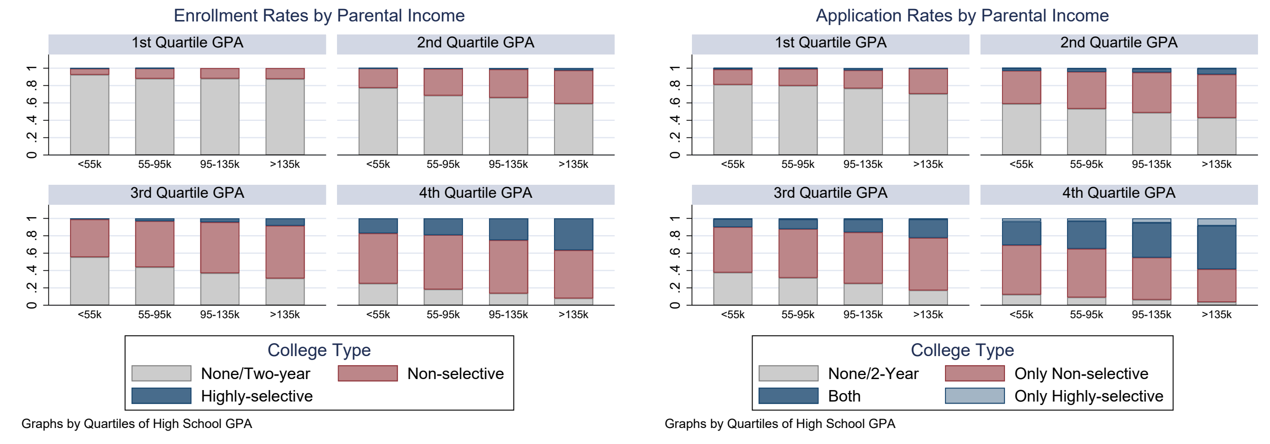


Figure A.1: College application and enrollment by parental income and student GPA. Left panel: Student enrollment. Right panel: Student applications. Source: HSLS.

Dep Var: Apply at all	**(1)**	**(2)**	**(3)**	**(4)**
Parental Income	0.101***	0.095***	0.093***	0.100***
	(0.012)	(0.014)	(0.014)	(0.016)
GPA (std)	1.208***		1.404***	
	(0.076)		(0.083)	
SAT (std)		0.745***		0.935***
		(0.098)		(0.104)
Parental Inc.*GPA/SAT	0.059***	0.067***	0.034*	0.046**
	(0.016)	(0.019)	(0.017)	(0.020)
Controls	No	No	Yes	Yes
N	15,738	12,007	14,711	11,429
R^2	0.262	0.151	0.314	0.194

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Controls include race, sex, parental education, parental relationship pattern, household size, and high-school characteristics including public vs. private, region, and percent of students receiving free or subsidized lunches.

Table A.1: Logit estimation results for whether or not students applied to any non-profit 4-year colleges. Data: HSLS.

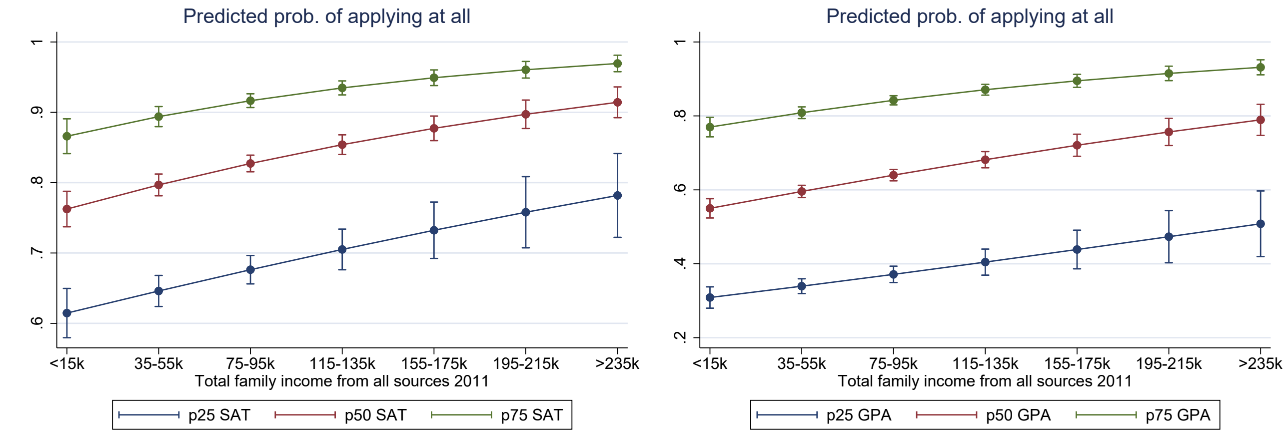


Figure A.2: Predicted relationship between SAT scores (left panel) / GPA (right panel), parental income, and application decisions implied by estimated logit model. See Table A.1 for estimation results.

Dep Var: Apply to selective|applied	**(1)**	**(2)**	**(3)**	**(4)**
Parental Income	0.032	0.003	-0.001	0.005
	(0.027)	(0.019)	(0.024)	(0.018)
GPA (std)	0.947***		1.104***	
	(0.166)		(0.146)	
SAT (std)		0.968***		1.106***
		(0.123)		(0.119)
Parental Inc.*GPA/SAT	0.079***	0.072***	0.071***	0.058***
	(0.024)	(0.018)	(0.022)	(0.018)
Controls	No	No	Yes	Yes
N	10,083	9,387	9,593	8,979
R^2	0.170	0.208	0.228	0.247

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Controls include race, sex, parental education, parental relationship pattern, household size, and high-school characteristics including public vs. private, region, and percent of students receiving free or subsidized lunches.

Table A.2: Logit estimation results for whether or not students applied to any Highly-selective colleges, conditional on applying to non-profit 4-year colleges. Data: HSLS.

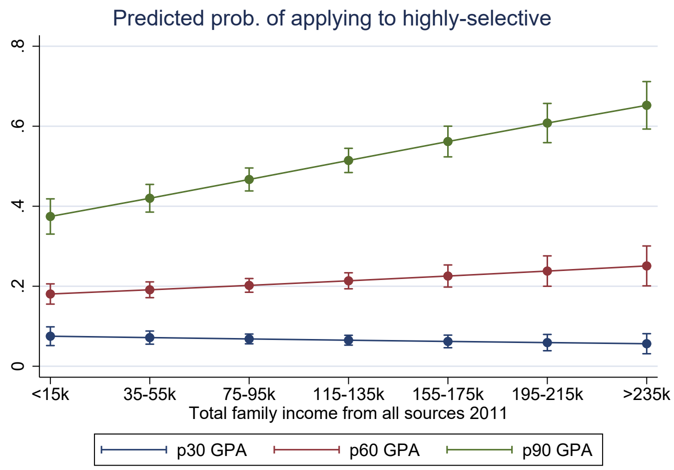


Figure A.3: Predicted relationship between GPA, parental income, and the decision to apply to a Highly-selective college implied by estimated logit model. See Table A.2 for estimation results.

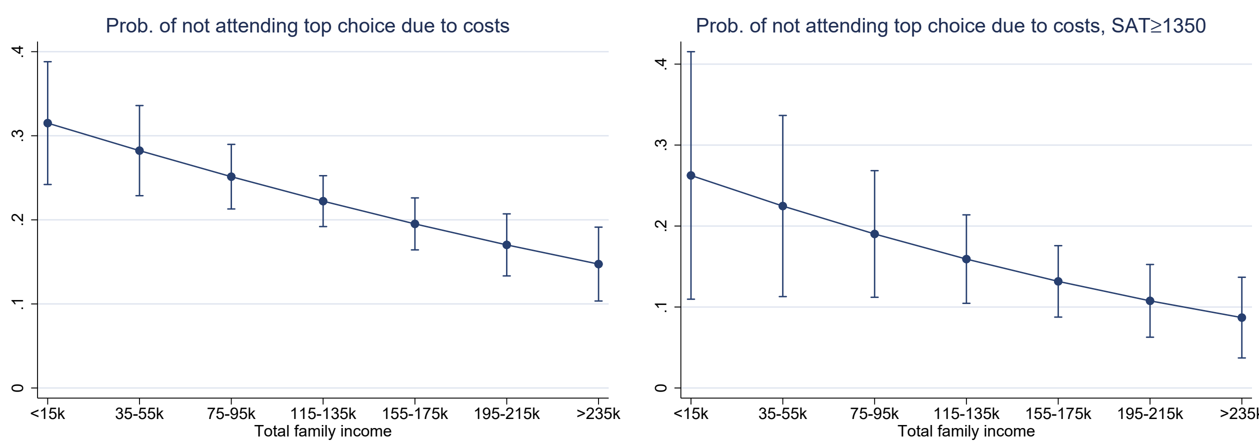


Figure A.4: Predicted probability that a student does not attend their preferred college they were admitted to due to costs. Predictions are from a probit model estimated from a subsample of students in the HSLS who were admitted to both Highly-selective and Non-selective colleges. The right panel further restricts to all such students who scored at least 1350 on their SAT.

B Data Appendix

B.1 Barron's Selectivity Index

List of Barron's Tier 1 and 2 Colleges and Universities (Alphabetical)
Tier 1: Amherst College, Barnard College, Bates College, Boston College, Bowdoin College, Brown University, Bryn Mawr College, Bucknell University, California Institute of Technology, Carleton College, Carnegie Mellon University, Case Western Reserve University, Claremont McKenna College, Colby College, Colgate University, College of Mount Saint Vincent, College of the Holy Cross, College of William & Mary, Colorado College, Columbia University/City of New York, Connecticut College, Cooper Union for the Advancement of Science and Art, Cornell University, Dartmouth College, Davidson College, Duke University, Emory University, Franklin and Marshall College, George Washington University, Georgetown University, Georgia Institute of Technology, Hamilton College, Hampshire College, Harvard University/Harvard College, Harvey Mudd College, Haverford College, Johns Hopkins University, Kenyon College, Lehigh University, Macalester College, Massachusetts Institute of Technology, Middlebury College, New York University, Northeastern University, Northwestern University, Oberlin College, Ohio State University at Marion, Pitzer College, Pomona College, Princeton University, Reed College, Rensselaer Polytechnic Institute, Rice University, Rose-Hulman Institute of Technology, Santa Clara University, Smith College, Southern Methodist University, Stanford University, Swarthmore College, The Ohio State University, Tufts University, Tulane University, Union College, United States Air Force Academy, United States Military Academy, United States Naval Academy, University of California at Berkeley, University of California at Los Angeles, University of Chicago, University of Miami, University of Missouri/Columbia, University of North Carolina at Chapel Hill University of Notre Dame, University of Pennsylvania, University of Richmond, University of Rochester, University of Southern California, University of Virginia, Vanderbilt University, Vassar College, Villanova University, Wake Forest University, Washington and Lee University, Washington University in St. Louis, Webb Institute, Wellesley College, Wesleyan University, Whitman College, Williams College, Yale University

Tier 2: Allegheny College, American University, Augustana College, Austin College, Babson College, Bard College, Bard College at Simon's Rock, Baylor University, Beloit College, Bennington College, Bentley University, Berea College, Berry College, Binghamton University/The State University of New York, Boston University, Brandeis University, Brigham Young University, California Polytechnic State University, Centre College, Christian Brothers University, Clark University, Clarkson University, Clemson University, College of New Jersey, College of the Atlantic, Colorado School of Mines, Cornell College, CUNY/City College, Denison University, Dickinson College, Drexel University, Elon University, Emerson College, Florida State University, Fordham University, Furman University, Gettysburg College, Gonzaga University, Grinnell College, Grove City College, Gustavus Adolphus College, Hendrix College, Hillsdale College, Illinois Institute of Technology, Indiana University Bloomington, Ithaca College, Kalamazoo College, Kettering University, Lafayette College, Lawrence University, Miami University, Mills College, Mount Holyoke College, Muhlenberg College, New College of Florida, New Mexico Institute of Mining and Technology, North Carolina State University, Pepperdine University, Polytechnic Institute of New York University, Providence College, Purdue University/West Lafayette, Rhodes College, Rollins College, Sarah Lawrence College, Sewanee: The University of the South, Skidmore College, St. John's College, Santa Fe, St. John's College-Annapolis, St. Lawrence University, St. Mary's College of Maryland, St. Olaf College, State University of New York / College of Environmental Science and Forestry, Stevens Institute of Technology, Stony Brook University / State University of New York, SUNY College at Geneseo, Syracuse University, Texas Christian University, Trinity College, Trinity University, Truman State University, United States Coast Guard Academy, United States Merchant Marine Academy, University of California at Davis, University of California at Santa Barbara, University of Connecticut, University of Florida, University of Illinois at Urbana-Champaign, University of Maryland, University of Michigan/Ann Arbor, University of Minnesota/Twin Cities, University of Pittsburgh at Pittsburgh, University of Puget Sound, University of San Diego, University of Texas at Austin, University of Texas at Dallas, University of Tulsa, University of Wisconsin/Madison, Virginia Polytechnic Institute and State University, Westmont College, Wheaton College, Wheaton College, Worcester Polytechnic Institute

B.2 High School Longitudinal Study of 2009 (HSLS:09)

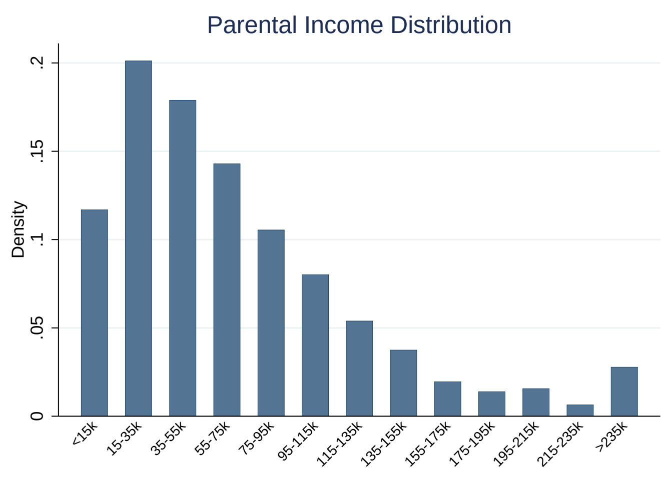


Figure B.5: Distribution of parental income in the HSLS.

B.3 IPEDS Sample Restrictions

The Integrated Postsecondary Education Data System is made publicly available through the National Center for Education Statistics. Since we are interested only in four-year nonprofit U.S. colleges and universities, we make the following restrictions to the universe of colleges in the system:

- U.S. only; Title IV participating; Degree-granting
- Undergraduate enrollment at least 100

- No Theological/faith related institutions
- No 2 year colleges
- No for-profit colleges

We restrict our IPEDS data sample to cover the 2013-2016 time frame, since it is the relevant 4 year period for our HSLS cohort who begin college in 2013. Our final sample includes 1,665 four-year colleges and universities.

C Estimation Appendix

C.1 EFC Calculation

For students who did not fill out the FAFSA, we calculate their EFC directly using the 2013-2014 EFC formula with data from the HSLS survey. To calculate EFC, one must first calculate Adjusted Available Income (AAI), which combines household income net of allowances (which depend on household size) and household assets (excluding the family's home). Since the HSLS does not report assets, we assume that the contribution from assets is 0.

The parents' contribution from AAI is then calculated from a (progressive) non-linear function of AAI, described in the table below:

If parents' AAI is –	**Parents' contribution from AAI is –**
Less than -\$3,409	-\$750
-\$3,409 to \$15,300	22% of AAI
\$15,301 to \$19,200	25% of AAI over \$15,300 + \$3,366
\$19,201 to \$23,100	29% of AAI over \$19,200 + \$4,341
\$23,101 to \$27,000	34% of AAI over \$23,100 + \$5,472
\$27,001 to \$30,900	40% of AAI over \$27,000 + \$6,798
\$30,901 or more	47% of AAI over \$30,900 + \$8,358

EFC is then the parents' contribution divided by the number of children that are enrolled in college. The HSLS specifies only if students have a sibling in college at the same time, so to find EFC we divide the parents' contribution by two if the student does have a sibling in college. We are thus able to construct EFC for the students in our sample even if they did not complete the FAFSA.

C.2 Distribution of Student Characteristics

In the HSLS, we find that approximately 30% of the students have an EFC of 0. In fitting our distribution, we therefore assign a mass point of 30% for $y = 0$. For the remaining 70% of the distribution with $y > 0$, we assume that y follows a log-normal distribution, $y \sim LogN(\mu_y, \sigma_y^2)$.

We use the students' SAT score in the HSLS as a proxy for ability to determine the distribution of ℓ. Again, we assume that ℓ follows a log-normal distribution, $\ell \sim LogN(\mu_\ell, \sigma_\ell^2)$. Since y and ℓ are correlated, we estimate these parameters separately for the case where $y = 0$ and where $y > 0$.

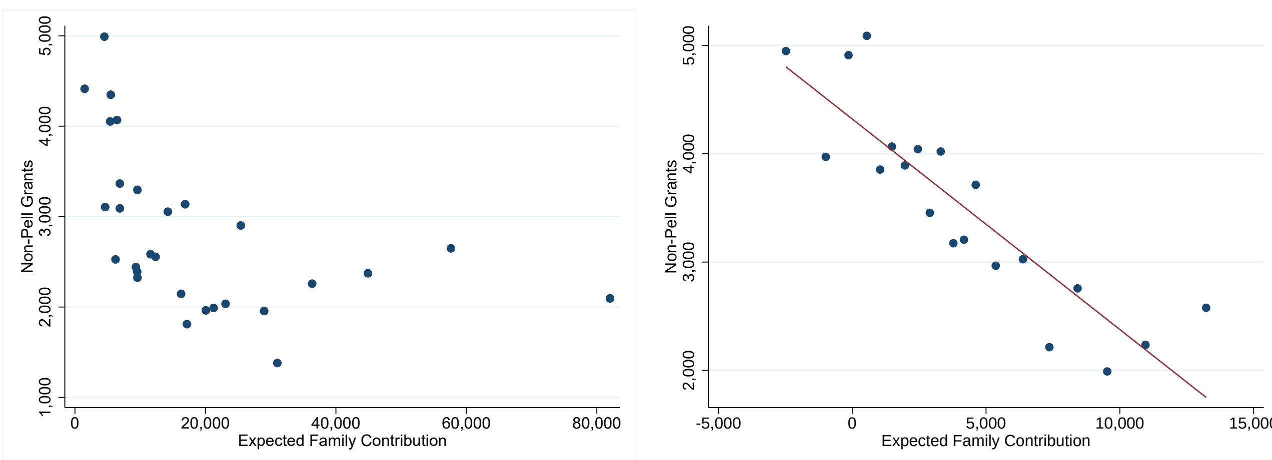
To summarize, the joint distribution of (y, ℓ) is given by:

$$\begin{cases} (y = 0, \ell) \sim LogNormal(\mu_{\ell 0}, \sigma_{\ell 0}^2) & w.p. \;\; 30\% \\ (y > 0, \ell) \sim LogNormal\left(\begin{bmatrix} \mu_a \\ \mu_{\ell 1} \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{a\ell} \\ \sigma_{a\ell} & \sigma_\ell^2 \end{bmatrix} \right) & w.p. \;\; 70\% \end{cases}$$

Using the HSLS, we find $\mu_{\ell 0} = -0.144, \sigma_{\ell 0}^2 = 0.155$ for $y = 0$, and $\mu_{\ell 1} = 0.135, \sigma_{\ell 1}^2 = 0.138$ for $y > 0$. We also estimate $\mu_y = -0.301, \sigma_y^2 = 2.825$. Finally, to account for this correlation for the case where $y > 0$, we estimate the covariance between ℓ and a to be $\sigma_{a\ell} = 0.174$.

C.3 Grant Function Estimation

Since all college-specific grants (both need-based and merit-based) and federal Pell Grants are accounted for in our model, we add the exogenous grants, denoted by $Gr(y)$, to stand in for all other grants that students receive. We identify all such grants in our BPS data by adding up all grants and aid excluding loans, Pell Grants, and institutional aid. As we can see from the binned scatter-plot below, these grants vary substantially with Expected Family Contribution (EFC).



These grants are decreasing in EFC and tend to level off at around \$15,000. For EFC below this level, we estimate a downward sloping linear relationship between EFC and non-Pell Grants using OLS (and controlling for SAT score). We find $g_0 = 0.43, g_1 = 0.19$ in (27), corresponding to right panel of the plot above. We pick $\underline{g} = 0.15$ to capture that the grants level off at a sufficiently high EFC level.

C.4 College Cost Function Estimation

Before estimating the college-specific parameters using our IPEDS sample, we first confirm that the sample includes the relevant colleges relative to the HSLS. We compare the share of total student enrollment accounted for by each type of college[12] in our IPEDS sample with the enrollment shares observed in the HSLS. For students in our HSLS sample, we find the following fractions at each type of college:

$$\kappa^1 = 0.069 \qquad \kappa^2 = 0.357.$$

To confirm that these numbers are consistent with our IPEDS data, we first add up all full-time equivalent undergraduate students across all colleges in our sample. We then infer the total number of high school graduates using the 42% 4-year enrollment rate reported by NCES[13]. We find that of this population, 7% attend College 1 and 35% College 2, similar to the HSLS enrollment patterns.

Confident that our IPEDS sample includes all colleges relevant for our analysis, we use variation in enrollment size for colleges to estimate the parameters of the cost function introduced in (29). We start by estimating the parameters in the following regression separately for each type of college:

$$cost_i^s = \hat{O}_0^s + \hat{O}_2^s (k_i^s)^2 + \varepsilon_i, \tag{31}$$

for college i of college type s, where k is the enrollment level of college i. In order to measure the costs for each college in our sample, we use the detailed financial data available from IPEDS. There are two methods to measure costs: one is by directly adding up all non-instructional expenditure including academic support, student services, and institutional support. The other is by using the budget constraint in Equation (18), where we calculate cost by adding up all tuition revenue, net grant revenue, government appropriations, unrestricted revenue from private sources, and subtracting off total instructional expenditure. We rely on both methods by defining costs in the left hand side of equation (31) as the maximum over both methods (this helps us deal with cases where the second procedure produces small or negative numbers). The results of the estimation are provided in Table C.3 below.

Next, we follow Epple et al. (2006) in their aggregation procedure to transform our cost function parameter estimates from (31) into our corresponding model parameters. If there are n_s colleges in type s, we assume they are identical in the model so that total enrollment K^s simply scales up the individual enrollments: $K^s = \sum_{i=1}^{n_s} k_i^s = n_s k^s$. We can then sum up the individual cost functions:

$$C(K^s) = \sum_{i=1}^{n_s} C(k_i^s) = \sum_{i=1}^{n_s} \left[O_0^s + O_2^s (k_i^s)^2 \right]$$

$$C(K^s) = n_s O_0^s + n_s O_2^s (k^s)^2$$

$$C(K^s) = o_0^s + o_2^s (K^s)^2$$

where $o_0^s = n_s O_0^s$, and $o_2^s = O_2^s / n_s$. We have 182 colleges in type 1, and 1,483 in type 2, so we find that $(o_2^1, o_2^2) = (5.77, 0.44)$, and $(o_0^1, o_0^2) = (0.15, 0.25)$ based on our estimates of the parameters from (31).

College subsidies. Similar to the cost function estimation above, we estimate Tr_1^s for each College type s using our IPEDS sample. We rely on variation in enrollment within each college type, and

[12] Where college types are based on the Barron's ranking as described in Section 4.3.

[13] See here: `https://nces.ed.gov/programs/coe/indicator_cpa.asp`.

	(1)	(2)
	$costs^1$	$costs^2$
Enrollment squared	1057.6***	624.2***
	(120.4)	(14.85)
Private college		-0.0000919***
		(0.0000153)
Constant	0.000841***	0.000222***
	(0.000111)	(0.0000126)
N	182	1445
adj. R^2	0.296	0.604

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.3: Estimates for the cost parameters in (31) for each college type. Note that all variables have been scaled down by the total student population, and the costs have been further scaled down by $40,000 to fit the model units.

measure subsidies to colleges by the level of state government appropriations and federal funding they receive. We simply estimate the slope in the following regression relating government transfers to enrollment levels:

$$transfer_i^s = \hat{Tr}_1^s k_i + \varepsilon_i.$$

The results of the estimation are given in Table C.4 below.

C.5 Net tuition variation within college types

We use the BPS to show how net tuition, defined as sticker-price tuition minus college-specific grants (need- and merit-based financial-aid) varies with student Expected Family Contribution (EFC). Figure C.6 below provides binned-scatter plots which show how net-tuition levels off for students with high EFC who are less likely to receive financial-aid from the colleges.

	(1)	(2)
	$transfer^1$	$transfer^2$
$Enrollment$	1.044***	0.620***
	(0.0516)	(0.0112)
N	182	1445
adj. R^2	0.691	0.679

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table C.4: Estimates for the relationship between government transfers to colleges and their enrollment level.

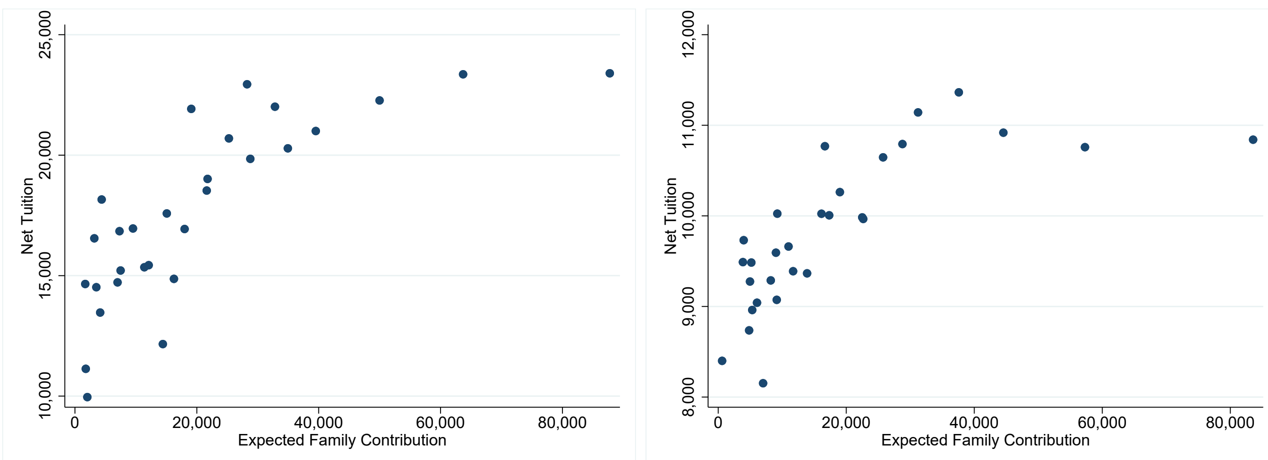


Figure C.6: Binned-scatter plots for net tuition (sticker-price tuition minus college-specific grants) for highly-selective colleges (left panel) and non-selective colleges (right panel). The data come from the 2012 cohort of the Beginning Postsecondary Students Longitudinal Study (BPS).

D Proofs of Propositions

D.1 Derivation of (20)

Let $\lambda_I, \lambda_L, \lambda_\kappa$ be the Lagrange multipliers on the budget constraint (18), average ability identity constraint (16), and enrollment identitiy constraint (15) respectively. First order conditions for $I_\mu, L_\mu, \kappa, T(y, \sigma)$ are then

$$
\begin{aligned}
&[\kappa] && -\lambda_\kappa = \lambda_L L + \lambda_I [I + C'(\kappa) - Tr'(\kappa)] \\
&[I_\mu] && \xi_s Q_I = \lambda_I \kappa \\
&[L_\mu] && \xi_s Q_L = \lambda_L \kappa \\
&[T(y,\sigma)] && T(y,\sigma) = -\frac{\lambda_\kappa}{\lambda_I} - \frac{\lambda_L}{\lambda_I} \frac{\int \ell \frac{\partial \tilde{q}}{\partial T(y,\sigma)} g(\sigma|\ell)\mu(y,d\ell)}{\int \frac{\partial \tilde{q}}{\partial T(y,\sigma)} g(\sigma|\ell)\mu(y,d\ell)} \qquad (32) \\
& && \qquad - \frac{\int \tilde{q}\ g(\sigma|\ell)\mu(y,d\ell)}{\int \frac{\partial \tilde{q}}{\partial T(y,\sigma)} g(\sigma|\ell)\mu(y,d\ell)} + \frac{\lambda_T(y,\sigma)}{\lambda_I \int \frac{\partial \tilde{q}}{\partial T(y,\sigma)} g(\sigma|\ell)\mu(y,d\ell)}
\end{aligned}
$$

where $\lambda_T(y,\sigma)$ is the Lagrange multiplier for the constraint on tuition. Combining these equations gives (20).

D.2 Derivation of (22)

Let $\tilde{q}(\underline{\sigma}_s) = \tilde{q}(y,\ell,\underline{\sigma}_s, T^s(y,\underline{\sigma}_s), T^{-s}(y,\underline{\sigma}_s))$. The first order condition the admissions standard is

$$
\frac{\int T(y,\underline{\sigma}_s)\tilde{q}(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell)}{\int \tilde{q}(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell)} \geq -\frac{\lambda_\kappa}{\lambda_I} - \frac{\lambda_L}{\lambda_I} \frac{\int \ell\ \tilde{q}(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell)}{\int \tilde{q}(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell)}, \qquad (33)
$$

which holds with equality when $\underline{\sigma}_s > 0$. Combining with (32), we arrive at (22).

D.3 Proposition 1

Proof. Combining (22) and (20), we find

$$
\begin{aligned}
&\frac{Q_{L_\mu}}{Q_{I_\mu}} \int \left(\ell - \frac{\int \ell \frac{\partial \tilde{q}^s(\underline{\sigma}_s)}{\partial T^s(y,\underline{\sigma}_s)} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)}{\int \frac{\partial \tilde{q}^s(\underline{\sigma}_s)}{\partial T^s(y,\underline{\sigma}_s)} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)} \right) \tilde{q}^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell) \\
&- \int \frac{\left[\int \tilde{q}^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell)\mu(y,d\ell)\right]^2}{\int \frac{\partial \tilde{q}^s(\underline{\sigma}_s)}{\partial T^s(y,\underline{\sigma}_s)} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)} dy + \frac{\int \lambda_T(y,\underline{\sigma}_s)\tilde{q}^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell) d\mu(y,\ell)}{\lambda_I \int \frac{\partial \tilde{q}^s(\underline{\sigma}_s)}{\partial T^s(y,\underline{\sigma}_s)} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)} \geq 0. \qquad (34)
\end{aligned}
$$

Since the constraint on tuition is never binding, we have $\lambda_T(y,\sigma) = 0$. Rewrite the left hand side of (34) to obtain

$$
\begin{aligned}
&\frac{Q_{L_\mu}}{Q_{I_\mu}} \int \left[\int q^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell)\mu(y,d\ell) \right] \underbrace{\left[\frac{\int \ell q^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell)\mu(y,d\ell)}{\int q^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell)\mu(y,d\ell)} - \frac{\int \ell \frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)}{\int \frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s} g(\underline{\sigma}_s|\ell)\mu(y,d\ell)} \right]}_{\geq 0} dy \\
&+ \underbrace{\int \frac{\left[\int q^s(\underline{\sigma}_s) g(\underline{\sigma}_s|\ell)\mu(y,d\ell;y)\right]^2}{\int -\frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s(y,\underline{\sigma}_s)} g(\underline{\sigma}_s|\ell)\mu(y,d\ell;y)} dy}_{>0}, \qquad (35)
\end{aligned}
$$

where $q^s(\underline{\sigma}_s) = \tilde{q}^s(y, \ell, T(y, \underline{\sigma}_s))$. We show that (35) is strictly positive for all $\underline{\sigma_s} > \underline{\sigma}$, so that the college does not choose an interior admissions standard.

Clearly, the term on the bottom is strictly positive provided there is a positive mass of students for which $q^s(\underline{\sigma}_s) > 0$ and $\frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s} < 0$. To see that the bracketed term on the top is positive (for each a), note that it is the difference in means of ℓ according to two distributions, H_0 and H_1, with associated densities: $h_0(\ell) = \frac{q^s(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)\mu(y,\ell)}{\int q^s(\underline{\sigma}_s)g(\underline{\sigma}_s|\ell)\mu(y,d\ell)}$ and $h_1(\ell) = \frac{\frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s}g(\underline{\sigma}_s|\ell)\mu(y,\ell)}{\int \frac{\partial q^s(\underline{\sigma}_s)}{\partial T^s}g(\underline{\sigma}_s|\ell)\mu(y,d\ell)}$. The term being positive is then implied by the fact that H_0 has first-order stochastic dominance over H_1.

To see that this is the case, define

$$\begin{aligned} F(\hat{\ell}) &= H_0(\hat{\ell}) - H_1(\hat{\ell}) \\ &= \int_0^{\hat{\ell}} h_0(\ell)d\ell - \int_0^{\hat{\ell}} h_1(\ell)d\ell \end{aligned}$$

Note that F is continuously differentiable, and $\lim_{\ell\to 0} F(\ell) = 0$, $\lim_{\ell\to\infty} F(\ell) = 0$. The interior solution to the first order condition $F'(\ell^*) = 0$ uniquely minimizes F:

For College 1,

$$\begin{aligned} F'(\ell) &= h_0(\ell) - h_1(\ell) \\ &= \left\{ q^1(\underline{\sigma}_1) - \left[\frac{\int q^1(\underline{\sigma}_1)g(\underline{\sigma}_1|\ell)\mu(a,d\ell)}{\int \frac{\partial q^1(\underline{\sigma}_1)}{\partial T^1}g(\underline{\sigma}_1|\ell)\mu(a,d\ell)}\right]\frac{\partial q^1(\underline{\sigma}_1)}{\partial T^1}\right\} g(\underline{\sigma}_1|\ell)\mu(a,\ell) \\ &= \Bigg\{ \underbrace{\frac{1}{1-q^1(\underline{\sigma}_1)} - \left[\frac{\int q^1(\underline{\sigma}_1)g(\underline{\sigma}_1|\ell)\mu(a,d\ell)}{\int \frac{\partial q^1(\underline{\sigma}_1)}{\partial T^1}g(\underline{\sigma}_1|\ell)\mu(y,d\ell)}\right]\lambda_C V_T^{C1}(y,\ell^*,T(a,\underline{\sigma}_1))}_{G(\ell)} \Bigg\} q^1(\underline{\sigma}_1)g(\underline{\sigma}_1|\ell)\mu(y,\ell)[1-q^1(\underline{\sigma}_1)] \end{aligned}$$

We have a unique interior minimum since,

1. $\lim_{\ell\to 0} G(\ell) < 0$
2. $\lim_{\ell\to\infty} G(\ell) = \infty$
3. $G'(\ell) > 0$

Thus we have $F(\ell) < 0$ for $\ell \in (0, \infty)$.

Since (35) is strictly positive, the lower bound constraint on $\underline{\sigma}_s$ binds, so the college does not use the admissions threshold to screen students. □

E Lifecycle Model

In this section, we describe how a simple lifecycle model used for the calibration maps easily into the two period model introduced in Section 3. Consider an individual who lives for $T + 1$ periods, where a period is four years and the first years are spent in college:

$$\max_{\{c_j,a_{j+1}\}_{j=0,\dots,T}} \quad u(c_0)+\sum_{j=1}^{T}\beta^j u(c_j) \tag{36}$$

$$\text{s.t.} \quad c_0+a_1+T=a_0$$

$$c_j+a_{j+1}=a_jR+w(1-\tau)\gamma_s\ell^{\alpha}, \quad j=1,\dots,T$$

$$a_1\geq \underline{a}_s \tag{37}$$

CRRA preferences imply that from the Euler equation we can write $u(c_{1+t})=(\beta R)^{\frac{t(1-\sigma)}{\sigma}}u(c_1)$, which means that

$$\tilde{\beta}u(c_1)=\sum_{j=1}^{T}\beta^j u(c_j).$$

We also combine the budget constraints for $j=1,\dots,T$, and plug in the result from the Euler equation to get

$$c_1\sum_{j=1}^{T}\frac{(\beta R)^{\frac{j-1}{\sigma_c}}}{R^{j-1}}=a_1R+w(1-\tau)\gamma_s\ell^{\alpha}\sum_{j=1}^{T}\frac{1}{R^{j-1}}.$$

Putting everything together, we have the same two period model from Section 3:

$$\max_{c_0,c_1} \quad u(c_0)+\tilde{\beta}u(c_1) \tag{38}$$

$$\text{s.t.} \quad c_0+a_1+T=a_0$$

$$c_1=a_1\tilde{R}+\tilde{w}(1-\tau)\gamma_s\ell^{\alpha}$$

$$a_1\geq \underline{a}_s \tag{39}$$

where

$$\tilde{\beta}=\sum_{j=1}^{T}\beta^j(\beta R)^{\frac{(j-1)(1-\sigma_c)}{\sigma_c}}$$

$$\tilde{R}=\frac{R}{\sum_{j=1}^{T}\frac{(\beta R)^{\frac{j-1}{\sigma_c}}}{R^{j-1}}}$$

$$\tilde{w}=\frac{w\sum_{j=1}^{T}\frac{1}{R^{j-1}}}{\sum_{j=1}^{T}\frac{(\beta R)^{\frac{j-1}{\sigma_c}}}{R^{j-1}}}$$

F Additional Model Results

F.1 Effect of Equalizing Application Patterns

		College 1		College 2	
		Baseline	*More Applicants*	*Baseline*	*More Applicants*
L_μ	Avg Student Ability	1.86	1.82	1.28	1.3
I_μ	Instr Spending per Student	2.1	2.04	0.98	0.97
Γ_s	Value-added	2.01	1.96	1.79	1.81
κ	% Enrolled	7.15	6.8	32.09	31.09
	Income Distribution				
	Q1 Income	0.1	0.09	0.22	0.22
	Q2 Income	0.12	0.11	0.2	0.21
	Q3 Income	0.26	0.25	0.31	0.31
	Q4 Income	0.52	0.55	0.27	0.26
	Ability Distribution				
	0-30% Ability	0.0	0.0	0.01	0.02
	30-60% Ability	0.0	0.01	0.41	0.39
	60-90% Ability	0.15	0.25	0.4	0.38
	Top 10% Ability	0.84	0.74	0.17	0.21

Table F.5: Effect on college market of fixing application choices for all students to be the same as the application choices of high-wealth students in the baseline estimation

F.2 Effect of Less Informative Signals

		College 1		College 2	
		Baseline	$6 * \sigma_g^2$	*Baseline*	$6 * \sigma_g^2$
L_μ	Avg Student Ability	1.86	1.67	1.28	1.19
I_μ	Instr Spending per Student	2.1	2.4	0.98	1.14
Γ_s	Value-added	2.01	1.87	1.79	1.71
κ	% Enrolled	7.15	8.93	32.09	39.72
	Income Distribution				
	Q1 Income	0.1	0.1	0.22	0.25
	Q2 Income	0.12	0.12	0.2	0.2
	Q3 Income	0.26	0.27	0.31	0.29
	Q4 Income	0.52	0.51	0.27	0.26
	Ability Distribution				
	0-30% Ability	0.0	0.0	0.01	0.14
	30-60% Ability	0.0	0.09	0.41	0.43
	60-90% Ability	0.15	0.3	0.4	0.29
	Top 10% Ability	0.84	0.61	0.17	0.14

Table F.6: Effect on colleges of making signals less informative by increasing the variance of the signal distribution

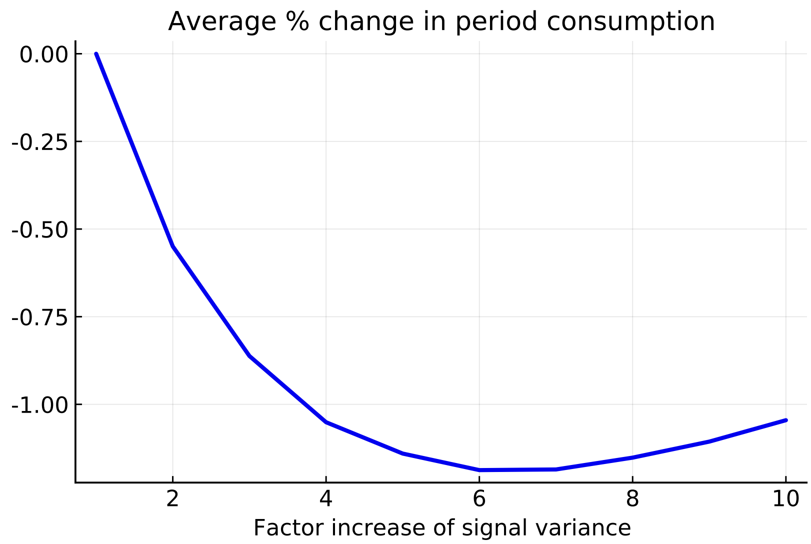


Figure F.7: Welfare changes as a function of increases to the signal variance.

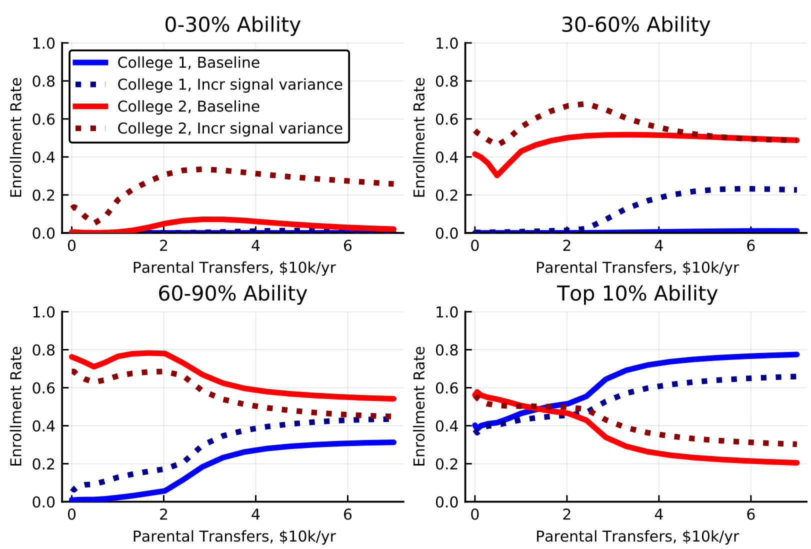


Figure F.8: Student enrollment rates in baseline equilibrium vs. equilibrium with less informative signals.

F.3 Effect of Perfectly Informative Signals

		College 1		College 2	
		Baseline	*Perfect Info*	*Baseline*	*Perfect Info*
L_μ	Avg Student Ability	1.86	1.91	1.28	1.36
I_μ	Instr Spending per Student	2.1	1.87	0.98	0.93
Γ_s	Value-added	2.01	2.02	1.79	1.86
κ	% Enrolled	7.15	6.5	32.09	29.8
	Income Distribution				
	Q1 Income	0.1	0.08	0.22	0.2
	Q2 Income	0.12	0.1	0.2	0.21
	Q3 Income	0.26	0.22	0.31	0.33
	Q4 Income	0.52	0.6	0.27	0.26
	Ability Distribution				
	0-30% Ability	0.0	0.0	0.01	0.0
	30-60% Ability	0.0	0.0	0.41	0.27
	60-90% Ability	0.15	0.08	0.4	0.53
	Top 10% Ability	0.84	0.92	0.17	0.2

Table F.7: Effect on colleges of making signals perfectly informative

F.4 Policy Experiments: Increasing Pell Grants

		College 1		College 2	
		Baseline	*Pell Grant*	*Baseline*	*Pell Grant*
L_μ	Avg Student Ability	1.86	1.87	1.28	1.26
I_μ	Instr Spending per Student	2.1	2.67	0.98	1.04
Γ_s	Value-added	2.01	2.09	1.79	1.77
κ	% Enrolled	7.15	7.25	32.09	32.14
τ	Tax rate (%)	1.51	3.93	1.51	3.93
	Income Distribution				
	Q1 Income	0.1	0.14	0.22	0.23
	Q2 Income	0.12	0.16	0.2	0.22
	Q3 Income	0.26	0.31	0.31	0.31
	Q4 Income	0.52	0.38	0.27	0.24
	Ability Distribution				
	0-30% Ability	0.0	0.0	0.01	0.06
	30-60% Ability	0.0	0.0	0.41	0.38
	60-90% Ability	0.15	0.14	0.4	0.39
	Top 10% Ability	0.84	0.86	0.17	0.17

Table F.8: Effect on colleges of increasing the Pell Grant maximum.

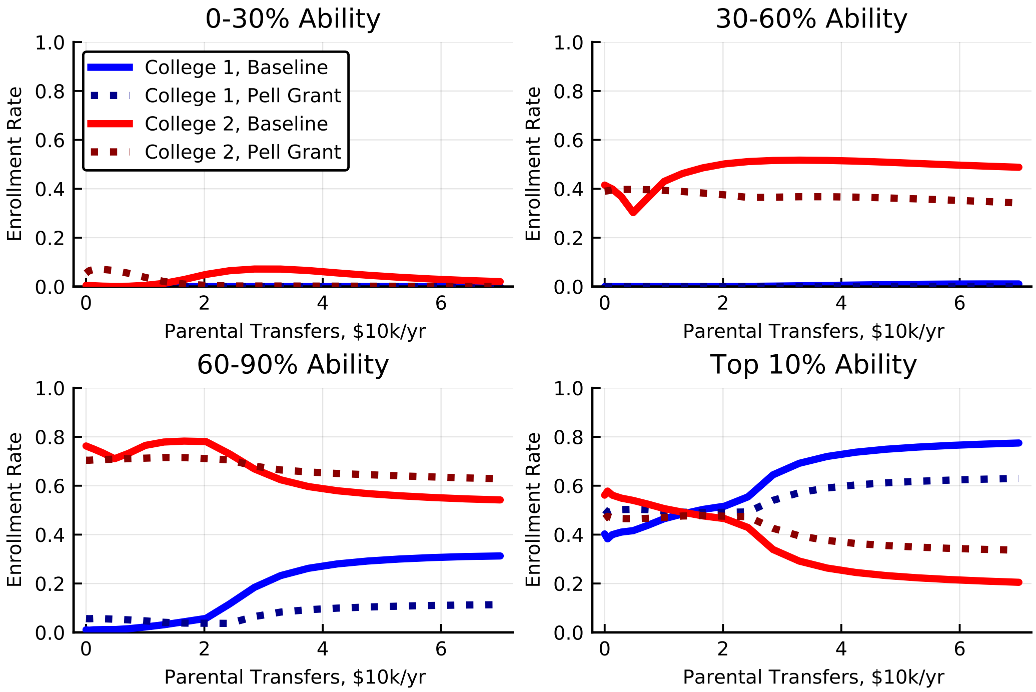


Figure F.9: Student sorting by EFC and ability in baseline equilibrium and equilibrium where Pell Grants have doubled.

G Computational Appendix

This section briefly describes the general procedure for how the equilibrium of the model is solved numerically. We start with a guess for College policies $\{\kappa_s, I_s, L_s, \Gamma_s, T_s(a,\sigma), \underline{\sigma_s} = 0\}_{s\in\{1,2\}}$ student application probabilities $\{p_i(a,\ell)\}_{i\in\{AB,A1,A2\}}$, and tax rate τ. We then update all choices so that they are consistent with optimal agent behavior, and iterate until our guess is consistent with itself. Our procedure is detailed in the following steps:

1. Using guesses $\{p_i(a,\ell)\}_{i\in\{AB,A1,A2\}}$, τ, Γ_s, solve for the total enrollment probability functions at each College, $\tilde{q}^s(a,\ell,T^1,T^2)$, $s\in\{1,2\}$. We find these functions using (14) and (8).

2. Given guess $\{\kappa_s, I_s, L_s, T_s(a,\sigma)\}$, $\tilde{q}^s(a,\ell,T^1,T^2)$, for $s\in\{1,2\}$, update the tuition and admissions standards for each college (we iterate until we find a fixed-point in the college policies).

   (a) For each College s, use guess for own aggregates $\{\kappa_s, I_s, L_s\}$, and guess for other Colleges tuition policy, $T_{-s}(a,\sigma)$ to find updated tuition policy $\hat{T}_s(a,\sigma)$ using (20). To handle the tuition cap, set $\hat{T}_s(a,\sigma) = \bar{T}_s$ if $\hat{T}_s(a,\sigma) > \bar{T}_s$.

   (b) For each College s, use updated tuition policy $\hat{T}_s(a,\sigma)$ to solve for the updated admissions standard $\underline{\hat{\sigma}}_s$ in (22)

   (c) Check for convergence

   - If $\sup|T_s(a,\sigma) - \hat{T}_s(a,\sigma)| > 10^{-5}$, set $T_s(a,\sigma) = \hat{T}_s(a,\sigma)$ and repeat (a)

- Otherwise, continue

3. Using updated tuition and admissions, find updated College aggregates $\{\hat{\kappa}_s, \hat{I}_s, \hat{L}_s, \hat{\Gamma}_s\}$ using (15), (16), (18), (13)
4. Using updated guesses, find updated tax rate $\hat{\tau}$ using (23)
5. Using all updated values and functions, find updated application decisions $\{\hat{p}_i(a, \ell)\}_{i \in \{AB, A1, A2\}}$ as in (25)
6. Check for convergence. Let $\hat{X} \equiv \left(\{\hat{\kappa}_s, \hat{I}_s, \hat{L}_s, \hat{\Gamma}_s\}_{s \in \{1,2\}}, \hat{\tau}\right)$, and $X \equiv \left(\{\kappa_s, I_s, L_s, \Gamma_s\}_{s \in \{1,2\}}, \tau\right)$
   - If $\sup |X - \hat{X}| > 10^{-5}$, set $X = \hat{X}$, $\{p_i(a, \ell)\}_{i \in \{AB, A1, A2\}} = \{\hat{p}_i(a, \ell)\}_{i \in \{AB, A1, A2\}}$, and repeat (1)
   - Otherwise, exit