Need-Based Aid from Selective Universities and High School Achievement

Sunha Myong*†

October 19, 2015

Abstract

Over the last decade, the amount of financial aid from selective universities increased by more than 60%, most of which was awarded based on students' financial need. This paper argues that not only increasing need-based aid can promote diversity, but also it can improve the academic preparation level of attendees at selective universities from *all* income backgrounds. If the tournament effect–the additional effort applicants put in to increase the probability of admission–is more elastic to student ability than to family income, elevated competition could raise the effort level of all high ability high school students. To quantify this effect, I develop a structural model of students' learning, application, and admission processes, and estimate it with the Education Longitudinal Study of 2002, a nationally representative sample. If selective universities provide $10,000 more in grants per attending student from the bottom quintile of the income distribution, the average effort level of attendees, measured by the number of Advanced Placement (AP) classes, increases by 1.3% for low-income students and by 0.5% for other income students. However, selective universities might face an efficiency-equality tradeoff because extending need-based aid to middle-income students, while raising tuition to keep the revenue, can further increase academic qualification of attendees by 0.4% but decrease the number of low-income students by 1.3%.

*Economics, Washington University in St. Louis; email:sunhamyong@wustl.edu

†I am deeply grateful to B. Ravikumar, Limor Golan, and Carl Sanders for their guidance and support. I also thank to Goerge-Levi Gayle, Juan Pantano, Young-Seok Shin, Jungho Lee, JungJae Park, Roberto Samaniego, and Yu Zheng for valuable discussions and suggestions.

1 Introduction

Over the last decade, the amount of financial aid per student offered by selective universities increased by more than 60%. Most of aid from selective universities was awarded according to students' financial need.[1] Promoting economic diversity and improving upward social mobility are claimed as primary goals. However, the academic preparation level of a high school graduate is not exogenous but depends on student's effort choice at the high school level, for example, the number of Advanced Placement (AP) classes. Also, the marginal return and the marginal cost of additional effort choice could differ by student ability and family income. Therefore, the extent of progressivity in the redistributive financial aid policies might not only have an implication on the inequality in the extensive margin—diversity, but also have an implication on the efficiency in the intensive margin—the average academic preparation level of attendees in selective universities.

For example, suppose that selective universities provide \$10,000 more in grants to attending students from the bottom quintile of income distribution, while raising tuition to keep the revenue. Some high ability low-income students, who otherwise would not apply for selective universities, might change their application decisions, take more AP classes, and achieve higher test scores in preparation for the competitive admission process. However, in response, some middle-income students might be discouraged from applying to selective universities, take less AP classes, and attend nonselective universities that offer generous merit-based aid. This, however, could be a loss from selective universities' point of views, if the students from middle-income families at the margin of application decision have higher ability than those from the bottom income quintile. In this case, spending \$10,000 on the middle-income students could have been more effective per dollar to increase the average academic preparation level of attendees. Therefore, selective universities might face an efficiency-equality tradeoff when they craft their financial aid packages.

In this paper, I ask to what extent redistributive aid policies from selective universities could increase economic diversity without lowering the academic quality of attendees. In doing so, I first consider a counterfactual policy that provides \$10,000 more grants from selective universities to students from the bottom quintile of income distribution. I evaluate its impact on the effort choice of students, measured by the number of AP classes, from different income backgrounds. Then I consider the following two counterfactual analyses to examine an equality-efficiency tradeoff associated with different forms of need-based aid policies of selective universities: (i) increasing the amount of grants awarded to low-income students from \$10,000 to \$35,000 gradually, and (ii) including middle-income students as the recipients of \$10,000 more grants. In both exercises, I raise the tuition to keep the total tuition revenue.

[1]Selective universities refer to four-year colleges in the United States that belong to top two categories of the NCES-Barron's Admissions Competitiveness Index among seven categories. Such universities account for about 20% of the entire four-year college enrollment in the United States. The data on the average amount of financial aid come from the Integrated Postsecondary Education System (IPEDS).

I argue that not only can increasing need-based aid from selective universities promote diversity, but also it can improve the academic preparation level of attendees at selective universities from *all* income backgrounds. Because the tournament effect—the additional effort choice made by applicants to increase the probability of admission—is more elastic to student ability than to family income, elevated competition could raise the effort level of all high ability high school students. To elaborate this point, it is useful to think about the asymmetric impact of the rising admission cutoff score on the effort choice of high and low ability students. Increased competition counteracts the diminishing marginal returns of the effort choice of high ability students. However, taking more AP classes is too costly for low ability students because the chance of getting admission is too small. Importantly, this paper claims that under the current learning technology and the admission criteria, less able high-income students cannot reverse this effect by paying for private high school.

Also, I find that the average test scores of students in selective universities increases until the amount of need-based aid toward the bottom quintile income group becomes more than double the current aid, even though I raise tuition to keep the tuition revenue. However, extending need-based aid to middle-income classes is more effective per dollar to raise the academic preparation level of attendees in selective universities, although it slightly decreases the number of low-income students in selective universities. This shows a potential tradeoff between diversity and academic quality of attendees in selective universities.

To quantify students' behavioral responses, I develop a structural model of the learning, application, and admission process. I estimate the model based on the three data sets: (i) the Education Longitudinal Study of Youth of 2002, (ii) the Integrated Postsecondary Education System Data (IPEDS 2004), and (iii) the NCES-Barron's Admissions Competitiveness Index. Then I discuss quantitative results of counterfactual policies. To validate the implication of the model, I conduct an out-of-sample prediction based on the National Education Longitudinal System of 1988.

In the model, attending selective universities provides pecuniary and nonpecuniary benefits. However, its direct costs are much greater than those of attending nonselective universities. A highly competitive admissions process exists for the limited seats in selective universities, and that process focuses mainly on students' academic achievements. Students decide (i) whether to attend private high school, (ii) how many Advanced Placement (AP) classes to take during high school, and (iii) whether to apply for selective universities. Tuition, the admission result, and the amount of financial aid are determined exogenously, and then the student takes a loan to finance the cost of college education. The liquidity constraint could limit the student's borrowing capacity as a fraction of her future income.

The main data set is the ELS2002, which provides a wide range of students' demographic characteristics, high school curriculum choices and test scores, the college application and admission results, and wage rates. I use the IPEDS to get tuition information and use the NCES-Barron's Admissions Competitiveness Index to define college selectivity. Four particular variables are useful to identify the model. First, the geographic distributions of tuition of home-state nonselective universities and the cost for the room and book are important exogenous variations to identify students' responses to financial incentives. Second, the total amount of student loans is used to identify the intertemporal preference over the consumption between college and work period. Then the different choices of student loans between rich and poor students help to identify the extent of the liquidity constraint. Third, the number of information sources about the college application process helps to control potential information barrier facing poor students. Fourth, the number of AP classes offered by high schools is used to account for different learning environments across students.

The estimated model provides three sets of quantitative results. First, if selective universities provide $10,000 more in grants per attending student from the bottom quintile of income distribution, the average effort level of low-income attendees increases by 1.3%, whereas that of other income groups increases by 0.5%. Not surprisingly, increasing need-based aid from selective universities can also promote diversity. The number of applicants from the bottom quintile of income distribution increases by 12% among which 53% attend selective universities. Reducing the liquidity constraint explains 51% of the increase in low-income students' attendance rates, and the remaining 49% is explained by the income effect[2]. Second, I find that until the aid to the lowest income quintile students increases by $30,000, the average academic quality from all income backgrounds increases, whereas the proportion of students from the bottom quintile income distribution in selective universities increases by 2.2%. Third, however, providing $10,000 additional grants to both low- and middle-income attendees further increases the average academic quality of attendees in selective universities by 0.4% but it decreases the number of low-income attendees by 1.3%.

Need-based aid from selective universities also affects the effort level of all students. If the amount of grants toward the students from the bottom quintile of income distribution decrease by $10,000, the overall effort level of low-income high school students also increases by 18%, and the achievement gap between rich and poor students increases. The aggregate effort level of all students increases by 0.8%.

To examine the validity of the structural model, I consider an out-of-sample prediction by using the National Education Longitudinal Study of 1988, a representative sample with almost identical survey instruments to the ELS2002. Assuming that changes in the financial aid policy from selective universities are exogenous, I compare the observed pattern in the NELS1988 data with the predicted outcomes in the estimated model. When it comes to the composition rate of applicants and attendees in selective universities conditional on family income quintile, the difference between the model and the NELS1988 data is less than 3%.

The paper is organized as follows. I discuss related literature in Section 2. The data and

[2]Because pecuniary benefit is too small to compensate for expensive tuition, only those who already have a lot to consume want to attend selective universities and obtain nonpecuniary benefit

motivating facts are described in Section 3. I explain the model and choices of high school students in Section 4. Section 5 discusses the identification of the model. Section 6 describe the estimation method. Section 7 documents the estimation result. Section 8 presents the counterfactual policy analysis. Section 9 shows an out-of-sample prediction. Section 10 concludes.

2 Related Literature

There have been increasing studies on the college application, admission, and matriculation processes and on the equilibrium assignment between students and colleges. Many studies are motivated by actual policies, such as race-based affirmative action (Arcidiacono, 2005) and [3], Texas Top 10 Law (Kapor, 2014)[4]. Others focus on the strategic behaviors of students and colleges. For example, Fu (2014) estimates the equilibrium matching model between students and colleges, focusing on the processes by which colleges set tuition and admission rules and students make application and enrollment decisions. Fillmore (2014) shows how colleges capture a large share of matching surplus through price discrimination, because they can collect student information from the Free Application Federal Student Aid (FAFSA). In those studies, students' academic achievement levels at the moment of college application are taken as given. In contrast, the main focus of this paper is to show how test scores can change as students adjust their effort choices at the high school level.

Hickman (2013) also studies how changes in college admissions policies can affect high school students' effort levels. This paper differs from Hickman (2013) in focusing on need-based aid from selective universities rather than on race-based affirmative action. In doing so, it focuses on the respective roles of family income and ability, and examines the impact of the liquidity constraint on application decisions (extensive margin) and effort choices (intensive margin). Because the ability cutoff of a student at the margin of the application decision for each income group is determined in equilibrium, it is difficult to have a monotone relationship between the student's initial characteristics and the final test scores that is invariant to policy changes. While the strong monotonicity between a student's test scores and the unobservable effort cost is the key identification strategy in Hickman (2014), this paper uses the rich data to identify the model. For example, the number of AP classes offered by high schools can be used to separately identify the cost of effort choice from the benefit. Because the financial aid only affects the benefit of additional effort, controlling the cost side is useful to understand the impact of tuition subsidy on student's effort choices.

This paper also relates to the literature studying the impact of the liquidity constraint on students' educational choice. Becker and Tomes (1994) shows how liquidity constraints can explain

[3]For example, the use of race-based affirmative action in college admission process has been banned in several states such as California, Washington, Michigan, Nebraska, Arizona, Oklahoma after the 1990s

[4]Since 1997, every high school student in Texas who ranks in the top 10% of her class is a guaranteed admission at state-funded universities

the different educational investment between rich and poor students, and how government tuition subsidy can affect the distribution of human capital through the endogenous response of families. As an extension, this paper studies the impact of increasing need-based aid from selective universities, accounting for students' monetary and effort investment during high school and competitive admissions processes for the limited seats in selective universities. Heckman and Cameron (1998) and Lochner and Monge (2002) evaluate the importance of the liquidity constraint on college enrollment decisions. Different from previous studies, this paper studies the differences between selective and nonselective colleges.

3 Data

3.1 Sample Description

I used the Education Longitudinal Study of 2002 (ELS2002), the Integrated Postsecondary Education Data System (IPEDS), and NCES-Barron's Admissions Competitiveness Index 2004 to estimate the model. The ELS2002 is a nationally representative sample of the United States high school students, and it followed the sample up to eight years after high school graduation. The ELS2002 was initially surveyed in 2002 with 10th graders, and there were three follow-ups in 2004, 2006, and 2012. Publicly available data includes comprehensive information such as students' academic achievements as measured by standardized test scores, school characteristics, family background, college attendance/graduation, and the wage rate. However, the information on which university the student applied to, admitted to, or attended is not available in the public data. For this reason, I used a restricted data set. Restricted data of the ELS2002 includes a complete history of college education: all institutions the student ever applied to, admitted to, and attended are listed. For students who attended more than one college, I focus on the first college the student attended.

The original sample size of the ELS2002 is about 35,000 of 10th graders in 2002. This paper only focuses on four-year college attendees. I dropped data with missing information such as initial math score, parents' educational attainment, family income, high school type, SAT, GPA, and AP test scores, or the history of college application, admission and attendance, and the wage rate. The final sample size is 3,080.[5].

The transcript data in the ELS2002 allows me to observe the entire history of a student's high school curriculum choices. For simplicity, I track the total number of AP classes taken by students regardless of grade or subject. Also, I can observe the number of AP classes offered by the high school, which helps to account for the different learning environments across high schools. For high school achievement, I use three scores: SAT, GPA, and the AP score.[6] Observing students'

[5]The wage rate is observed in the 3rd follow up, so that attrition rate is quite high.

[6]There are more than twenty AP tests, each of which has a scale of 1-5, but I use the average AP test score using the equal weight for each subject. I allow that students can take AP test without attending AP classes.

initial academic achievement as measured by the standardized math score in 10th grade[7] is useful to control different achievement level at the beginning of high school period. Having early academic achievement allows me to examine the importance of the endogenous learning process.

The total student loan amount reported in 2013 (eight years after high school graduation) is used to infer annual loan amount. The ELS2002 has information on the fraction of total grants for tuition students receive at the first-attended postsecondary institution. I use the cost of tuition reported by colleges to the Integrated Postsecondary Education System (IPEDS) to infer the amount of grants each student receives. I include the number of information sources available to students about the college application process. This helps to control information disparity between rich and poor, which has received increasing public and policy interests.[8]. The ELS2002 also has data on students' preference for certain features of college education reported in 10th grade. For example, students ranked the importance of college's reputation on a scale 1 to 3. I include student's preference for college's reputation, location, and whether it is the same attended by one of her parents to account for different application behaviors driven purely by differences in preference.

The IPDES has two important information. First, I use the posted tuition of each university along with the cost of room, board and books to account for the direct costs of college education. Also, I assume that the tuition of nonselective universities is the average tuition of nonselective universities in students' home states. I also account for home state discounts as reported by each institution in the IPEDS (Figure 15, 16, 17).[9] Second, I include the number of selective universities in students' home states. Including this exogenous geographic variation, presumably independent of student ability, could be useful to isolate preference from ability when explaining different application behaviors across students.

I use NCES-Barron's Admissions Competitiveness Index Files 2004, to define college selectivity. It covers all four-year colleges across the United States and provides an admission-competitive index for each university based on specific criteria. There are seven categories of selectivity level in NCES-Barron's index, but in this paper, selective universities are defined as the universities with top two NCES-Barron's Index in 2004. This leads to 103 selective universities out of 1,247 post secondary institutions in 2004, and 20% of all four-year college enrollees in my sample attend selective universities.

[7]This is the earliest achievement test score available in the sample. I check how this score could change between 8th and 10th grade from the comparable data set, the National Education Longitudinal Study of 1992. Those two samples share almost the same survey instruments. I found that the correlation between the standardized math score in 8th grade and in 10th grade is 0.89 and the t-test shows that the growth rate does not differ by whether student attends private high schools or not.

[8]Low application rate for selective universities by high-achieving low-income students has motivated a policy interruption to reduce information barriers facing low-income students such as sending out a packet about college application process (Hoxby et al. ,2013)

[9]For the tuition of private high school, I use the aggregate data in 2002 from National Center for Educational Statistics (NCES).

3.2 Descriptive Analysis

Need-Based Aid from Selective Universities

Table 1 documents tuition, room, and board at different types of colleges. Tuition of selective universities is 65% more expensive than that of nonselective universities for students from out of state. If I include tuition (before subtracting financial aid) and the cost of room, board and books, the average direct cost of attending selective universities was $35,240 dollars per year, in 2004, as compared to $22,240 dollars per year for nonselective colleges. [10]

Figure 6 and Table 6 show the composition of students in selective universities from each quintile of income distribution conditional on student's family income. Students from the highest income quintile families comprise 42.6% of attendees in selective universities, whereas only 8.4% of students come from the bottom quintile of income distribution. Figure 9 shows the attendance rate (attendees/all), application rate (applicants/all), and admission probability (attendees/applicants) of students by income quintile. Figure 10(a) shows the measures of students at all achievement levels, whereas Figure 10(b) shows the measures for high-achieving students who have top 20 math scores at the 10th grade. Family income is positively correlated with all three rates if I do not control for initial ability. For high-achieving students, it shows an inverted U-shape: students from the bottom quintile of income distribution have equally high attendance, application, admission rates as the highest income quintile students. The admission probability upon application is actually higher for students from the lowest income quintile distribution than those from the highest income quintile distribution.

Financial Aid from Selective and Nonselective Universities

Figure 4 shows the average amount of institutional grant per student at selective and nonselective universities in 2012. Students with a family income of $30,000 received more than twice as much aid as students with a family income of $110,000 at selective universities. At nonselective universities, students from the richest families receive more aid from nonselective universities because nonselective universities also have merit-based aid. Table 13 shows the regression result of the amount of grants students received in ELS2002 sample. The amount of aid decreases faster by family income increases in selective universities. Thus, although the sticker price of attending selective universities is much more expensive, attending selective universities can be less expensive for low-income students than high-income students.

Figure 2 shows the trend in institutional grants from selective and nonselective universities from

[10] A more drastic difference is found if I compare tuition of colleges by finer degree of selectivity as defined by the NCES-Barron's Admissions Competitiveness Index Files 2004. Table 1 shows that among seven selectivity categories, the most selective universities cost almost twice as much as colleges with a median level of selectivity, and almost three times as expensive as the least selective universities.

2000 to 2012 (the IPEDS). The institutional grants offered by selective universities increased by more than 60% over the last decade, whereas there is only a mild increase in institutional grants offered by nonselective universities. Figure 3 shows the trend in the average Pell Grant amounts per student at selective and nonselective universities over the same period. There is no significant difference between selective and nonselective universities, and the average amount of a Pell Grant award increased less than $1,000 over the ten years per student. Table 4 shows the no-loan policy introduced by top-ranked universities. Table 2 documents four different types of no-loan policies. In most cases, many prestigious universities replaced loans with grants in financial aid packages during the mid-2000s. As a result, students from families with less than $60,000 annual income now get a full tuition discount if they can attend one of those wealthiest universities.

Diversity vs. Academic Qualification

Figure 5 shows the proportion of Pell Grant recipient in selective universities after 2008.[11] The proportion of Pell Grant recipients in selective universities is a widely used metric, because 97% of Pell Grant recipients come from families with lower than $50,000 family income. Thus, this table suggests that economic diversity in selective universities increased after the mid-2000s.

Figures 10 and 11 document the average SAT scores of attendees in selective and nonselective universities during a similar period. Math, verbal and writing SAT scores increase in selective universities, whereas they do not change very much in nonselective universities.

Advanced Placement Classes

Figure 14 describes the number of AP classes taken by students conditional on their family's income quintile. It also shows how applicants/attendees take different number of AP classes from others. The average number of AP classes students take is 0.6 credit units higher for students from the highest income distribution than those from the bottom income distribution quintile. However, if I look at those who attend selective universities, students from the lowest income quintile take more AP classes than those from the highest income quintile. On average, applicants and attendees of selective universities take 2.7 and 3.5 AP classes respectively, whereas non-applicants and non-attendees take 0.9 and 1.2 AP classes.

Figure 13 describes the number of AP classes offered by high school of students who do not take any AP class. Of those nonparticipants, more than 25% of he students from the bottom quintile of income distribution do not have available AP classes in high school, whereas the corresponding number for the highest income quintile group is 15%.[12]

[11] 2007-2008 academic year is when many top-ranked universities introduced no-loan policy. Also, the proportion of Pell Grant recipients is not available before 2008 data from the IPEDS.

[12] This motivated the government to increase policy interventions that aim to increase AP participation by low-income students. For example, during 2002 $\sim$ 2011, the Federal government granted $22 million in Advanced Place-

Table 14 shows the regression result of test scores on the number of AP classes students take, controlling for other observable characteristics such as race, sex, initial math score, family income, and the unit of math credits. It shows that students who take more AP classes obtain significantly higher test scores. Tables 12 and 15 document the regression result of students' college dropout rate and the wage rate based on the number of AP classes and other observable characteristics. Again, AP classes have a significant positive correlation with students' higher graduation rate and wage return.

High School Achievement and Wage Rate

Table 15 shows the original linear regression of the wage rate on high school achievement, family backgrounds, and curriculum choice. Column (3) indicates that log SAT (0.242) and log GPA (0.207) scores significantly increase the log wage rate, after controlling race, gender, high school curriculum choice, initial math score, family income, parents' education, and two personality traits: motivation and action control. Motivation variable measures student's attitude toward the financial success in life. Table 16 shows that log SAT (0.166) and log GPA (0.125) remain significant after controlling college major, dropout, and further education. These results suggest that high school achievements are significantly determining factor of early period wage rate.

Table 17 shows the wage regression based on the National Longitudinal Study of Youth 1979. It is a pooling estimation of the panel data that include the hourly wage over more than 20 years after high school graduation. Column (2) and (3) suggest that including SAT instead of AFQT (The Armed Forces Qualification Test) increases the R-square by 0.03. Column (4) suggests that if I include both SAT and AFQT score and interactions with the age, log SAT score significantly increases the level of the wage rate (0.63), while its impact on the growth rate (SAT$\times$AGE) is not significant. This suggests that school achievement affects not only the early period wage rate, but also affects the life time earning even after controlling ability measure AFQT.

4 The Model

I consider an individual who is about to enter a high school, and live two periods. In the first period, she engages in the high school and college education, and in the second period she becomes a full time worker. I assume that everyone goes to the four-year college after high school graduation, and there are two types of colleges: selective and nonselective. Attending selective universities provides pecuniary and nonpecuniary benefits, but there are direct and indirect (effort) costs for preparing for and attending selective universities. To maximize the lifetime utility, the student decides how much effort (AP class) and money (whether to attend private high school) to invest in during the

ment Incentive Program grants to 20$\sim$30 high schools in certain districts (Advanced Placement Incentive Program Grants).

high school, and decides whether to apply for selective universities. In what follows, first, I describe the general description of the structural model, then I proceed to explain the empirical specification of each component. Finally, I characterize choices of students and discuss the implication.

Let $S = \{A, \theta, M_1\}$ be the vector of state variables that summarizes individual's initial characteristics at the moment of the high school entrance. A is the initial math score, θ is the unobservable ability, and M_1 is family income. Individual's utility consists of four components: the utility from consumption (U_c), the utility from nonpecuniary benefit from attending selective universities (U_{sel}), the utility cost of application (U_{apply}), and the utility cost of taking AP classes (U_{AP}). To maximize the lifetime utility, individuals choose the following choice variables $X = \{I_{private}, N_{AP}, I_{apply}, L, C_1, C_2\}$, where $I_{private}$ indicates whether to attend private high school or not, N_{AP} is the number of AP classes to take during high school, I_{apply} is whether to apply for selective universities or not, L is the amount of student loan, and C_1, C_2 are consumptions during college and working periods. Throughout the process, there are four outcome variables that affect student's labor income and utility. Let $Y = \{H, I_{sel}, I_{BA}, M_2\}$ be the vector of outcome variables, where H is the test score, I_{sel} is whether to attend selective universities or not, I_{BA} is whether to finish college education or drop out, and M_2 is labor income.

The test score H is determined following a learning technology, $H = F_H(x; A, \theta)$ where $x = \{I_{private}, N_{AP}\}$ are two inputs that affect the final test score. Note that the learning efficiency is affected by initial math score A and unobservable ability θ. Attending a private high school costs tuition (t). There are exogenous stochastic processes that determine (i) the admission probability from selective universities (P_{ad}), (ii) the success rate of college education (P_{BA}), and (iii) tuition and the amount of financial aid offered by selective and nonselective universities. I assume that once the test score H is realized, the initial math score A becomes irrelevant. Let $P_{ad}(\theta, H)$ be the probability of getting an admission from selective universities upon application. Let $g(\theta, H)$ be the admission criteria and let h^* be the cutoff value that equalizes the number of seats available in selective universities and the number of admitted students. Then $P_{ad}(\theta, H) = P(I_{sel} = 1 | I_{apply} = 1) = P(g(\theta, H) > h^*)$. Let P_{BA} be the probability of completing the college education, and it follows a stochastic process, $P_{BA} = F_{BA}(\theta, H, I_{sel})$. Let T_{sel} and T_{non} be the realized tuition of selective and nonselective universities respectively. Let G_{sel} and G_{non} be aid offered by selective and nonselective universities respectively. G_{sel} is a function of only family income M_1 (completely need-based), while G_{non} is affected by both family income and the test score (M_1, H) (both need and merit-based). Therefore, the direct cost of college education is determined by $T^s = T_{sel} - G_{sel}$ and $T^n = T_{non} - G_{non}$. Let $\overline{T} = I_{sel}T^s - (1 - I_{sel})T^n$ be the realized net tuition.

At the moment of college entrance, the student knows $(I_{sel}, P_{BA}, \overline{T})$. She has a correct information about the expected labor market earning $M_2 = F_M(\theta, H, I_{sel}, I_{BA})$. She has to finance $\overline{T}$, and she can take a student loan L from the capital market with a fixed interest rate r and the repayment plan. There is no difference between government loan and private loan. However, there is a friction in the capital market that reduces the maximum amount of loans available compared to the complete market case, thus the student potentially faces a liquidity constraint. In particular, I assume that the student can borrow only up to a certain fraction of her future expected income, $L \leq \lambda E[M_2]$.

Note that there are direct and indirect channels through which family income affect choices of students. First, the direct channel is through financing the high school (if she attends private high school) and college education. Second, the indirect channel is through the unobservable ability θ. In particular, I allow that the distribution of θ to be correlated with the family income, and θ is included in each process. Thus, besides the financial channel, all other correlations between family income, choices of students, and economic outcomes are attributed to θ. Based on this description, student's problem at the moment of high school entrance can be written as

$$\max_X E_{S'}[U_c + U_{sel}] - U_{AP} - U_{apply} \tag{4.1}$$

subject to

$$C_1(S') + T(S') \leq M_1 + L(S') - t \cdot I_{private}$$
$$C_2(S') \leq M_2(S') - (1+r)L(S')$$
$$L(S') \leq \lambda M_2(S')$$

,where S' is the set of all possible future states associated with the realization of $Y = (H, I_{sel}, I_{BA}, M_2)$. Note that the model does not include (i) student's portfolio choice within the same college selectivity and (ii) problems from the perspective of colleges regarding admission decision, financial aid and tuition level. Instead, I focus on whether the student ever applies for any of selective universities or not. Also, the admission probability and the amount of financial aids are assumed to follow exogenous stochastic processes, and students are assumed to be fully aware of the correct distribution when they infer their expected outcome. Now I describe the empirical specification of each component of structural model.

4.1 Preference

I assume a log utility from consumption, and β is the time preference between college and working period. Denoting the consumption during the college and working period as C_1, C_2, the utility from consumption can be written as

$$U_c(C_1, C_2) = \ln(C_1) + \beta \ln(C_2)$$

Next, I assume that the nonpecuniary benefit of attending selective universities (U_{sel}) depends on student's initial math score (A), unobservable ability (θ), the number of selective universities in student's home state ($N_{SelHome}$), and student's preference for certain features of college education such as location, reputation, and whether they want to attend the same college as her parents attended (RP_m). Ability measures in the nonpecuniary benefit are supposed to capture academic orientation or the capacity of handling peer pressures after attending selective universities. The number of selective universities in student's home state is supposed to capture the potential influence of growing up knowing more about elite universities and prestigious campus life. On the other hand, the preference regarding certain feature of college's characteristics is included to capture pure preference heterogeneity[13].

$$U_{sel} = \nu_{10} + \nu_{11}\ln(A) + \nu_{12}\theta + \nu_{13}q + \nu_{14}N_{SelHome} + \sum_{m=1}^{4}\nu_{RP_m}RP_m$$

The effort cost of taking AP classes is affected by student's ability and the number of AP classes offered by the high school $N_{offered}$. ϵ_e is the i.i.d. shock in the marginal effort cost.

$$U_{AP}(N_{AP}) = \left[\xi_0 + \xi_1\ln(A) + \xi_2\theta + \xi_3 ln(N_{offered}) + \epsilon_e\right]\cdot N_{AP}$$

Finally, the effort cost of application also depends on student's ability, high school type (private of public), the number of information sources available regarding college application process, N_{inf}.

$$U_{apply} = \psi_0 + \psi_1\theta + \psi_2\cdot I_{public} + \psi_3 N_{inf}$$

4.2 Unobservable Ability Heterogeneity

The unobservable ability θ is allowed to affect each part of structural component: the wage rate, learning efficiency, nonpecuniary benefit from attending selective universities, application cost, and the admission probability. It is known to individuals and colleges, but is not observed by econometrician. The aggregate distribution of θ is assumed to have a normal distribution, in which the mean depends on student's family income (M_1) and parent's education (Edu_p).

$$\theta = \psi_1\ln(M_1) + \psi_2\ln(Edu_p) + \epsilon_\theta$$

, where $\epsilon_\theta \sim N(0, \sigma_\theta^2)$

[13] One interesting find in Fu (2014) is that students' preferences over different types of colleges vary substantially even after controlling SAT and family income.

4.3 Learning Technology

I assume that $H = (H_{SAT}, H_{AP}, H_{GPA})$: there are three test scores - SAT, AP, and GPA scores, relevant in the admission process, the wage rate, and graduation rate from the college. The test scores depend on how many AP classes to take and whether the student attends private high school. Also, the learning efficiency is affected by student's characteristics $Z = \{A, \theta, sex, race\}$. Each test score has a unique learning technology. This could capture the difference between standardized test scores, SAT and AP, and local score GPA which is affected by achievement level of peers. In particular, if the average ability level of peers in private high school is higher than that of public high school, students who attend private high school may obtain lower GPA. For $j = \{1, 2, 3\}$,

$$lnH = \sum_{m=1}^{4} \beta_{mj} Z_m + \beta_{5j} \ln(N_{AP} + 1) + \beta_{6j} I_{private} + \epsilon_{H_j}$$

, where ϵ_{H_j} is i.i.d. random shock realized at the end of the high school period.

4.4 Admission Probability

I assume that the admission criteria is exogenously given and students are fully aware of the rule, that is , there is no information friction. Selective universities rank applicants based on a measure for student's merit, $g(D)$, which is a function of $D = \{race, H, \theta\}$. Then the cutoff value of admission is determined to equalize the number of seats in selective universities and the number of admitted students[14].

$$g(D) = \sum_{j=1}^{J} \beta_j I_{race,j} + \beta_{t1} \ln(SAT) + \beta_{t2} \ln(AP + 1) + \beta_{t3} \ln(GPA) + \theta_4 \ln(N_{AP} + 1) + \beta_{t5}\theta + \epsilon_p$$

Let $P_{ad}(D)$ be the admission probability and let Q_{sel} be the number of seats available in selective universities. Then $P_{ad}(D) = P(g(D) > h^*) = \Phi(g(D) - h^*)$, where h^* satisfies

$$\sum_{i=1}^{N_{applied}} I(g(Z) - h^* > 0) = Q_{sel}$$

4.5 Financial Aid

Let G_k be the grant from college with selectivity $k \in \{sel, non\}$. Assume that the amount of financial aid from selective universities are completely need-based, and it only depends on the income

[14] I assume that there is no asymmetric information regarding θ between students and admission committees. Or it can be considered that (i) colleges have a perfect screening device $\hat{\theta}$ to infer θ such that $\theta = f_\theta(\hat{\theta})$ where f_θ is a monotone function, and (ii) $g(H, \theta) = g(H, f_\theta^{-1}\theta)$ is known to everyone. Thus, the student can infer her admission probability correctly

quartile of student's family. On the other hand, the financial aid from nonselective universities has both need-based and merit-based components. In particular, the amount of aid depends on both income quartile and academic achievement (SAT, AP, and GPA). Let DF_k be the dummy variable indicating the student's family income belongs to $k-th$ highest quintile, and let ϵ_{ζ_s} and ϵ_{ζ_n} are i.i.d. random shocks. Thus the financial aid from each type of college is specified as follows.

$$G_{sel} = \zeta_{s0} + \sum_{k=1}^{4} \zeta_{sk} DF_k + \epsilon_{\zeta_s}$$

$$G_{non} = \zeta_{n0} + \sum_{k=1}^{4} \zeta_{nk} DF_k + \zeta_{nt1} \ln(SAT) + \zeta_{nt2} \ln(AP+1) + \zeta_{nt3} \ln(GPA) + \epsilon_{\zeta_n}$$

4.6 Graduation Rate

Once entered the college, the student faces a stochastic process determining whether she can graduate from the college or drop out from it. The probability depends on the ability (θ), college preparatory test scores (SAT, AP, and high school GPA), and student's sex, and college selectivity I_{sel}. Let P_{BA} be the probability of obtaining BA degree and it can be written as follows [15]

$$\begin{aligned} P_{BA} &= P(s_0 + s_1 I_{sel} + s_2 female + s_3 ln\theta + s_4 \ln(SAT) + s_5 \ln(AP+1) + s_6 \ln(GPA) + \epsilon_{BA} > 0) \\ &= \Phi(s_0 + s_1 I_{sel} + s_2 female + s_3 ln\theta + s_4 \ln(SAT) + s_5 \ln(AP+1) + s_6 \ln(GPA) \end{aligned}$$

4.7 Wage Rate

I assume that every student becomes a full time worker in the second period. Let $I_{drop} = 1 - I_{BA}$ be the dummy variable of the college dropout. The wage rate is determined by student's high school achievement (H), ability (θ), college selectivity (I_{sel}), drop out (I_{drop}), and the demographic characteristics (sex,race). Since there is only one period as a worker and every worker is at the same age, I do not include the return from potential experience. The coefficient of I_{drop} is assumed to capture the loss of returns from a college degree net of the return from actual labor market experience.

$$\begin{aligned} lnW = &\Gamma_0 \cdot Z_w + \gamma_1 \ln(SAT) + \gamma_2 \ln(AP+1) + \gamma_3 \ln(GPA) + \gamma_4 I_{sel} + \gamma_5 I_{nonsel} I_{drop} + \gamma_6 I_{sel} I_{drop} \\ &+ \gamma_6 \theta + \epsilon_w \end{aligned}$$

[15] Alternatively, I can endogenize the graduation decision, while estimating the utility cost of graduation as a function of $(sex, I_{sel}, \theta, lnSAT, lnAP, lnGPA)$. Unless I assume that the preference shock regarding drop out decision is i.i.d., then the implication of the model would not change a lot

, where $Z_w = \{Black, Asian, Hispanic, Female\}$ are the demographic characteristics that affects the wage rate, and $\Gamma_0 = \{\Gamma_{Black}, \Gamma_{Asian}, \Gamma_{Hispanic}, \Gamma_{Female}\}$ are the corresponding parameters. Finally, ϵ_w is the i.i.d. random shock on the wage rate.

4.8 Characterization

4.8.1 Family Income and Application (Extensive Margin)

Suppose that everyone graduates from the college. Denote M_2^{s*} and M_2^{n*} to be the labor market income when graduating from selective and nonselective universities respectively. Suppose that the student who attends $j \in \{sel, non\}$ type of college is not liquidity constrained. Then the optimal choices for the consumption and student loan are $C_1^* = \frac{1}{1+\beta}\left[M_1 + \frac{1}{1+r}M_2^{j*} - T^j\right] = \frac{1}{\beta(1+r)}C_2^*$, and $L^* = \frac{1}{1+\beta}\left[\frac{M_2^{j*}}{1+r} - \beta(M_1 - T^j)\right]$. If the constraint is binding, then $C_1^{**} = [1-(1+r)\lambda]M_2^{j*}$, $C_2^{**} = M_1 + \lambda M_2^{j*} - T^j$, and $L^{**} = \lambda M_2^{j*}$. Let V^{apply} be the value of applying for selective universities, and let V^s and V^n be the value of attending selective and nonselective universities without including U_{apply}. Since the admission is competitive, $V^{apply} = P_{ad} \cdot V^s + (1 - P_{ad}) \cdot V^n$. Therefore, applying for selective universities is optimal iff $P_{ad}(V^s - V^n) > U_{apply}$. It is straightforward to show the following.

Proposition. *If liquidity constraint is not binding, it is optimal to apply for selective universities iff* $M_1 + \frac{1}{1+r}M_2^{s*} - T^s > exp(K)\left[M_1 + \frac{1}{1+r}M_2^{n*} - T^n\right]$, *where* $K = \frac{1}{1+\beta}\frac{U_{apply}}{P_{ad}} - U_{sel}$.

I interpret $exp(K)$ as a relative price of the consumption conditional on the college selectivity. For instance, $exp(K) < 1$ would imply that living at a dormitory in a prestigious campus values more than living at a dormitory in less prestigious universities. There are two reasons why $exp(K)$ could be larger for the low-income students and why it is more expensive for low-income students to attend selective universities in utility term. For expositional simplicity, I assume that there is no effort choice in the following discussion. First, $exp(K)$ increases if U_{apply} increases, P_{ad} decreases and U_{sel} decreases. Consider a high-income and a low-income student who have the same ability measures (A, θ). U_{sel} is not directly affected by family income, thus two students would have the same nonpecuniary benefit U_{sel}. However, if only the rich student attends a private high school, two students have different application cost U_{apply} and probability of admission P_{ad}.[16] Then two students might choose different application decision because $exp(K)$ is higher for the low-income student.

Second, suppose that both the high-income and the low-income student attend private high school and they have the same ability measures (A, θ). In this case, the relative price $exp(K)$ is

[16] I discuss why the decision on attending private high school is directly affected by family income controlling for ability measures (A, θ) in Appendix H.

same for the rich and poor student. The above condition can be rewritten as $M_1(1 - exp(K)) > [T^s - exp(K)T^n] - \frac{1}{1+r}(M_2^{s*} - exp(K)M_2^{n*})$. It shows that if $exp(K) < 1$ then the student with higher M_1 would be more likely to apply for selective universities. I will call this as *the income effect associated with the nonpecuniary benefit* in this paper, because sufficiently high nonpecuniary benefit is necessary to have $exp(K) < 1$. This effect can be rephrased as follows: if the pecuniary benefit of attending selective universities is too small to compensate for expensive tuition ($[T^s - exp(K)T^n] > \frac{1}{1+r}(M_2^{s*} - exp(K)M_2^{n*})$), then only those who already have a lot of money to consume want to attend selective universities and enjoy the nonpecuniary benefit of prestigious campus life. [17]

Proposition. *The student decides not to apply for selective universities due to liquidity constraint, although the optimal choice is to apply for selective universities iff (i)* $M_1 < \left[\frac{1}{\beta(1+r)} - \lambda\frac{1+\beta}{\beta}\right]M_2^{s*} + \beta T^s$, *and (ii)* $\frac{\left[M_1+\lambda M_2^{s*}-T^s\right]\left[(1-\lambda(1+r))M_2^{s*}\right]^{\beta}}{\left[M_1+\frac{1}{1+r}M_{2*}^{n}-T^n\right]^{1+\beta}} < exp(K) < \left[\frac{M_1+\frac{1}{1+r}M_2^{s*}-T^s}{M_1+\frac{1}{1+r}M_2^{n*}-T^n}\right]^{1+\beta}$.

The first condition implies that low income students are more likely to be constrained. Second condition implies that as the extent of liquidity constraint becomes more stringent, a larger number of constrained students would change application decision and make suboptimal choices.

4.8.2 Tournament Effect (Intensive Margin)

Applying for selective universities increases the marginal benefit of taking additional AP classes because it can also increase the probability of admission. However, the magnitude would differ by student's ability and family income. Also, the admission probability depends on the equilibrium cutoff value h^*, which depends on the number and the academic quality of the applicant pool. Therefore, to understand how tuition subsidy would affect the intensive margin of the choices of students, I need to understand (i) how the marginal benefit of taking AP classes changes by student ability and family income for each given h^*, and (ii) how elevated competition (higher cutoff value $h^{**} > h^*$) affects the effort choices of students. Assume that the unit of credits for AP classes is continuous variable and assume that there is no private high school. Then the optimal choice N_{AP} satisfies

$$\underbrace{\frac{\partial P_{ad}}{\partial N_{AP}}(V^s - V^n)}_{(1)} + \underbrace{P_{ad}\frac{\partial(V^s - V^n)}{\partial N_{AP}}}_{(2)} + \underbrace{\frac{\partial V^n}{\partial N_{AP}}}_{(3)} = \frac{\partial U_{AP}}{\partial N_{AP}} \tag{4.2}$$

For students who do not apply for selective universities, the marginal benefit of effort is just Part (3). Other things being equal, it is decreasing as family income increases. If I focus on students

[17]Note that $exp(K) = 1$ if I do not consider the nonpecuniary benefit and application cost. Then family income would not affect the application choice differently by family income if students have the same M_s.

who do not apply for selective universities, they would take less AP classes than applicants, and rich students would take fewer AP classes than comparable poor students.

Part (1) captures how additional AP class can increase the marginal benefit by increasing the probability of admission. I will call it the *tournament effect*. It is positive if taking AP classes increases the test score ($\frac{\partial H}{\partial N_{AP}} > 0$) and attending selective universities provides higher value than attending nonselective universities ($V^s > V^n$). However, the magnitude of the tournament effect would depend on student ability and family income. Denoting $J = \beta_{sat}\kappa_{sat} + \beta_{AP}\kappa_{AP} + \beta_{GPA}\kappa_{GPA}$ where $\beta_j = \frac{\partial g}{\partial lnH_j}$ and $\kappa_k = \frac{\partial lnH_k}{\partial N_{AP}}$ for $j = \{SAT, GPA, AP\}$ and $k = \{SAT, GPA, AP\}$, Part (1) can be rewritten as

$$\frac{\partial P_{ad}}{\partial N_{AP}}(V^s - V^n) = \underbrace{J\phi(g(\theta, H) - h^*)}_{\text{(I)}} \frac{1}{N_{AP}} \underbrace{\left[ln\Big(\frac{M_1 + \frac{1}{1+r}M_2^s - T^s}{M_1 + \frac{1}{1+r}M_2^n - T^n}\Big) + U_{sel}\right]}_{\text{(II)}} \tag{4.3}$$

Part(I) of equation (4.3) increases in ability and family income if students with high ability (θ, A) learn faster, and attending private high school can increase test scores. Part (II) of equation (4.3) increases by family income if the pecuniary benefit from attending selective universities is smaller than the extra cost ($T^s - T^n > \frac{1}{1+r}(M_2^s - M_2^n)$). Thus, even without liquidity constraint, expensive tuition can be a reason why low-income students take fewer AP classes and achieve lower academic test scores. However, Part (II) of equation 4.3 increases in (A, θ) if $T^s - T^n exp(\gamma) > M_1(1 - exp(\gamma))$ where γ is the wage premium from attending selective universities. Therefore, it is an empirical question whether the tournament effect is more sensitive to student ability than to family income.

Increasing need-based aid from selective universities affects the tournament effect in two ways. First, for low-incomes students, it directly increases Part (II) of equation (4.3) and reduces the income effect. Second, it can raise h^* for everyone if the academic qualification of the low-income student at the margin of application is higher than that of the least competitive attendee in selective universities. This, in turn, directly affects Part (I) of equation (4.3) for everyone. Because of the bell-shape of the normal probability density distribution, increasing h^* has an asymmetric impact on high-achieving ($g(\theta, A) > h^*$) and low-achieving ($g(\theta, A) < h^*$) students. It raises the tournament effect of high-achieving students, whereas it has the opposite impact on low-achieving students. On the other hand, chaning h^* may not reduce the achievement gap between rich and poor students. Increased competition could encourage more affluent students to attend private high school and to put more efforts, thus some of less academiacally able high-income students could remain competitive in the admission process.

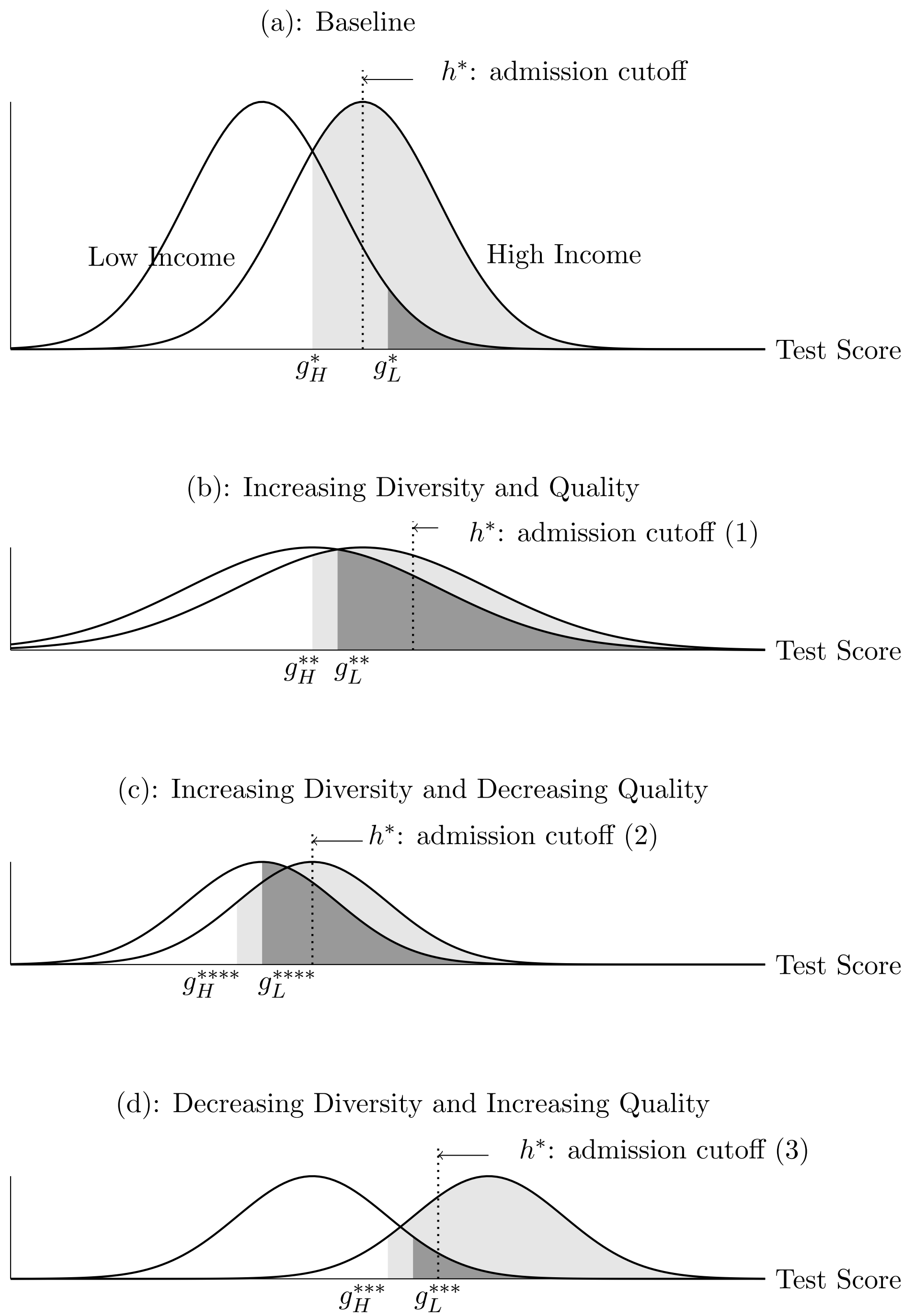


Figure 1: Tournament Effect and Changes in Test Scores

Note. Each graph indicates the distribution of test scores of high- and low-income students. g_H^* and g_L^* are the cutoff values of academic qualification of applicants from high- and low-income students respectively. Both student ability and family income can affect the tournament effect —additional effort applicants put to increase the probability of admission. If the tournament effect changes more sensitively to ability than family income, the academic quality attendees of selective universities could increase (b). If increasing competition and rising tuition discourage application of high-income students, the quality of attendees could decrease although diversity increase (c). If attending private high school is more important than student's own effort choice in the learning process, elevated competition induced larger number of less able high-income students to attend private high school and get higher test scores. In this case, diversity can decrease but quality can increase (d).

5 Identification

The model has seven structural components. Among them, three components have direct data counterparts: the learning technology $H = F_H(x; A, \theta)$, the wage rate $lnWage = F_w(H, I_{sel}, I_{BA}, \theta)$, the admission probability $P_{ad} = \Phi(H, \theta)$. The only issue to identify those components is controlling the selection bias driven by θ. I have four components describing individual's preference: the utility from consumption (U_c), the nonpecuniary benefit from attending selective universities (U_{sel}), the utility cost of application (U_{apply}), the utility cost of taking AP classes (U_{AP}). To identify those preference components, I use the variation in choices of students: the amount of student loan, admission results, the number of AP classes taken by students, whether to attend private high school or not, taking four outcome components as given.

I need to control the selection bias driven by unobservable learning ability that can be positively correlated with family income. As discussed in Cunha and Heckman (2007), noncognitive skills such as self-control or perseverance can affect academic achievement and the wage, and they seem to have a strong correlation with family background. To isolate the impact of monetary resource from that of those abilities, I impose a parametric assumption on the distribution of unobservable ability across individuals, allowing its mean to be affected by student's family income and parent's education. The structural model generates choices and outcomes of students given all other structural parameters, then the conditional mean of student's choices and outcomes are used to identify the distribution. Then the identification of the learning technology $H = F_H(x; A, \theta)$, the wage rate $lnWage = F_w(H, I_{sel}, I_{BA}, \theta)$, the admission probability $P_{ad} = \Phi(H, \theta)$ becomes straightforward.

To identify the intertemporal preference in U_c, I use three variables in data: family income, labor earning, and the amount of students loan borrowed. One potential problem is that the amount of student loan of low-income students could be affected by liquidity constraint. In the model all students have the same intertemporal preference. Thus I identify intertemporal preference from choices of high-income students who are not likely to be liquidity constrained. I use the application decision to identify U_{sel} and U_{apply}. The model predicts that $\frac{U_{apply}}{P_{ad}} - U_{sel}$ accounts for different application decision conditional on family income and labor earning. I can separately identify U_{sel} and U_{apply}, because U_{sel} is obtained only if the student attends selective universities, whereas all applicants have to pay U_{apply}. By comparing applicants with different P_{ad}, I can distinguish U_{apply} from U_{sel}. I include the number of selective universities in student's home state and student's preference for college characteristics reported in 10th grade in U_{sel} as exclusion restrictions. Finally, given the learning technology, the number of AP classes taken by students explains variations in the utility cost of taking AP classes (U_{AP}). I control how many AP classes are offered by the high school for each students to isolate the effect of policy interruption that aims to increase AP participation by low-income students.

Another challenge is to identify the extent of liquidity constraint in the financial market. First, liquidity constraint would decrease low-income students' the consumption at college period. Without constraint, the model predicts no correlation between family income and the intertemporal marginal rate of substitution (MRS) of consumption between two periods. The correlation between family income and the intertemporal MRS could be used to identify the extent of liquidity constraint λ. Second, liquidity constraint can also decrease the application rate of low-income students. If high-achieving low-income students did not apply for selective universities due to liquidity constraint, there are disproportionately more high ability students from low-income families among nonapplicants. Outcome measures, such as wage rate, test scores can be used to see whether there is a upward selection bias among low-income non-applicants.

6 Method of Simulated Moments

There are three components estimated in the first stage which are taken as given in the structural estimation: (i) the distribution of grants students could get when attending selective universities and nonselective universities conditional on family income and test scores, (ii) the number of AP classes offered by high school, (iii) the number of information sources about college application process. I assume that they are exogenous processes and taken as given from student's point of view. High school type affects those processes, thus students take into account those processes when they decide whether to attend private high school or not. I assume that everyone has the same repayment plan: repay the total student loan over the eight years after college graduation with a fixed interest rate 7%.[18]

Given the first stage estimation, I estimate the model with the method of the simulated moments. The criteria function for the structural parameter Θ, $Q(\Theta)$ is constructed as

$$Q(\Theta) = \Big[\sum_{i=1}^{n} Z_i(m_i - \tilde{m}_i(\Theta))\Big]' \hat{\Sigma}^{-1} \Big[\sum_{i=1}^{n} Z_i(m_i - \tilde{m}_i(\Theta))\Big] \tag{6.1}$$

where

$$\tilde{m}_i(\Theta) = \frac{1}{ns} \sum_{s=1}^{ns} m_i^s(\Theta). \tag{6.2}$$

$\tilde{m}_i^s$ indicates the simulated moment for individual i in simulation s, whereas m_i is directly computed from the data. In particular, I simulate the application decision and effort choices for

[18]The interest rate for unsubsidized Federal Stafford Loan is 6.8% as of 2015 (6.83% for PLUS loan). Although the interest rate of the subsidized loan is 3.4%, since private loan would cost more (http://www.finaid.org/loans/privatestudentloans.phtml), I pick the interest rate for unsubsidized loan.Also, I do not consider repayment during college education.

each given parameters, then compare the model generated moment to the data counterpart. On the other hand, Z is the set of instrumental variables orthogonal to error components. The moment is constructed based on the identification argument and orthogonality assumption. I chose the optimal weight matrix $\hat{\Sigma}$ constructed based on the sample variance of the moment.

7 Results

Figure 18-23 show the model fit in terms of the number of applicants and attendees in selective universities, and the number of AP classes students take conditional on their family income and initial math scores. The estimated model tracks the observed pattern in the data fairly well, especially the outcome variations across different income quintile groups. The model slightly understates the number of AP classes taken by students from the bottom quintile of income distribution. This might reflect the impact of policy interruptions that are not fully captured by the number of AP classes offered by high schools.

7.1 Distribution of Unobservable Ability

Table 30 documents the estimation result of the distribution of unobservable learning ability. The mean of unobservable ability increase by 0.03 for additional log family income and increases by 0.06 for additional log parents' educational attainment ($\hat{\phi}_1 = 0.03$ and $\hat{\phi}_2 = 0.06$). Note that the coefficient of θ in the wage rate is normalized to one so that θ increases the log wage rate by one-to-one. On average, students of family income \$10,000 with college graduated parents get 3.7% higher wage rate than students of family income \$50,000 with high school graduated parents, conditional on race, sex, test scores, college selectivity, and college graduation. The variance of the distribution of θ is estimated to be $\hat{\sigma}_\theta = 0.13$, which is quite sizable. Therefore, it is possible that substantial fraction of low-income students have higher learning ability than average high-income students. θ also affects test scores, wage rate, admission probability, application cost, and nonpecuniary benefit.

7.2 The Benefit of Attending Selective Universities

Table 24 shows the estimation result of the wage rate. Attending selective universities increases 0.076 log points of hourly wage rate with 5% significant level. This is after controlling SAT, AP, and GPA scores and the unobservable learning ability of those students. The model assumes that the wage premium from attending selective universities is the same for every attendee.[19]

Figure 24 shows the net pecuniary benefit of attending selective universities compared to attending nonselective universities. I calculate the pecuniary benefit of attending selective universities

[19] Table 20 and Table 21 show the reduced form regression of the wage rate, including the interaction term of family income and college selectivity. I do not find strong evidence from my sample that the wage premium of attending selective universities is significantly different by family income.

and nonselective universities by subtracting the direct cost (tuition and cost of room and book net out of financial aid) from the labor earnings. Then I subtract the pecuniary benefit of attending nonselective universities from the pecuniary benefit of attending selective universities. The result suggests that for most students, attending selective universities is a worse option than attending nonselective universities in terms of monetary return, except for students from the bottom quintile of income distribution who receive a large tuition discount. This implies that the wage premium is too small to compensate for the expensive cost of attending selective universities. Figure 25 shows that the direct cost of attending selective universities, rather than nonselective universities, is more than twice as expensive for high ability students than attending nonselective universities. Merit-based aid in nonselective universities is the reason why attending selective universities is relatively more expensive for high ability student conditional on family income. Need-based aid in selective universities is the reason why it is relatively more expensive for affluent students.[20]

Table 28 shows the estimation result of nonpecuniary benefit of attending selective universities. Ability measures (A, θ) do not significantly affect the nonpecuniary benefit. Instead, the number of selective universities in student's home state and whether the student values the reputation of the college are significant factors. This result seems consistent with the finding in Fu (2014) which finds that students with similar academic achievement scores have very different preference over different type of colleges.

7.3 The Extent of Liquidity Constraint

Table 30 shows the estimation result of the extent of liquidity constraint $\hat{\lambda} = 0.76$; students cannot borrow more than 76% of their expected income in each year of college education. This number is comparable to other papers; for example, it is not statistically different from the estimates that captures the garnishment cost in default state (Lochner and Monge, 2011).

Figure 27 shows the estimated loan amount of students conditional on the initial math score (A) and family income (M_1) quintile. It shows that students who attend selective universities take out more than $10,000 in student loans per year than those who do not attend selective universities. Also, the amount of the loan substantially decreases by family income. Students from the top quintile income distribution do not borrow. Among students who attend nonselective universities, those from the bottom quintile of income distribution borrow more than $10,000 compared to

[20]It is possible that the wage return at age 28 is too early to capture the lifetime earning of graduates from selective universities. For example, if the labor market earning grows faster for graduates from selective universities than those from nonselective universities, the model would understate the pecuniary benefit of attending selective universities. Table 19 shows the wage regression from other data set, the Baccalaureate and Beyond 92 which includes labor earnings 10 years after college graduation. While OLS estimate supports that graduating from selective universities significantly increases the growth rate of earning, IV regression suggests that the correlation between growth rate and college selectivity is not causal.I use the distribution of colleges with different selectivity in students' home state as instrumental variables

those from the second highest quintile of income distribution. Corresponding number for students who attend selective universities is $8,000. The model assumption that students' borrowing limit depends only on their future earning not family income may not be too misleading, because rich students need loans of smaller amounts compared to academically similar poor students.[21]

8 Counterfactual Analysis

First, I discuss the impact of increasing need-based aid from selective universities on the effort choices of low- and high-income high school students, and the academic quality of attendees at selective universities. In the baseline counterfactual policy, selective universities provide $10,000 more in grants per attending student from the bottom quintile of the income distribution. I assume that tuition and financial aid to other income groups do not change. Second, I examine to what extent increasing need-based aid can increase the academic qualifications of attendees. Third, I discuss the quantitative importance of students' initial characteristics to explain the choice differences between low- and high-income students. Third, I discuss the impact of the liquidity constraint on students' choices and outcomes. Fourth, I discuss the respective roles of features of selective universities on the choice difference between rich and poor students. Finally, I compare the impact of increasing need-based aid from selective universities to alternative policies such as increasing need-based aid from nonselective universities, offering more AP classes to low-income students, introducing merit-based aid in the financial aid package from selective universities, income-based affirmative action (income quota), and changes in the admission criteria (Texas Top Ten Law).

8.1 Increasing Need-Based Aid from Selective Universities

If selective universities provide $10,000 more in grants per attending student from the bottom quintile of income distribution, the overall effort level of those low-income high school students, measured by the number of Advanced Placement (AP) classes, increases by 12%. On the other hand, the overall effort choice of all other income backgrounds decreased by 0.4%. Thus, increasing need-based aid from selective universities can decrease the effort difference between the rich and poor students. If I only consider attendees of selective universities, the average effort level of low-income students increases by 1.3%, whereas that of other income groups increases by 0.5%. Finally, it can increase the aggregate effort level of all students by 0.7%.

Increasing need-based aid can also increase the application rate of students from the bottom quintile income distribution by 12%. Among those low-income students who changed their application decision, 53% attend selective universities. The liquidity constraint explains 51% of the increase in the application rate of low-income students. The remaining 49% is explained by the income effect. The income effect indicates that because pecuniary benefit is too small to compensate

[21] Further discussion is in Appendix H

for the expensive tuition, only those who already have a lot of money to consume want to attend selective universities and obtain the nonpecuniary benefit.

Table 43 shows how elevated competition affects the effort choice of students conditional on family income and initial math score. In this exercise, I decrease the average admission probability by 2% by increasing the cutoff value in the admission criteria (h^*). This is the case when the number of applicants increases without changing the academic qualification of the applicant pool. The effort reduction is much larger for students with low math score (8%) compared to students with high math score (0.9%). When I compare the elasticity conditional on students' initial math scores, there is no significant difference by family income. This shows that elevated competition has an asymmetric impact on high and low ability students, and family income cannot counteract this effect. Because only talented low-income students would change application decision, increased need-based aid from selective universities would also improve the academic quality of applicants, which would strengthen the asymmetric impact of elevated competition by students' initial math score.

By promoting competition, increasing need-based aid makes selective universities can become more selective in terms of students' initial academic achievement compared to their family income. This implies that not only will talented high school students have a better opportunity for attending selective universities but also promoting equal opportunities to talented students can lead to a higher level of aggregate effort and achievement score.

8.2 Impact of Redistributive Policy: Tuition vs. Need-Based Aid

Table 46 shows how the academic qualifications of attendees in selective universities changes if selective universities increase need-based aid to students from the bottom quintile income distribution continuously, while raising tuition to keep the total tuition revenue the same. I find that until the aid to low-income students increases by \$30,000, the average academic quality from all income backgrounds increases. After that, the academic quality of attendees starts to decrease. Second, a redistributive policy significantly increases the overall effort level of students from the bottom income quintile group. In particular, if the aid increases by \$30,000, the average effort choice of the lowest income quintile students increases by 20.87%. Of those low-income students, the effort level of attendees in selective universities increases by 1.80%. Third, until \$30,000 is reached, the average effort levels of attendees in selective universities from all other income backgrounds also increases. Fourth, except for students from the lowest quintile of income distribution, the effort level of students who do not attend selective universities decreases when the need-based aid increases. Fifth, the proportion of students from the bottom quintile income distribution in selective universities increases by 2.2% if the aid increases by \$30,000.

Therefore, the marginal return from a redistributive policy in terms of academic qualification of attendees, keeping total revenue equal by raising tuition, is positive until the need-based aid for students from the bottom quintile distribution increases more than double compared to the current policy. Although a redistributive policy would lower the effort level of less academically able high-income students, the aggregate effort level would increase.

8.3 Students' Initial Characteristics and Choice Difference between Rich and Poor Students

Students are mainly differentiated by their initial math score (A), unobservable ability (θ), and family income (M_1). How much students' ability measures (A, θ) explain the different choices of low- and high-income students would show the upper bound of the impact of need-based aid from selective universities. To quantify the respective roles of students' initial characteristics in explaining the choice difference between the rich and poor students, I consider the following counterfactual.

I divide the initial characteristics into four parts: (i) characteristics except initial math score, unobservable ability, and family income (Z), (ii) initial math score (A), (iii) unobservable ability (θ), and (iv) family income. I consider two representative students from the bottom and the top quintile of income distribution. I examine how the choices of students from the top quintile of income distribution change if I substitute one type of initial characteristic with that of a student from the bottom quintile of income distribution.

Table 42 and Figure 26 summarize results. The number (percent) indicates to what extent the model can explain the different choices of rich and poor students based on the true estimated parameters, if one characteristic of rich students were the same as that of poor students. First, different initial math score is the most important reason why rich students take more AP classes and apply more for selective universities than their poor counterparts. It explains 65.9% of the gap in the number of AP classes, 57.7% of the gap in the probability of attending private high school, and 37.4% of the gap in the application rate. Second, the decision whether to attend a private high school is most sensitive to family income (63.1%), whereas the number of AP classes is least sensitive to the income difference (35.8%). Finally, the number of AP classes students take change very sensitively to all of (A, θ, M), whereas the application rate is least sensitive to changes in each type of characteristics. This might suggest that those initial characteristics are a substitute for AP class taking, but they act as a complement when it comes to the application decision.

8.4 The Impact of Liquidity Constraint on Students' Choices

Table 37 summarizes the impact of liquidity constraints on students' college education outcome. I consider four discrete types of college education outcome: dropping out of nonselective universities, graduating from nonselective universities, dropping out of selective universities, and graduating from selective universities. The element in ith row and jth column (i, j) indicates the number of students who choose educational choice i in true equilibrium, but who would have chosen j were it not for the liquidity constraint. It suggests that only 0.8% of the entire population changes their college education choice because of liquidity constraint. Also, an additional 0.8% of students has to consume suboptimal level during their college period due to the liquidity constraint. If I focus on students whose optimal choice is graduating from selective universities without liquidity constraint, 2% attend nonselective universities because of liquidity constraint. This suggests that the liquidity constraint is not empirically important to explain low level of economic diversity in selective universities.

Table 39 and Figure 28 compare the number of students who apply for selective universities in the baseline model and with the counterfactual case without the liquidity constraint. The liquidity constraint reduces the percentage of applicants from the lowest income quintile students by 12%, whereas it increases the application of highest income quintile students by 1.5%. Given the fixed capacity in selective universities, some of high-income students benefit from the presence of the liquidity constraint because it discourages applications by high-achieving, low-income students.

Table 40 and Figure 29 show the impact of the liquidity constraint on the effort choice. Although less than 1% students attended nonselective universities because of the liquidity constraint, its impact is amplified by students' effort choices. In particular, the average AP classes taken by lowest income quintile students decreased by 4.4% because of the liquidity constraint. Figure 30 and Table 41 show that liquidity constraint has larger impact on whether students attend private high school. Without the liquidity constraint, twice as many students from the bottom quintile income distribution would attend private high school than in those in the current equilibrium.

Finally, I discuss the role of current need-based financial aid from selective universities in reducing potential liquidity constraint faced by low-income students. Table 39-41 and Figure 28-30 show the changes in students' choices if the financial aid from selective universities were granted on a completely random basis. Under this assumption, 2.7% of the entire student population would not apply for selective universities because of the liquidity constraint. The number of students from the bottom quintile of income distribution who would not apply for selective universities because of the liquidity constraint would increase from 12% to 20%. The number of constrained students from lowest income quintile increases from 16 to 26 (63%). Therefore, the current financial aid policy from selective universities plays an important role in reducing the liquidity constraint.

8.5 Selective Universities and Choices of Students

Selective universities have several features that can affect students' choices: pecuniary/nonpecuniary benefit, application cost, competitive admission process based on test scores, need-based financial aid, and expensive tuition. Table 44 shows how the choice difference between rich and poor students (the top and bottom quintile of income distribution) changes if I change a certain feature of selective universities to a counterfactual one.

First, the nonpecuniary benefit and test-based competition in the admission process are two of the most relevant features. If the admission process becomes random, the number of AP classes taken by students from the bottom (top) quintile of income quintile decreases by 87% (96%). A similar result would hold were it not for the nonpecuniary benefit from attending selective universities. However, the nonpecuniary benefit is not sufficient to motivate students to take advanced classes if the test score is not valued in the admissions process. Second, if the tuition of selective universities were the same as that of nonselective universities, the difference in the application rate between rich and poor students would have decreased by 12%. A reduced tuition would increase the number of applications of poor students by 15% and decrease the number of applications by high income students by 4%. This suggests that some of less academically able high-income students benefit from expensive tuition of selective universities. Finally, if the financial aid from selective universities were to be granted on a random basis, it would increase the effort difference between rich and poor students by 5%.

8.6 Alternative Policies

I compare a couple of policies in terms of its impact on low-income student's academic achievement. First, increasing need-based aid from selective universities is more effective per dollar than the same policy implemented from nonselective universities or offering more AP classes to low-income students, because it ties financial incentive to students' intrinsic motivation to put more effort. Second, income-based affirmative action (income quota) and changing admission criteria such that only GPA affects the admission result (Texas Top 10 Law) are much more effective to promote economic diversity than increasing need-based aid from selective universities. However, both policies would substantially decrease the overall effort level of students, which suggests the importance of merit-based admission criteria when it comes to students' academic achievement. Table 47 to Table 50 compare the impact of those policies on student's choices.

8.6.1 Increasing Need-Based Aid from Nonselective Universities

I consider the case in which nonselective universities raise need-based aid for students from the bottom quintile of income distribution by \$10,000. The lowest income quintile students take 3∼14% less AP classes if nonselective universities increase need-based aid for them. The number of AP classes taken by highest income quintile students increases by 1∼ 10%. If selective universities raised need-based aid it would decrease the effort level of low-income students by 10∼21% and would have decreased the effort level of high-income students by 1∼2 %. Considering the number of low-income students who attend nonselective universities, this policy requires much larger budget than that for the baseline counterfactual because increasing need-based aid from selective universities provides additional grants only to talented low-income students. When it comes to increaseing academic achievement of low-income students, increasing student incentive to attend selective universities is more effective than reducing the tuition burden of attending nonselective universities.

8.6.2 Offering More AP Classes to Low-Income Students

There has been an increase in policy interruption to encourage more low-income students to participate in the AP program. To compare the impact reducing effort cost with the impact of increasing need-based financial aid from selective universities, I consider the following counterfactual. I compute the total budget required to implement the baseline counterfactual. Based on the estimated start-up cost for the average AP class documented by the College Board,[22] I compute how many AP classes can be newly offered by high schools in which students from the lowest income quintile families attend. I find that the same budget can offer 1.58 more AP classes to every student from the bottom quintile of the income distribution.

Low-income students takes 5∼13% more AP classes if more AP classes are offered by the high school, but offering more AP classes to low-income students does not affect choices of students from other income quintile groups. It is less effective per dollar to increase economic diversity in selective universities or to increase the number of AP classes taken by low-income students compared to increasing need-based aid from selective universities. Offering more AP classes targets all low-income students, whereas increasing grants from selective universities targets more selective groups who have a higher marginal benefit and a lower marginal cost of taking more AP classes.

8.6.3 Income Quota and Texas Top 10 Law

As an alternative to race-based affirmative action in college admission process, income-based affirmative action is often discussed. [23]Also, from 1997, Texas introduced the Texas Top 10 Law, which guarantees admission to flagship universities in Texas for students who ranked in the top 10% of their classes in a Texas high school. Motivated by these discussions, I consider following the two counterfactual policies that change admission criteria. One is the income quota system that gives additional points to the lowest income quintile students in the admission process so that the number of lowest income quintile students doubles. The other is similar to the Texas Top 10 Law such that only the GPA is considered in the admission process. These policies do not need extra resources to be implemented.

Changing admission criteria is more effective at increasing economic diversity than increasing need-based aid from selective universities. However, both policies substantially decrease the academic achievement of the overall population. Although the income quota system can increase the proportion of the lowest income quintile students by 20∼55%, high-achieving low-income students take 6% fewer AP classes and high-achieving, high-income students take 2% fewer AP classes. Similarly, the Texas Top 10 Law can increase diversity, although it is less effective than income quota

[22]4,343 dollar per class with 25 students

[23]http://www.nytimes.com/roomfordebate/2014/04/27/should-affirmative-action-be-based-on-income

system. However, most students take less than half of the AP classes compared to what they take in the true estimated case if only the GPA is considered in the admission process.

9 Out-Of-Sample Prediction

To examine the validity of structural model, I consider an out-of-sample prediction. The National Education Longitudinal Study of 1988 is a nationally representative sample of 8th graders and it has four follow-up surveys in 1990, 1992, 1994, and 2000. The ELS2002 and the NELS1988 have almost identical survey instruments. Assuming that changes in financial aid policy from selective universities are exogenous from students' point of views, I consider the following exercise. First, I estimate the financial aid policy from selective universities in the NELS1988 sample. Second, I substitute the financial aid policy of the ELS2002 cohort with that of the NELS1988 cohort. Based on the estimated structural model and by using the data of the ELS2002, I can predict students' application decisions, admission results conditional on family income and initial math scores. Third, I look at the raw data of the NELS1988 and calculate the joint distribution of family income, the initial math score, the application, and the admission result. Then I compare how much the predicted outcome tracks the actual data observed in the earlier cohort, the NELS1988.

Figure 31-34 show the result of out-of-sample prediction. Because the aggregate application rate increased from 18% to 42% during this period, I focus on the composition rates of applicants and attendees in selective universities conditional on family income and initial math score rather than focusing on the application and admission rate of each group. Also, I do not consider the number of AP classes taken by students because the number of AP classes offered by the high school increased drastically during this period.

The estimated model predicts the NELS1988 data fairly well in terms of the composition rate by family income quintile. The disparity between the predicted model and the NELS1988 data is less than 3% for each income quintile group. However, the model predict much smaller disparity between students with high and low math scores. In particular, the data show a stark difference in the composition rates between students from the first and the second highest quintile of math score distribution. More difficult application process and much fewer number of available AP classes in the NELS1988 cohort may explain why the selection into selective universities in earlier cohort was more strongly driven by student's initial math score than the sorting pattern predicted based on the ELS2002 sample.

10 Conclusion

Despite their expensive sticker prices, selective universities provide far more generous financial aid to low-income students than nonselective universities. This paper argues that increasing need-based aid from selective universities may not only decrease the achievement gap between rich and poor but also improve the academic quality of attendees of selective universities from *all* income backgrounds. The estimated tournament effect-the additional effort applicants put in to increase the probability of their admissions-shows more elasticity in relation to student ability than in relation to family income. Thus, increased competition will raise the effort levels of high-achieving high school students.

Therefore, increasing need-based aid from selective universities actually promotes meritocracy, insofar as it is measured by the quality of attendees in selective universities. As long as the college admissions process is based on test scores, in which the student's own ability is the determining factor, the further increase in need-based aid from selective universities will not lead to a mismatch, resulting in fewer well-prepared students attending selective universities, as has been considered to be the case in the past due to race-based affirmative action.

Important things remain for the future studies. First, this paper does not model problems from the perspective of the colleges. However, competition between two selective universities would be an important reason for selective universities to care also about attracting middle- and high-income students. Thus, incorporating competition between colleges could be valuable extension. Second, this paper does not distinguish between different types of loans. However, the cost of financing varies substantially according to whether or not it is supplied by a private loan and according to the type of Federal loan. Therefore, further study of those features would provide relevant policy implications. Finally, incorporating job search capability, college major, and occupational sorting would be another relevant extension, because the benefit of attending selective universities and taking more advanced classes may depend on those margins.

References

[1] Brant Abbott, Giovanni Gallipoli, Costas Meghir, and Giovanni L.Violante, *Education policy and intergenerational transfer in equilibrium*, working paper (2013).

[2] Peter Arcidiacono, *Affirmative action in higher education: How to admission and financial aid rules affect future earnings?*, Econometrica **75(3)** (2005), 1477–1524.

[3] Peter Arcidiacono, Esteban Aucejoand Patrick Coate, and Joseph Hotz, *Affirmative action and university fit: Evidence from proposition 209*, working paper (2012).

[4] David Autor and David Scarborough, *Does job testing harm minority workers? evidence from retail establishments*, The Quarterly Journal of Economics **February** (2008), 219–277.

[5] Christopher Avery, Caroline Hoxby, Clement Jackson, Kaitlin Burek, Glenn Poppe, and Mridula Raman, *Cost should be no barrier: An evaluation of the first year of harvard's financial aid initiative*, working paper (2006).

[6] Anthony Carnevale and Jeff Strohl, *Separate and unequal - how higher education reinforces the intergenerational reproduction of white racial privilege*, Center on Education and Workforce, Goergetown University (2013).

[7] Satvaiit Chatterjee and Felicia Ionescu, *Financial aid policy: Lessons from research*, working paper (2010).

[8] Flavio Cunha and James Heckman, *The technology of skill formation*, working paper (2007).

[9] Stacy Dale and Alan Krueger, *Estimating the payoff to attending a more selective college: An applicaion of selection on observables and unobservables*, The Quarterly Journal of Economics **November** (2002), 1491–1527.

[10] Sujan Dynarski and Judith Scott-Clayton, *Insuring student loans against the risk of college failure*, working paper (2013).

[11] Ian Fillmore, *Price discrimination and public policy in the u.s. college market*, working paper (2014).

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

[13] David Hawkins and Melissa Clinedinst, *State of college admission*, National Association for College Admission Counseling (2006).

[14] Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante, *Optimal tax progressivity: An analytic framework*, working paper (2014).

[15] Brent Hickman, *Effort, race gaps, and affirmative action: A game-theoretic analysis of college admissions*, working paper (2010).

[16] Brent Hickman, *Pre-college human capital investment and affirmative action: A structural policy analysis of us college admissions*, working paper (2013).

[17] Mark Hosekstra, *The effect of attending the flagship state university on earnings: A discontinuity-based approach*, The Review of Economics and Statistics **91(4)** (2009), 714–724.

[18] Caroline Hoxby, *The changing selectivity of american colleges*, The Journal of Economic Perspectives **23(4)** (2009), 95–118.

[19] Caroline Hoxby and Sarah Turner, *Expanding college opportunities for high achieving, low income students*, working paper (2013).

[20] Adam Kapor, *Distributional effects of race-blind affirmative action*, working paper (2015).

[21] Kristin Klopfenstein, *Advanced placement: Do minorities have equal opportunity*, Economics of Education Review **23** (2004), 115–131.

[22] Edward.P Lazear and Sherwin Rosen, *Rank-order tournaments as optimal labor contracts*, The Journal of Political Economy **89(5)** (1981), 841–864.

[23] Lauren Linzmeier, *Assessing the effects of advanced placement policies*, working paper (2012).

[24] Lance Lochner and Alexander Monge-Naranjo, *The nature of credit constraints and human capital*, American Economic Review **101** (2011), 2487–2529.

[25] Taylor Nadauld Lucca, David and Karen Shen, *Credit supply and the rise in college tuition: Evidence from the expansion in federal student aid programs*, Federal Reserve Bank of New York Staff Report (2015).

[26] C. Manski and D Wise, *College choice in america*, Harvard University Press (1983).

[27] Diego Restuccia and Carols Urrutia, *Intergenerational persistence of earning: The role of early and college education*, American Economic Review **94(5)** (2004), 1354–1378.

[28] Sherwin Rosen and Robert Willis, *Education and self-selection*, The Journal of Political Economy **87(5)** (1979), 7–36.

[29] Carl Sanders, *Skill accumulation, skill uncertainty, and occupational choice*, working paper (2014).

[30] Michael Sattinger, *Assignment models of the distribution of earnings*, Journal of Economic Literature **31(2)** (1993), 831–880.

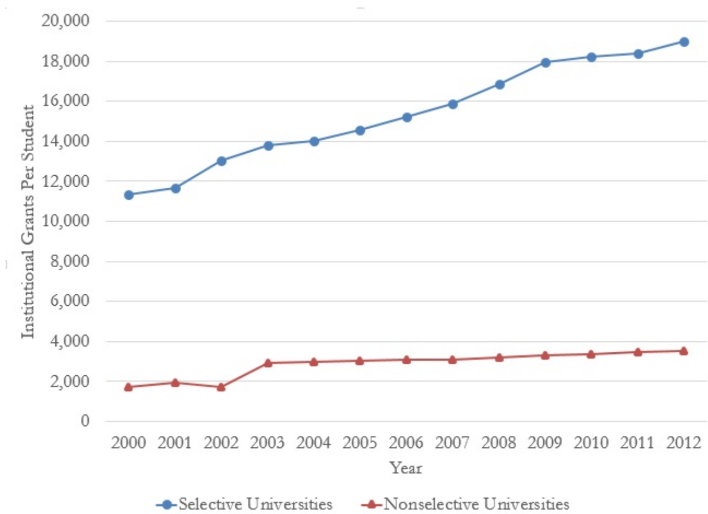
[31] Ralph Stinerbrickener and Todd Stinerbrickner, *The effect of credit constraint on the college drop-out decision: A direct approach using a new panel study*, American Economic Review **98(5)** (2008), 2163–2184.

[32] Ralph Stinerbrickener and Todd Stinerbrickner, *Learning about academic ability and the college drop-out decision*, working paper (2009).

[33] Jenni Woo and Matthew Soldner, *Degrees of debt: Student borrowing and laon repayment of bachelor's degree recipients 1 year after graduation: 1994, 2001, and 2009*, U.S.Department of Education **NCES 2014-011** (2014).

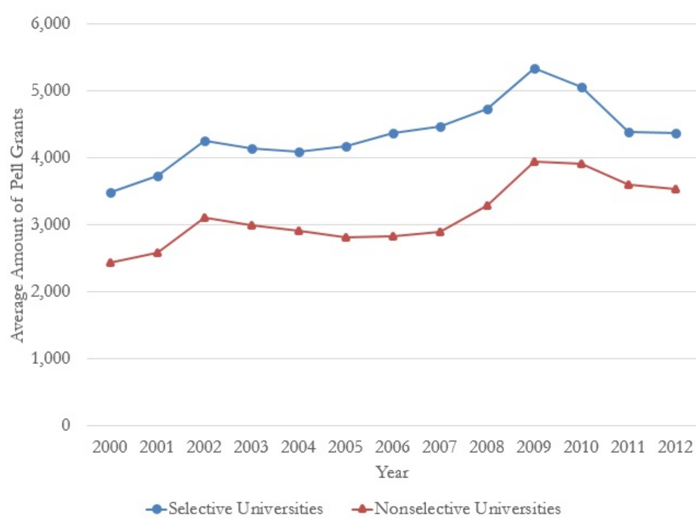
Appendix A: Motivating Facts

Figure 2: Trend of Institutional Grant Per Student from Selective and Nonselective Universities



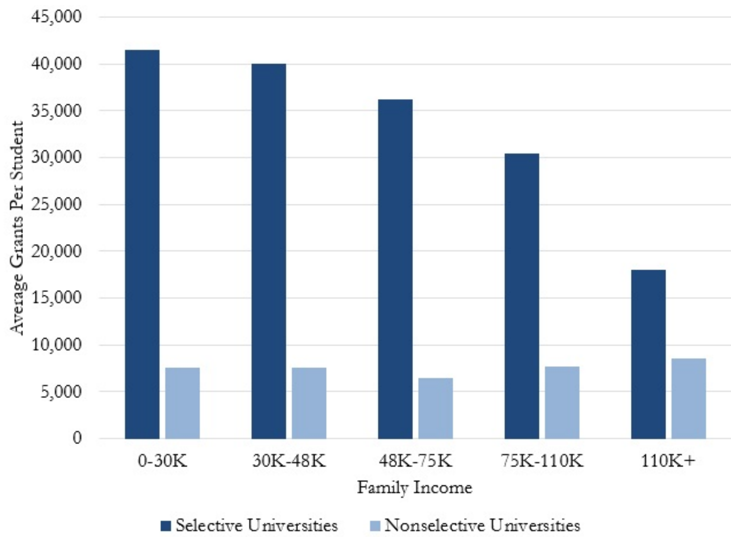
Note. Data are from Integrated Postsecondary Education System (IPEDS) 2000-2012. The unit is 2004 Dollar. Selective universities are 4-year colleges with top 2 categories of NCES-Barron's Admission Competitive Index, which account for about 20% of entire enrollees of 4-year colleges.

Figure 3: Trend in Average Federal Grants



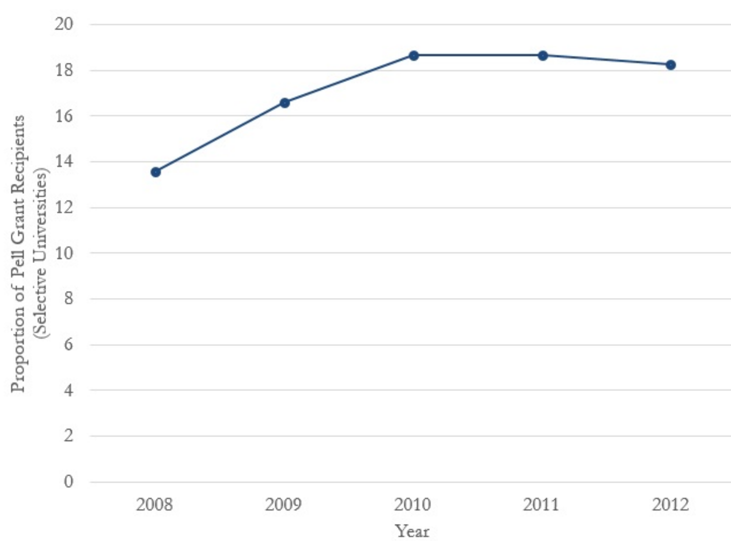
Note. Data are from Integrated Postsecondary Education System (IPEDS) 2000-2012. The unit is 2004 Dollar. Federal grants includes Pell Grants, Federal Supplemental Educational Opportunity Grants (FSEOG), Teacher Education Assistance for College and Higher Education (TEACH) Grants, Iraq and Afghanistan Service Grants.

Figure 4: Average Aid Given to Attendees from Different Income Category (2012-2013 Academic Year



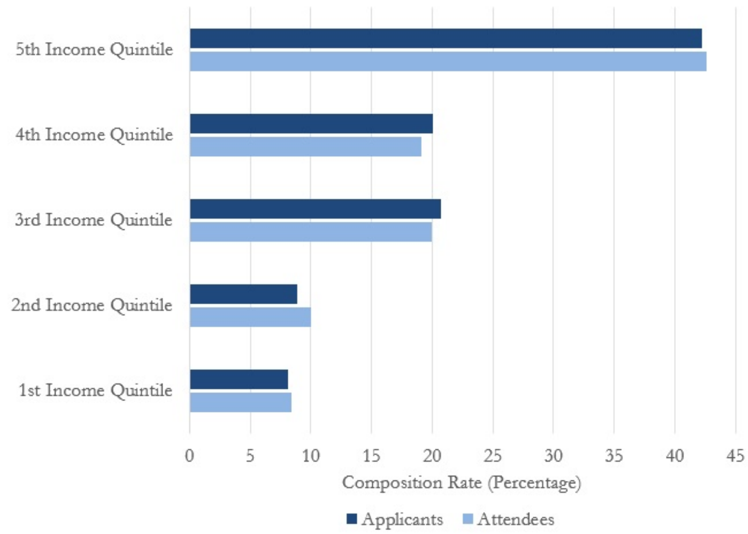
Note. Data are from IPEDS of 2012-2013 academic year. The unit is 2012 dollar.

Figure 5: Trend in the Proportion of Pell Grants Recipient in Selective Universities (Source:IPEDS)



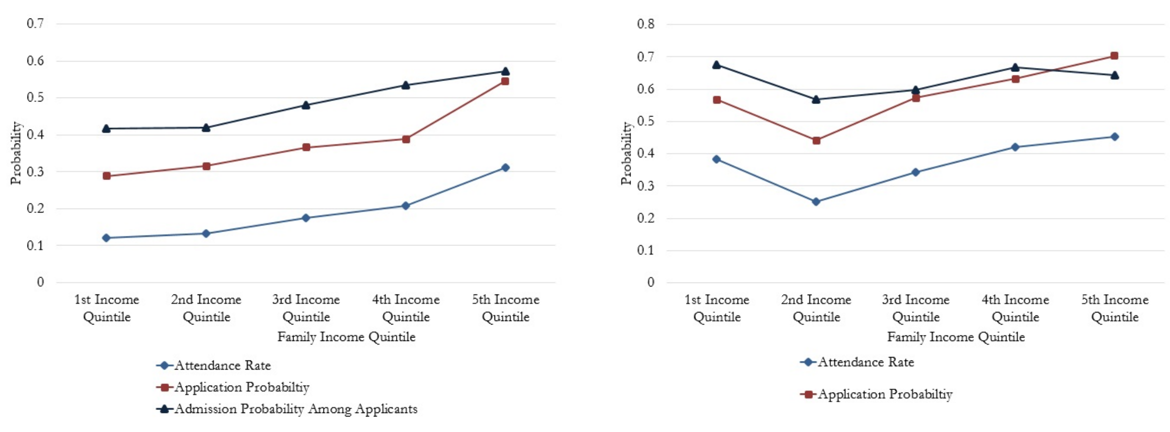
Note. Data come from the IPEDS of 2012-2013 academic year. 97% of Pell Grant recipient comes from families with lower than $50,000 annual income.

Figure 6: Composition Rate of Students of Attendees and Applicants of Selective Universities By Income Quintiles



Note. Data come from the Education Longitudinal Study of 2002 (ELS2002) and NCES-Barron's Admission Competitive Index 2004. The unit is percent. It is the proportion of attendees (applicants) from each quintile of income distribution among all attendees (applicants) in selective universities.

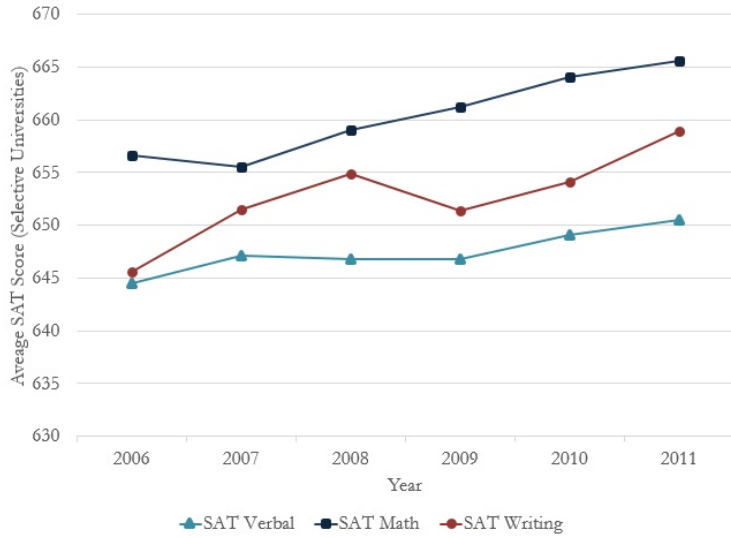
Figure 9: Attendance Rate, Application Rate, and Admission Rate upon Application



(a) All Students (b) High-Achieving Students (Top 20th Math Score)

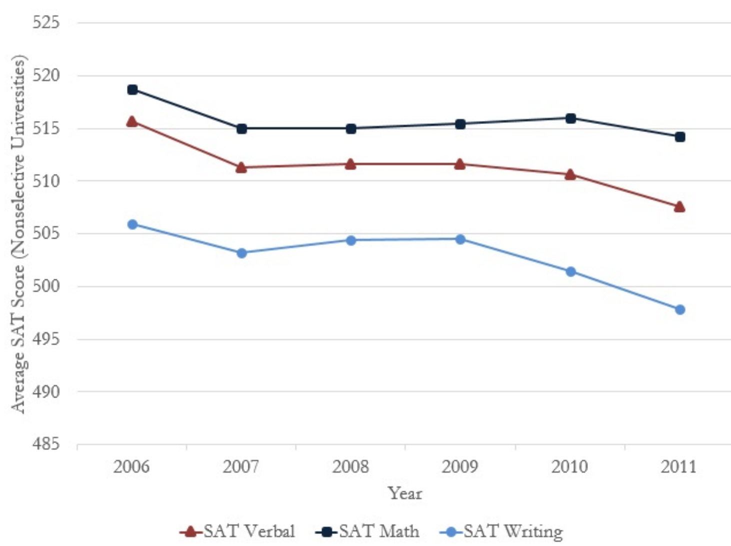
Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). The unit is percent. Figure 9(a) is based on the entire sample, and Figure 9(b) is based on high-achieving students who belongs to top 20 math score at the sophomore year of the high school. Attendance rate (Application probability) is the ratio of the number of students who attend (apply for) selective universities to the number of all students conditional on income quintile. Admission rate is the ratio of the number of attendees to the number of applicants conditional on income quintile.

Figure 10: Trend in Average SAT Scores of Attendees in Selective Universities



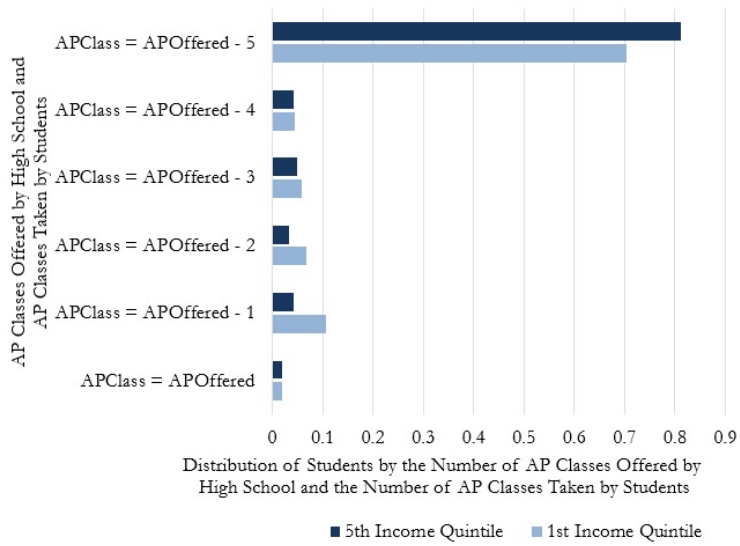
Note. Data come from the Integrated Postsecondary Education Data System (IPEDS) 2006-2012. It is the average SAT score of attendees in selective universities. SAT has three sections: verbal, math, and writing. Selective universities belong to top two categories of the NCES-Barron's Admissions Competitiveness Index, which account for about 20% of 4-year college enrollees

Figure 11: Trend in Average SAT Scores of Attendees in Nonselective Universities



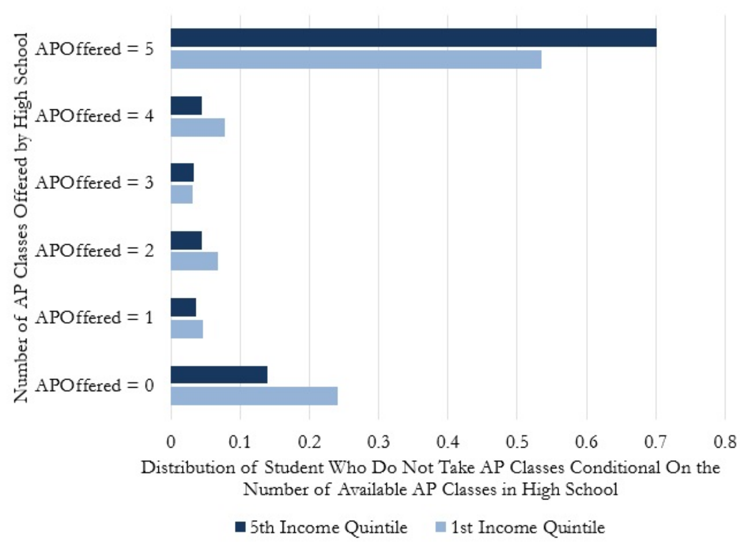
Note. Data come from the Integrated Postsecondary Education Data System (IPEDS) 2006-2012. It is the average SAT score of attendees in nonselective universities.SAT has three sections: verbal, math, and writing. Selective universities belong to top two categories of the NCES-Barron's Admissions Competitiveness Index, which account for about 20% of 4-year college enrollees

Figure 12: The Number of AP Classes offered by High School and the Number of AP Classes Taken By the Student (Among Those Who Take At Least One AP Class)



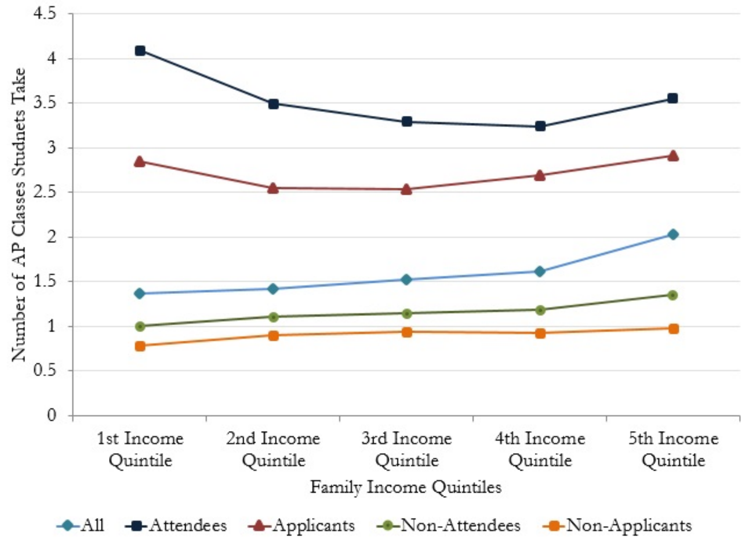
Note. Data come from the High School Transcript of the Education Longitudinal Study of 2002. First income quintile indicates the bottom quintile of income distribution. It is the distribution of students by the relationship between the number of AP classes offered by high school and the number of AP classes students take. I only consider students who take at least one AP class.

Figure 13: The Number of AP Classes offered by High School Among Those Who Do Not Take AP Class



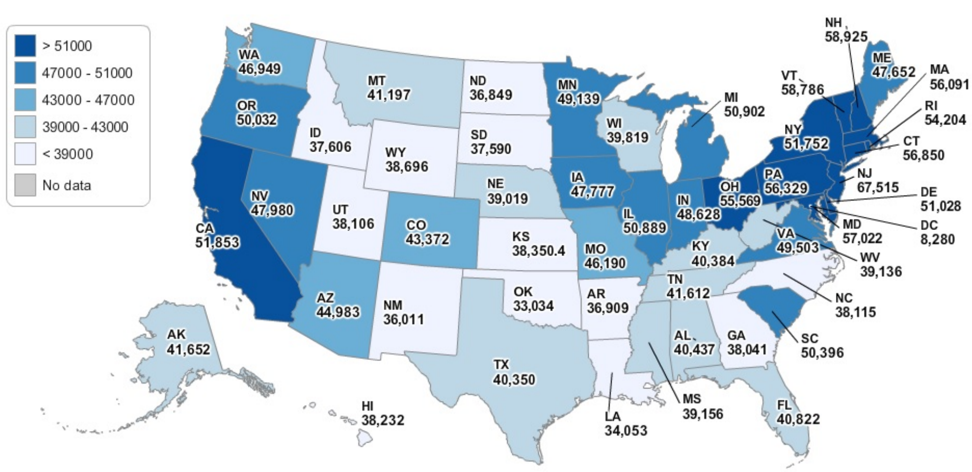
Note. Data come from the the High School Transcript of the Education Longitudinal Study of 2002. First income quintile indicates the bottom quintile of income distribution. It is the distribution of students who do not take AP classes conditional on the number of AP classes available in high school.

Figure 14: The Number of AP Classes Taken by Students



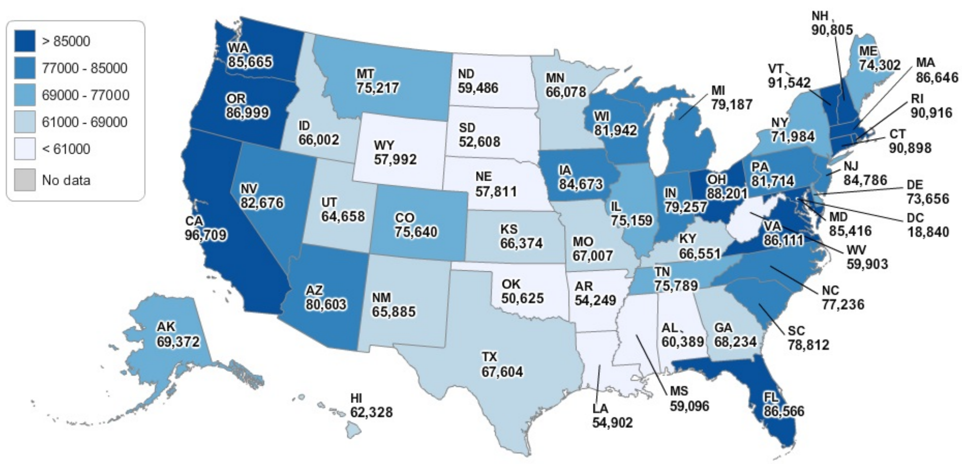
Note. Data come from the the Education Longitudinal Study of 2002. First income quintile indicates the bottom quintile of income distribution. It is the average number of AP classes taken by each group conditional on family income quintile. All indicates the average AP classes students take by family income quintile. Attendees (Applicants) are those who attend (apply for) selective universities. Non-attendees (Non-applicants) are those who do not attend (apply for) selective universities.

Figure 15: Total Cost of Attending Nonselective Universities in Home States for 4 Years



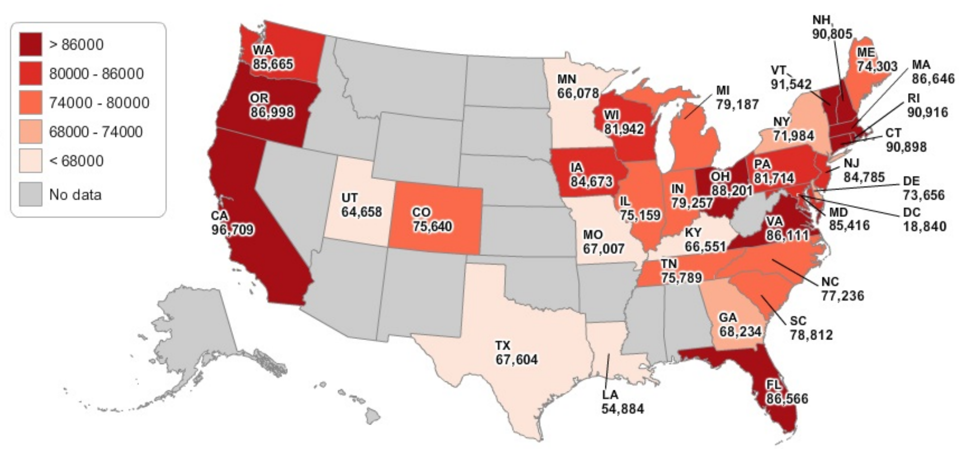
Note. Data come from the the Integrated Postsecondary Education Data System 2004. The total cost includes tuition/fees and the cost for room and book. I do not use the cost of living with family but the cost of living in campus.

Figure 16: Total Cost of Attending Nonselective Universities in Out of Home States for 4 Years



Note. Data come from the the Integrated Postsecondary Education Data System 2004. The total cost includes tuition/fees and the cost for room and book.

Figure 17: Total Cost of Attending Selective Universities in Out of Home States for 4 Years



Note. Data come from the the Integrated Postsecondary Education Data System 2004. The total cost includes tuition/fees and the cost for room and book.

Appendix B: Motivation (Tables)

Table 1: Average Tuition by College Selectivity for Students

College Selectivity	(1) Tuition (Out of State)	(2) Tuition (In State)	(3) Room and Book	(1) + (3)
Barron 1	26,911(7,671)	25,461 (9,814)	9,301 (2,130)	34,762
Barron 2	23,460(7,093)	20,976 (10,002)	8,476 (1,747)	29,452
Barron 3	18,564(5,198)	16,119 (7,677)	7,653 (1,489)	23,772
Barron 4	14,817(4,194)	12,025 (6,599)	7,079 (1,550)	19,105
Barron 5	13,040(3,917)	10,415 (5,940)	6,901 (1,527)	17,316
Barron 6	9,994(3,787)	7,097 (4,534)	5,902 (1,455)	12,999
Barron 7	17,160(6,517)	16,622 (7,130)	9,025 (2,164)	25,648
Selective Univ	24,824(7,497)	22,749 (10,140)	12,492 (7,131)	35,241
Nonselective Univ	15,061(5,098)	12,492 (7,131)	7,180 (1,651)	22,241

Note. Data come from the Integrated Postsecondary Education System (2004) and Barron's Admission Competitive Index (2004). Units is 2004 Dollar. Standard errors are in parentheses.

Table 2: Types of No-Loan Policies

Type	Description	Universities
No Loans	Eliminate loans from financial aid package	Princeton(all students) Rice, U Penn (of low income students)
Loan Caps	Low cap on loans for low-income students	Brown
No Parental Contribution	Eliminate the parental contribution, but retain the student contribution (self-help level)	Yale Stanford
Pell Grant Match	Match the student's Federal Pell Grant reduce self-help level	MIT U of Minnesota (previously)

Note. Data come from http://www.finaid.org/questions/noloansforlowincome.phtml. Description of types of no-loan policies.

Table 3: Changes in Financial Aid Policy from Top Universities (I)

University	Description	Family Income	Year
Brown	Caps loan at $7,000 (4 year)	Less than $30,000	1999-2000
	Caps loan at $11,500 (4 year)	Less than $50,000	1999-2000
	No loan in financial aid package	Less than $100,000	2008-2009
	Limit 4 year debt to $12,000	$100,000 to $125,000	2008-2009
	Limit 4 year debt to $16,000	$125,000 to $150,000	2008-2009
	Limit 4 year debt to $20,000	More than $150,000	2008-2009
	No parental contribution	Less than $60,000	2008-2009
Columbia	Replace loans with grants	Less than $50,000	2007-2008
	Replace loans with grants	All Columbia College/SEAS	2008-2009
	No parent contribution	Less than $60,000	2008-2009
	Reduced parent contribution	$60,000 to $100,000	2008-2009
Cornell	Replace loans with grants	Less than $60,000	2008-2009
	Caps need based loans at $3,000	$60,000 to $120,000	2008-2009
	Replace loans with grants	Less than $75,000	2009-2010
	No parental contribution	$60,000 to $100,000	2009-2010
	Caps need based loans at $3,000	$75,000 to $120,000	2009-2010
	Replace loans with grants	less than $60,000	2013-2014
	No parental contribution	less than $60,000	2013-2014
	Caps need based loans at $2,500	$60,000 to $75,000	2013-2014
	Caps need based loans at $5,000	$75,000 to $120,000	2013-2014
	Caps need based loans at $7,500	More than $120,000	2013-2014
Dartmouth College	No loan in financial aid package	All	2008-2009
	Free tuition	less than $75,000	2008-2009
	Free tuition with no loan	less than $75,000	2011-2012
	Caps on annual loan ($2,500 to $5,500)	$75,000 to $200,000	2011-2012
	Free tuition with no loan	Less than $100,000	2012-2013
	Caps on annual loan ($2,500 to $5,500)	$100,000 to $200,000	2012-2013
Harvard	No parent contribution	Less than $40,000	2004-2005
	No parent contribution	Less than $40,000	2006-2007
	Replace loans with grants	All	2008-2009
	Zero to 10 Percent Standard		
	at most 10% of their income	$120,000 to $180,000	2008-2009
	at most 0% to 10% of Income	$60,000 to $120,000	2008-2009
	0%	Less $60,000	2008-2009

Note. Data come from http://www.finaid.org/questions/noloansforlowincome.phtml

Table 4: Changes in Financial Aid Policy from Top Universities (II)

University	Description	Family Income	Year
Princeton	Replace loan with grants	low income family	1998-1999
	Replace loan with grant	All with financial aid	2001-2002
Yale	No parent contribution	less than $45,000	2005-2006
	Reduce parent contribution	$45,000 to $60,000	2005-2006
	Replace loan with grants	All students	2008-2009
	No parent contribution	less than $60,000	2008-2009
	Limit parent contribution (1 to 10%)	$60,000 to $120,000	2008-2009
	Limit parent contribution (10%)	$120,000 to $200,000	2008-2009
	Increase grants	child in college $\geq$ 2	2008-2009
	Replace loans with grants	All students	2010-2011
	No parent contribution	Less than $65,000	2010-2011
	Limit parent contribution (1 to 10%)	$65,000 to $130,000	2010-2011
Stanford	No parent contribution	Less than $45,000	2006-2007
	Replace loan with grants ($4,500 contribution by earning from work)	All Families	2008-2009
	No parent contribution, No tuition/room/board	Less than $60,000	2008-2009
	No tuition	Less than $100,000	2008-2009
University of Pennsylvania	No loan in financial aid package	Less than $50,000	2006-2007
	No loan in financial aid package	Less than $60,000	2007-2008
	No loan in financial aid package	Less than $100,000	2008-2009
	No loan in financial aid package	All	2009-2010

Note. Data come from `http://www.finaid.org/questions/noloansforlowincome.phtml`

Table 5: Changes in the Percentage of Freshmen Who Receive Merit-Based Aid from Selective Universities from 2007-2008 to 2014-2015

College	Tuition/Fees	Freshman getting merit aid	Change from 2007-8	Average merit aid	Change from 2007-8	Undergraduate enrollment enrollment
Goerge Washinton University (DC)	$44,148	20%	30%	$18,495	-9%	9,680
University of Chicago (IL)	$42,783	16%	19%	$10,636	-19%	5,311
Johns Hopkins University (MD)	$42,780	1%	14%	$29,310	0%	5,038
St.Lawrence University (NY)	$42,726	33%	2%	$14,978	15%	5,238
Bowdoin College (NY)	$42,396	6%	58%	$1,020	-5%	1,750
Washington University in St. Louis (MO)	$41,992	19%	2%	$8,803	-13%	6,322
Mount Holyoke College (MA)	$41,842	24%	-9%	$8,803	-13%	6,322
University of Notre Dame (IN)	$40,717	2%	30%	$13,889	106%	8,411

Note. Data come from Education Life of the New York Times
`http://www.nytimes.com/interactive/2012/07/08/education/edlife/8edlife_chart.html`.

Table 6: Student Composition in Selective Universities

	All Students		Top 20 Math Scores	
	Attendees	Applicants	Attendees	Applicants
1stIncomeQuintile	8.4	10.3	8.7	8.1
2ndIncomeQuintile	10.0	12.2	7.9	8.8
3thIncomeQuintile	19.9	21.2	19.5	20.8
4thIncomeQuintile	19.1	18.3	21.1	20.0
5thIncomeQuintile	42.6	38.0	42.9	42.2
Total	100.0	100.0	100.0	100.0

Note. Data come from the Education Longitudinal Study of 2002. The unit is percent. It is the composition rate of students from each quintile of income distribution. High-achieving students are those with top 20 math score in the sophomore year of the high school. Selective universities are 4-year colleges with top 2 categories of Barron's Admission Competitive Index, which account for about 20% of entire enrollees of 4-year colleges.

Table 7: Admission and Application Probability (ELS 2002)

	All			Top 20 Ability		
	(1)	(2)	(3)	(1)	(2)	(3)
	Admitted/All	Applied/All	Admitted/Applied	Admitted/All	Applied/All	Admitted/Applied
1stQuintile	12.0	28.9	41.6	38.3	56.7	67.6
2ndQuintile	13.2	31.5	41.9	25.0	44.0	56.8
3thQuintile	17.6	36.5	48.1	34.2	57.2	60.0
4thQuintile	20.7	38.8	53.4	42.1	63.2	66.7
5thQuintile	31.2	54.5	57.2	45.2	70.2	64.4

Note. Data come from the Education Longitudinal Study of 2002. The unit is percent. (1): attendance rate, (2): application rate, and (3): admission probability. Attendance (Application) rate is obtained by dividing the number of attendees (applicants) of selective universities by the number of students (four-year college attendees) in each quintile of income distribution. Admission probability is obtained by dividing the number of attendees by the number of applicants from each quintile of income distribution. Selective universities are 4-year colleges with top two categories of Barron's Admission Competitive Index, which account for about 20% of entire enrollees of 4-year colleges.

Table 8: The Number of AP Classes Taken by Students and Application Decision (OLS)

	(1)	(2)
Black	-0.143	0.285***
Asian	1.215***	1.179***
Hispanic	0.325***	0.517***
Female	0.280***	0.109*
lnMath	0.674***	-0.045
lnFincome	-0.024	-0.038
lnParEdu	0.842***	0.217
lnSAT		4.481***
lnGPA		1.881***
I_{Apply}	1.213***	0.786***
$NumAPOffered$	0.062***	0.067***
Constant	-4.374***	-33.091***

Note. Data come from the Education Longitudinal Study of 2002. This is an original linear regression result of the number of AP classes students take on the application decision (I_{Apply}). I also control students' demographic characteristics, family income (lnFincome, log value) ,parents' education (lnParEdu, log value), test scores(log value), and the number of AP classes offered by high school ($NumAPOffered$).

Table 9: Attending Private High School and Application Decision (Probit)

	(1)	(2)
Black	-0.210**	-0.235***
Asian	-0.787***	-0.801***
Hispanic	0.036***	0.023
Female	-0.212***	-0.145***
lnMath	-0.094***	-0.174***
lnFincome	0.314***	0.293***
lnParEdu	1.186***	1.064***
lnSAT		1.351***
lnGPA		-0.996***
I_{Apply}	0.186***	0.131***
Constant	-6.747***	-14.150***

Note. Data come from the Education Longitudinal Study of 2002. This is probit estimation of whether to attend private high school on the application decision (I_{Apply}). I also control students' demographic characteristics, family income (lnFincome, log value) ,parents' education (lnParEdu, log value), and test scores(log value).

Table 10: Average Financial Aid

Family Income	Less than 30,000	30,000-48,000	48,000-75,000	75,000-110,000	More than 110,000
Selective Univ 2008	36,245	34,115	29,082	23,372	13,429
Nonselective Univ 2008	6,428	5,803	4,907	5,732	5,963
Ratio	5.80	5.88	5.93	4.08	2.25
Selective Univ 2012	41,455	40,069	36,180	30,381	18,007
Nonselective Univ 2012	7,539	7,558	6,410	7,646	8,527
Ratio	5.50	5.30	5.64	3.97	2.11
Increase in Aid (Sel)	5,209 (114)	5,954 (130)	7,098 (155)	7,009 (153)	4,578 (100)
Increase in Aid (NonSel)	1,291 (50)	1,755 (68)	1,503 (59)	1,914 (75)	2,564 (100)
Ratio	4.03	3.39	4.72	3.66	1.79

Note. Data come from the IPEDS 2008 and the IPEDS 2012. Selective universities are 4-year colleges with top 2 categories of Barron's Admission Competitive Index, which account for about 20% of entire enrollees of 4-year colleges.

Table 11: Application,Attendance and Admission Rate (Probit Estimation)

	Parameter Estimates			Marginal Effect		
Specification	Application	Attendance	Admission	Application	Attendance	Admission
	(1)	(2)	(3)	(1)	(2)	(3)
Black	0.524***	0.415***	0.236	0.157***	0.092***	0.081
Asian	0.513***	0.357***	0.173	0.158***	0.078***	0.060
Hispanic	0.578***	0.517***	0.284**	0.176***	0.117***	0.097**
Female	-0.010	0.063	0.088	-0.003	0.013	0.031
lnfincome	0.136***	0.212***	0.164***	0.040***	0.042***	0.057***
lnParEdu	1.163***	0.840***	0.470*	0.341***	0.168***	0.163*
lnSAT	2.368***	2.719***	1.779***	0.694***	0.544***	0.617***
lnGPA	0.051	0.694***	1.088***	0.015	0.139***	0.377***
lnAPscore	0.227***	0.180***	0.081	0.067***	0.036***	0.028
lnAPclass	0.367***	0.383***	0.248***	0.108***	0.077***	0.086***
Home Sel 10+	0.110**	0.164***	0.162***	0.032**	0.033***	0.056**
Constant	-22.287***	-26.262***	-17.631***			

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a probit estimation of application, attendance, and admission (upon application) on students' demographic, family income, test scores, and the number of selective universities in student's home state. Home Sel 10+ is the dummy variable whether there are more than 10 selective universities in student's home state. The marginal effect is calculated at mean.

Table 12: College Drop Out, Test Scores, and AP class (Probit Estimation)

	Parameter Estimates			Marginal Effect		
Specification	(1)	(2)	(3)	(1)	(2)	(3)
Black	0.320***	0.325***	0.080	0.096***	0.097***	0.023
Asian	0.063	0.094	0.071	0.021	0.028	0.021
Hispanic	-0.051	-0.037	-0.140	-0.017	-0.011	-0.041
Female	-0.014	-0.020	0.035	-0.006	-0.006	0.010
lnMath	-0.127***	-0.134***	-0.057	-0.040***	-0.040***	-0.016
lnfincome	-0.165***	-0.154***	-0.160***	-0.049***	-0.046***	-0.046***
lnMomEdu	-0.301	-0.276	-0.186	-0.091	-0.083	-0.054
lnDadEdu	-0.303*	-0.263	-0.187	-0.089*	-0.079	-0.054
lnAPclass	-0.206***	-0.167***	0.014	-0.062***	-0.050***	0.004
lnUNITMath	-0.463***	-0.436***	-0.287*	-0.138***	-0.130***	-0.083*
Motivation	-0.023	-0.024	-0.037	-0.016	-0.007	-0.011
ActionControl	-0.049	-0.047	0.004	-0.006	-0.014	0.001
$Dummy_{Sel}$		-0.211***	-0.111		-0.063***	-0.032
lnSAT			-0.528***			-0.153**
lnGPA			-1.478***			-0.427***
lnAPscore			-0.095			-0.028
Constant	4.174***	3.875***	8.266***			

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a probit estimation college drop out, OLS of test scores, and OLS of the number of AP classes students take on students' demographic, family income, test scores, and whether the student attended selective universities ($Dummy_{Sel}$).

Table 13: Proportion of Grants out of Tuition

	Selective		Nonselective	
	All	More than half	All	More than half
Black	1.128***	0.812***	1.01	-0.115
Asian	1.037***	0.557***	-0.069	-0.117
Hispanic	1.083***	0.785***	0.783*	0.640*
lnfincome	-0.357***	-0.285***	-0.240	-0.652***
lnParEdu	-0.553	-0.247	-1.219	0.118
lnSAT	0.932**	0.721*	0.551	1.501
lnGPA	1.230***	0.738**	1.375	-0.012
lnAPscore	0.230*	0.114	0.040	0.210
lnAPclass	0.017	-0.008	0.250	-0.143
Constant	-3.573	-2.713	-1.249	-4.261

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a multiple probit estimation regarding the fraction of financial aid a student gets out of total tuition. The baseline outcome is getting less than half of tuition as aid. I estimated separately by college selectivity, thus it does not include information on the financial aid offered by other colleges than the student attended.

Table 14: Test Scores and AP Classes

Parameter	lnSAT	lnGPA	lnAPscore
Black	-0.049***	-0.097***	-0.004
Asian	-0.016***	-0.017**	-0.002
Hispanic	-0.030***	-0.053***	0.007
Female	-0.002	0.069***	0.012
lnMath	0.112***	0.070***	0.060***
lnfincome	0.010***	-0.015***	0.044***
lnParEdu	0.132***	0.013	0.185***
lnAPclass	0.066***	0.054***	0.570***
lnUNITMath	0.031***	0.103***	-0.025
Constant	5.993	0.817***	-1.145***

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a OLS estimation of log test scores on students' demographic, family background, initial math score (lnMaht), and curriculum choice (lnUNITMath: log of the number of total math credit (Carnegie unit), and lnAPclass: log number of AP classes students take).

Table 15: Log Wage Rate, (OLS)

Parameter	(1)	(2)	(3)
Black	-0.079**	-0.081**	-0.029
Asian	-0.007	-0.013	-0.007
Hispanic	0.022	0.016	0.039
Female	-0.063***	-0.062***	-0.065***
lnMath	0.020*	0.020*	0.003
lnfincome	0.081***	0.078***	0.078***
lnMomEdu	-0.013	-0.024	-0.053
lnDadEdu	-0.059	-0.070	-0.095*
lnAPclass	0.069***	0.057***	0.026
lnUNITMath	0.083	0.074	0.046
Motivation	0.019*	0.020*	0.023*
ActionControl	0.002	-0.0004	-0.007
$Dummy_{Sel}$		0.064***	0.042*
lnSAT			0.242***
lnGPA			0.207***
lnAPscore			0.025
Constant	1.912***	2.016***	0.361

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a OLS estimation of log wage rate on students' demographic, family background, initial math score (lnMaht), and curriculum choice (lnUNITMath: log of the number of total math credit (Carnegie unit), and lnAPclass: log number of AP classes students take), college selectivity ($Dummy_{Sel}$), and other traits such as Action Control measure, and Motivation measure (how important monetary return is in your future.

Table 16: Log Wage Rate, (OLS)

Parameter	(1)
Black	-0.070***
Asian	0.010
Hispanic	0.007
Female	-0.043***
lnSAT	0.166***
lnGPA	0.125**
lnAPscore	0.010
$Dummy_{Sel}$	0.044***
Business	0.187***
Social Science	0.030
Engineering	0.264***
Science and Math	0.010
Health	0.282***
Humanity	-0.094***
lnFincome	0.039***
lnParEdu	-0.086
Dropout BA	-0.189***
Graduate School	0.087***
Constant	1.302***

Note. Data come from the the Education Longitudinal Study of 2002 (ELS2002). This is a OLS estimation of log wage rate. I further control major choice, college dropout, and whether to attend graduate school or not.

Table 17: Log Wage Rate (NLSY79)

Parameter	(1)	(2)	(3)	(4)
Black	0.021	0.021	0.030	0.063
Hispanic	0.061***	0.060***	0.110***	0.116***
Female	-0.169***	-0.168***	-0.0004	-0.001
Highest Grade Completed		0.036***	-0.001	-0.002
Age		0.049***	-0.086	
Age Square		-0.00044***	-0.0001	
AFQT	-0.85e-06***	-3.54e-07		-0.000013***
AFQT×Age	3.40e-07***	1.06e-07***		3.85e-07***
lnSAT			-0.0007**	0.63***
lnSAT ×Age			0.000041***	-0.00083
Constant	5.933***	4.781***	6.305***	2.580***
R-square	0.1930	0.2070	0.2321	0.2300

Note. Data come from the National Longitudinal Study of Youth 1979. This is a pooling OLS estimation of log wage rate.

Table 18: Majoring in STEM and AP Classes (Probit)

Parameter	STEM
Black	0.372***
Asian	0.050
Hispanic	0.010
Female	-0.507***
lnMath	0.122***
lnfincome	-0.049
lnMomEdu	0.259
lnDadEdu	0.297
lnAPclass	0.215***
lnUNITMath	1.029***
Motivation	0.047
ActionControl	0.060
Constant	-4.100***

Note. Data come from the Education Longitudinal Study of 2002 (ELS2002). This is a probit estimation choosing STEM major conditional on students' demographic, curriculum choice, test scores, and personality traits.

Table 19: Difference in Salary between 1992 and 2003 (B&B 92)

	OLS	IV
I_{sel}	8521***	14128
female	-9066***	-8819***
lnSAT	7654***	4993
lnGPA	3356**	3058***
Business	3450***	3674***
Education	-7095***	-6326***
Engineering	8402***	8029***
Health	-3234	-3047
Science	-1383	-1728
Math	-1638	-2729
SocialScience	3173***	1986
Constant	-19505	-1629

Note. Data come from the Baccalaureate and Beyond 1992. This is a OLS estimation of the log wage rate on test scores, college selectivity (I_{sel}), and college major.

Table 20: Potential Heterogeneity in Returns from Attending Selective Universities by Family Income (ELS2002)

	lnWage	Graduation Rate
Black	-0.028	0.069
Asian	0.020	-0.026
Hispanic	0.034	0.132
Female	-0.063***	0.008
lnSAT	0.209***	0.887***
lnGPA	0.355***	1.782***
lnAPscore	0.006	0.071
lnfincome	0.053***	0.127***
lnParEdu	-0.069	0.713***
I_{sel}	0.041*	-0.104
$I_{sel} \times D_{Income1Q}$	0.050	-0.134
Constant	0.566	11.154

Note. Data come from the ELS2002. This is a OLS estimation of the log wage rate and a probit estimation of college graduation rate, including the interaction between college selectivity and family income (the dummy variable for the bottom quintile of income distribution)

Table 21: Matching Efficiency (ELS2002)

	lnWage	Graduation Rate
Black	-0.028	0.067
Asian	0.020	-0.025
Hispanic	0.033	0.131
Female	-0.062***	0.009
lnSAT	0.204***	0.878***
lnGPA	0.375***	1.827***
lnAPscore	0.006	0.072
lnfincome	0.052***	0.127***
lnParEdu	-0.069	0.712***
$I_{sel} \times D_{Income1Q}$	0.052	0.139
$I_{sel} \times lnSAT$	0.045	0.111
$I_{sel} \times lnGPA$	-0.222	-0.559
Constant	0.579	11.139***

Note. Data come from the ELS2002. This is a OLS estimation of the log wage rate and a probit estimation of college graduation rate, including the interaction between college selectivity and family income (the dummy variable for the bottom quintile of income distribution), the interaction between college selectivity and test scores.

Appendix C: First Stage Estimation

Table 22: First Stage Estimates - The Amount of Financial Aids

Amount of Financial Aid	Selective Univ.	Nonselective Univ.
$Dummy_{FQ2}$	-5845.9	-2918.7
$Dummy_{FQ3}$	-12366.4	-6583.1
$Dummy_{FQ4}$	-17935.1	-6891.6
$Dummy_{FQ5}$	-27148.2	-10627.9
lnSAT		21899.5
lnAPscore		4192.1
lnGPA		16279.7
Constant	48019.3	-145204.3

Note. Data come from the ELS2002. This is a OLS estimation of the amount of grant a student receives conditional on her family income and test scores. From IPEDS 2004, I match the tuition of each institution. From ELS2002, I observe the fraction of all grants out of total tuition. Based on two variables, I imputed the amount of grants students received, and regress it on family income and test scores. I assume that selective universities provide only based on students' financial aid, whereas nonselective universities have both need-based and merit-based aid.

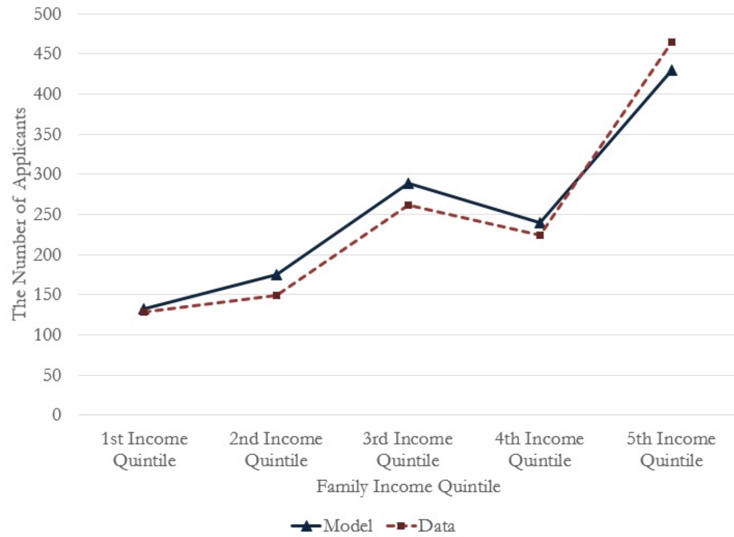
Table 23: First Stage Estimates - The Number of AP Classes Offered by High School/The Number of Information Sources Available to Students

Parameter	Number of AP classes offered by school	Number of Information Sources
Black	0.0278	0.228
Asian	0.482***	0.274*
Hispanic	0.325***	-0.173
Private	-1.456*	0.242***
lnfincome	0.167***	-0.0057
lnParEdu	0.435***	0.285
Private × lnfincome	0.163***	
Private × lnParEdu	-0.238	
Constant	-1.008**	2.858

Note. Data come from the ELS2002. This is a OLS estimation of the number of AP classes offered by high school, and the number of information sources available to students regarding college application process.

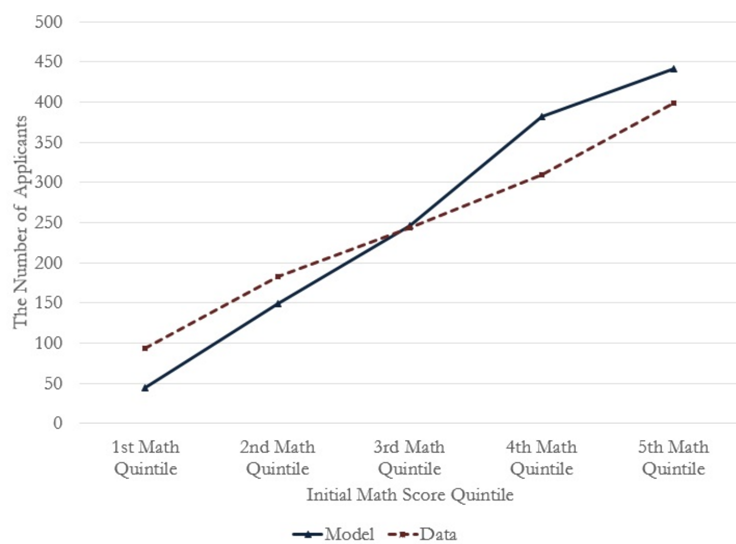
Appendix D: Results (Figures)

Figure 18: The Model Fit of the Number of Applicants for Selective Universities by Family Income Quintile



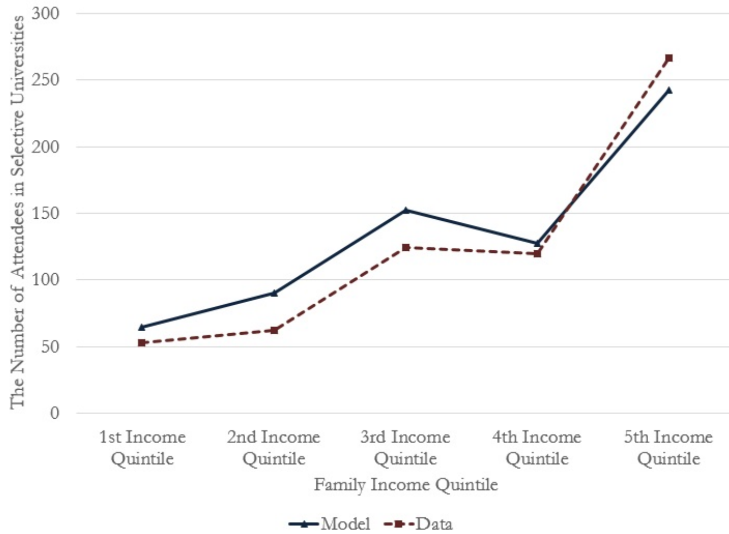
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 19: The Model Fit of the Number of Applicants for Selective Universities by Initial Math Score Quintile



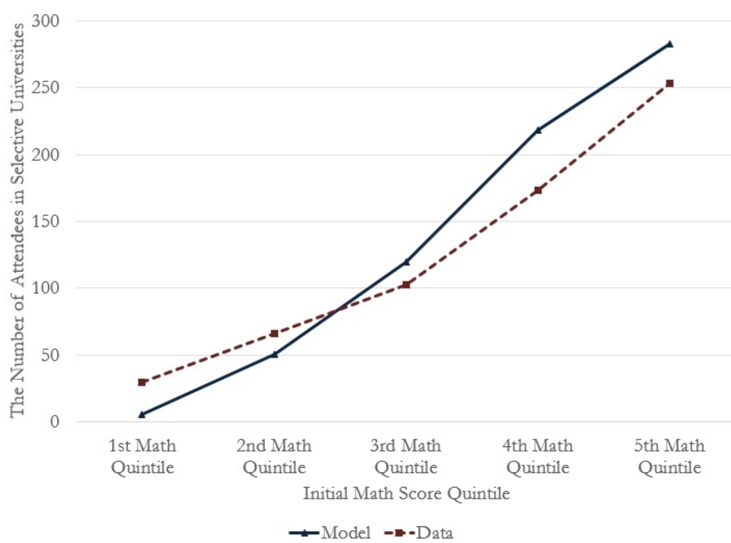
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 20: The Model Fit of the Number of Attendees in Selective Universities by Family Income Quintile



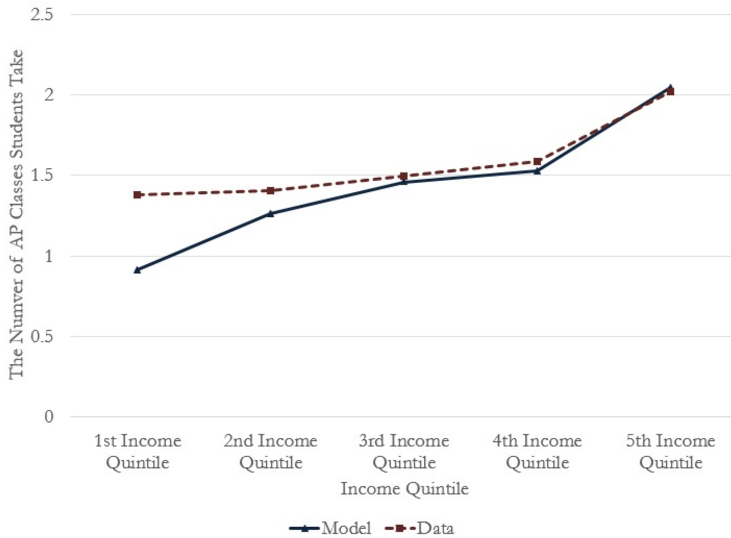
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 21: The Model Fit of the Number of Attendees in Selective Universities by Initial Math Score Quintile



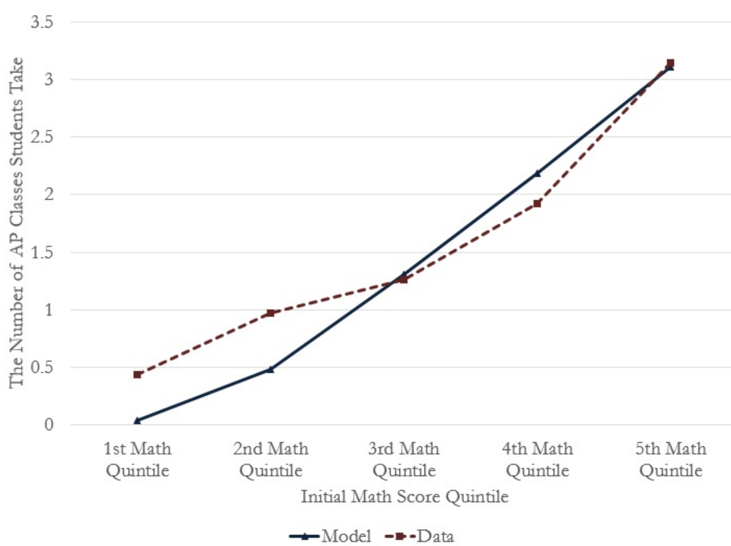
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 22: The Model Fit of the Number of AP Classes Students Take by Family Income Quintile



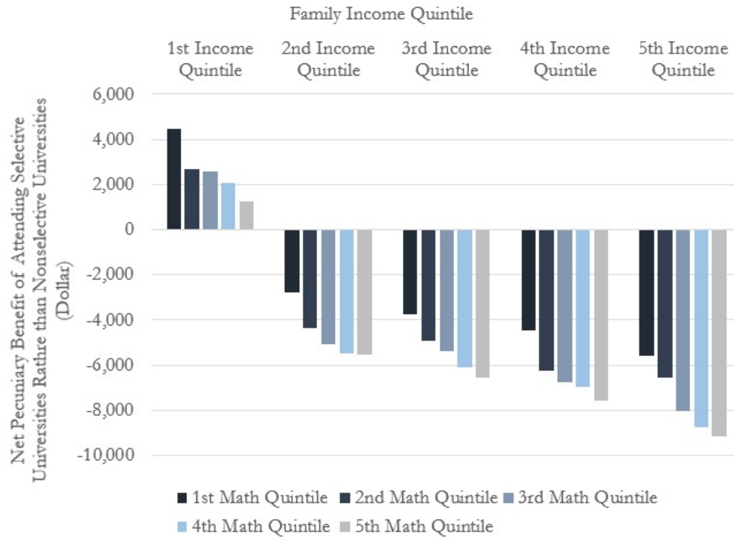
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 23: The Model Fit of the Number of Applicants for Selective Universities by Initial Math Score Quintile



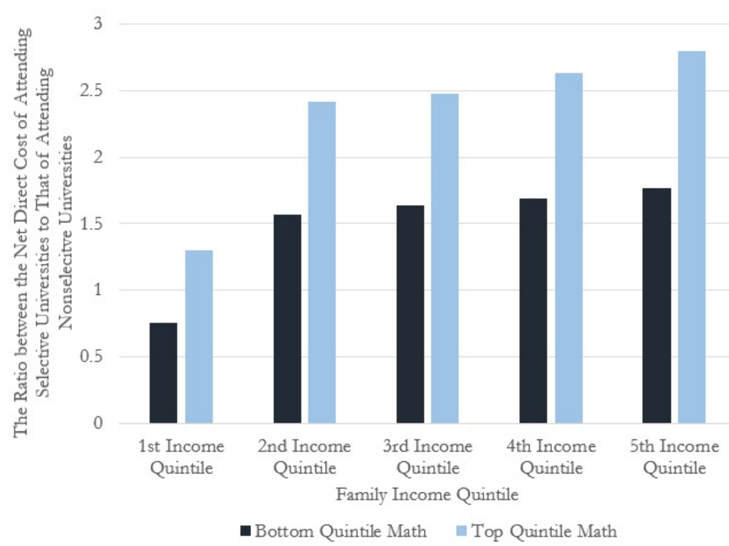
Note. The graph shows the predicted value based on the estimated model and the value observed in data.

Figure 24: The Estimated Net Pecuniary Benefit of Attending Selective Universities Compared to Attending Nonselective universities



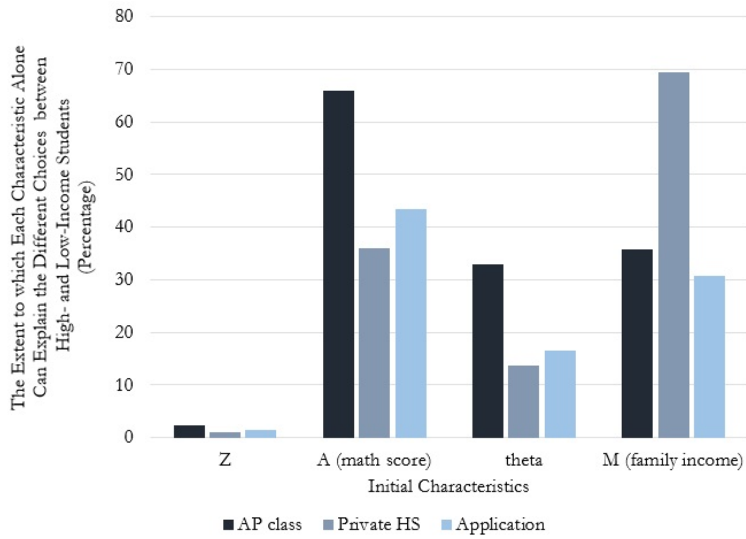
Note. This is the predicted net pecuniary benefit of students from the estimated model. First, I subtract the direct cost of college education from the expected labor market earning conditional on college selectivity. Then I subtract the pecuniary benefit of attending nonselective universities from the pecuniary benefit of attending selective universities.

Figure 25: The Estimated Ratio of the Direct Cost of Attending Selective Universities to Attending Nonselective Universities



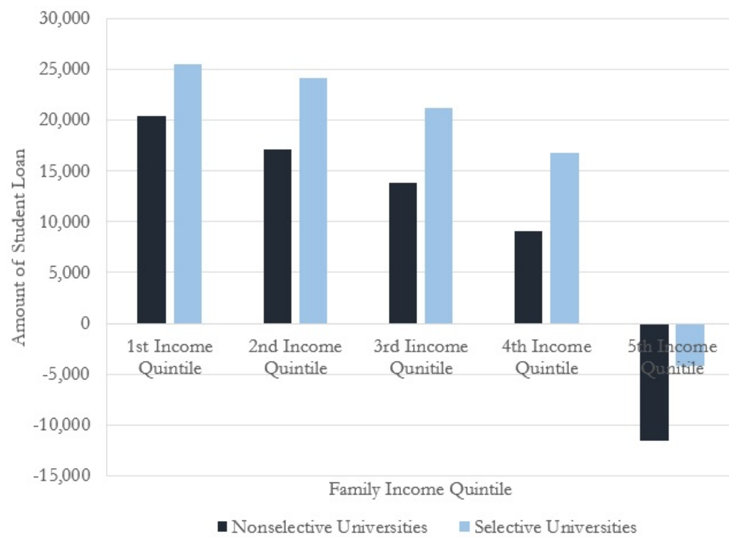
Note. This is the ratio between the direct cost of attending selective universities to that of attending nonselective universities. I compare students with top and bottom quintile of math score distribution.

Figure 26: Percentage Explained By Initial Characteristics Regarding Choice Difference Between Rich and Poor (5th and 1st Income Quintile)



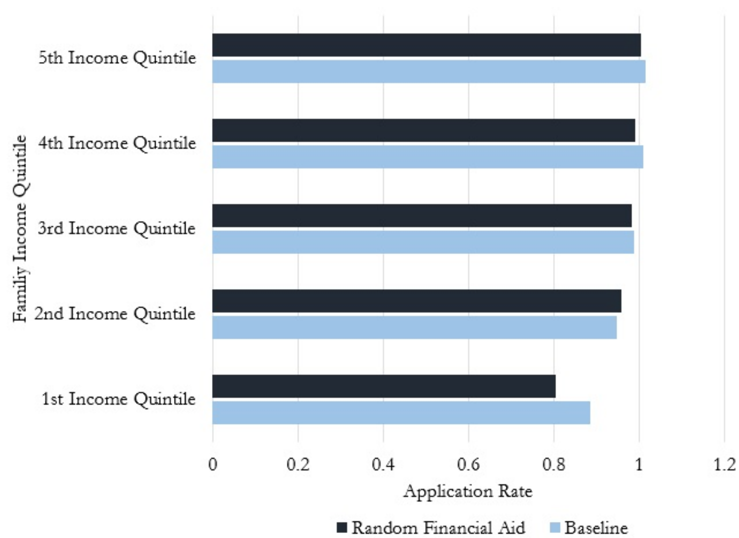
Note. I classify students' characteristics into 4 categories. A: initial math score, θ: unobservable ability, M:family income, Z: all other characteristics including race, sex, parents' education, home state, preference for college characteristics. I obtain the average characteristics of students from the highest quintile and the bottom quintile of income distribution. For each exercise, I consider a student who has the exactly same characteristics as students from the highest income quintile distribution, then replace one characteristic with the one of the bottom quintile of income distribution. The y-axis implies to what extent difference in one type of characteristics between rich and poor students can explain the estimated true choice difference between rich and poor students (percent).

Figure 27: Amount of Student Loan by Family Income when Attending Selective/Nonselective Universities



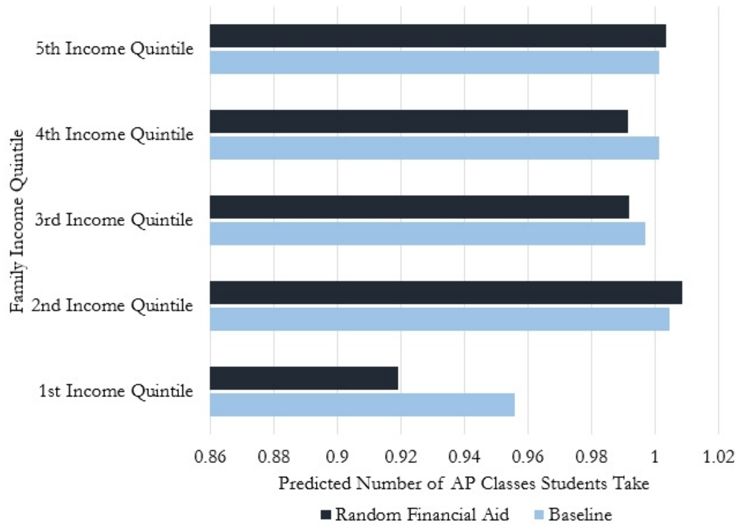
Note. This is the predicted value of the amount of student loan conditional on the highest (rich) and the lowest (poor)quintile of income distribution.

Figure 28: Liquidity Constraint and the Number of Applicants



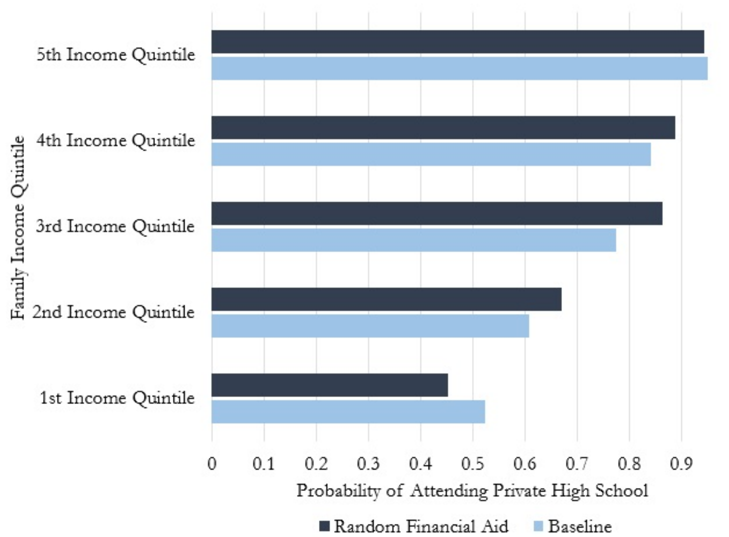
Note. I compare the number of applicants for selective universities from each quintile of income distribution with and without liquidity constraint.

Figure 29: Liquidity Constraint and the AP classes



Note. I compare the number of AP classes students take of students from each quintile of income distribution with and without liquidity constraint.

Figure 30: Liquidity Constraint and the Probability of Attending Private High School



Note. I compare the probability of attending private high school of students from each quintile of income distribution with and without liquidity constraint.

Appendix E: Parameter Estimates

Table 24: Log Wage Rate

Parameter	ln(Wage)
Black	-0.0350 (0.1988)
Asian	0.0300 (0.1992)
Hispanic	0.0276 (0.2384)
Female	-0.0342 (0.0366)
lnSAT	0.1305 (0.0169)
lnAP	0.0625 (0.1415)
lnGPA	0.1064 (0.0268)
I_{sel}	0.0759 (0.0360)
I_{drop}	-0.0823 (0.0185)
θ	1
Const	1.2982 (0.2107)

Note. The coefficient of unobservable ability θ is normalized to one so that it increase the wage rate one-to-one. The distribution of θ is estimated and documented in Table 30.

Table 25: Log Test Scores

Parameter	ln(SAT)	ln(AP+1)	ln(GPA)
Black	-0.0135 (0.4411)	0.0021 (0.1532)	-0.0400 (0.1128)
Asian	0.0073 (0.4605)	0.0023 (0.2156)	-0.0103 (0.1370)
Hispanic	0.0199 (0.5334)	0.0153 (0.1089)	-0.0408 (0.1639)
Female	0.0137 (0.2175)	0.0329 (0.0440)	0.0098 (0.0571)
lnA_{math}	0.1933 (0.0203)	0.0629 (0.0191)	0.0798 (0.0130)
θ	0.3150 (3.8265)	0.1271 (0.0848)	-0.1631 (0.7949)
I_{public}	-0.0367 (0.1152)	-0.0455 (0.1128)	0.2272 (0.0530)
$ln(APclass + 1)$	0.0973 (0.0668)	0.2763 (0.0238)	0.1456 (0.0336)
Const	6.0568 (1.9464)	0.0218 (0.5128)	0.6959 (0.3787)

Note. Since almost half of students did not take AP tests, I use $ln(AP + 1)$ instead of $ln(AP)$.

Table 26: Admission Probability

Parameter	Admission Probability
Black	0.4413 (2.5617)
Asian	0.2485 (0.8593)
Hispanic	0.4058 (1.1169)
lnSAT	2.1445 (0.3609)
ln(AP+1)	0.1161 (1.2017)
lnGPA	0.9760 (0.5326)
ln(APclass+1)	0.0427 (0.8281)
θ	0.0980 (4.4716)
Const	-16.7763 (1.3818)

Note. Since almost half of students did not take AP tests, I use $ln(AP+1)$ instead of $ln(AP)$.

Table 27: Graduation Rate from the College

Parameter	Graduation Probability
Female	0.2509 (1.4902)
I_{sel}	0.1346 (7.9805)
θ	2.5476 (9.7945)
lnSAT	0.3018 (0.1562)
ln(AP+1)	0.3331 (1.3538)
lnGPA	0.2847 (0.7397)
Const	-3.2515 (8.1200)

Note. Since almost half of students did not take AP tests, I use $ln(AP+1)$ instead of $ln(AP)$.

Table 28: Nonpecuniary Benefit of Attending Selective Universities

Parameter	Nonpecuniary Benefit
lnA_{math}	0.3236 (0.7989)
θ	2.1578 (6.3472)
$ln(Num_{sel,home})$	1.2983 (0.8621)
$Imp_{Reputation}$	1.5759 (0.3876)
$Imp_{Location}$	-0.1060 (0.9757)
$Imp_{Same_College_asparent}$	0.8345 (7.0267)
Const	-0.6247 (7.1526)

Note. $ln(Num_{sel,hom})$ indicates the log number of selective universities in student's home state. $Imp_{Repuation}$ is student's reported preference at the sophomore year of the high school over the importance of college's reputation in her college choice with scale 1 to 3. Similarly, $Imp_{Location}$ and $Imp_{Same_College_asparent}$ are the reported preference over college's location and whether it is the same college as one of parents in student's college choice.

Table 29: Application Cost

Parameter	Application Cost	Effort Cost of Taking AP class
lnA_{math}	0.222 (1.3641)	-1.6313 (0.6952)
θ	0.1769 (2.8192)	-5.0512 (2.4755)
I_{public}	-0.7225 (0.4556)	
$Num_{InformationSource}$	0.7898 (0.4149)	
$Num_{APclass}$		-0.5420 (0.5573)
Const	-5.7963 (2.7308)	9.6778 (1.3310)

Note. $Num_{InformationSource}$ is the log number of information sources available to students regarding college application process ($ln(Num_{sources} + 1)$). $Num_{APcalss}$ is the number of AP classes offered by the high school.

Table 30: Other Parameters

Parameter	Other Parameters
β (Intertemporal Preference)	0.2562 (0.0381)
λ (Liquidity Constraint)	0.7608 (0.2520)
ψ_1 (Learning Ability $\psi_{fincome}$)	0.0282 (0.015)
ψ_2 (Learning Ability $\psi_{ParentEdu}$)	0.0599 (0.0340)
σ_θ (Standard Error of Learning Ability θ)	0.1301 (0.0747)
σ_{wage}	0.4973 (0.0393)
σ_{lnSAT}	0.3495 (0.1724)
σ_{lnAP}	0.3024 (0.1098)
σ_{lnGPA}	0.3816 (0.0859)
$\sigma_{\epsilon_{AP}}$	1.4338 (1.0745)

Note. β captures the intertemporal preference of consumption between college and working period. λ captures the extent of liquidity constraint such that students cannot borrow up to λ fraction of her expected future income. ψ_1 and ψ_2 are the conditional mean of distribution of the unobservable ability θ. σ_j $j \in \{wage, lnSAT, lnAP, lnGPA\}$ are the standard error of the i.i.d. error shock on the wage rate, test scores. $\sigma_{\epsilon_{AP}}$ is the standard error of i.i.d. component in the effort cost of taking AP classes. σ_θ is the standard error of the distribution of θ.

Appendix F: Estimation Result

Table 31: Estimated Income of Graduates from Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	29,063	30,095	31,464	31,187	33,348
Math2Q	31,644	32,235	33,560	33,512	34,550
Math3Q	33,536	33,129	33,901	34,159	36,050
Math4Q	33,357	34,333	35,463	35,201	36,859
Math5Q	34,479	36,355	37,185	36,934	38,032

Note. Predicted labor income if students attend nonselective universities

Table 32: Estimated Income of Graduates from Selective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	31,354	32,269	33,946	33,647	35,978
Math2Q	34,140	34,777	36,206	36,155	37,274
Math3Q	36,180	35,741	36,574	36,853	38,893
Math4Q	35,988	37,040	38,260	37,977	39,765
Math5Q	37.198	39,222	40,117	39,847	41,031

Note. Predicted labor income if students attend selective universities

Table 33: Cost Difference between Graduates of Selective Universities and Graduates of Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	5,014	4,790	5,670	7,393	8,077
Math2Q	7,118	6,691	6,818	9,535	9,247
Math3Q	6,080	8,976	7,347	9,617	10,088
Math4Q	8,577	8,620	8,783	9,211	11,557
Math5Q	8,667	8,521	8,733	10,865	11,008

Note. Predicted difference of annual cost of attending selective universities and attending nonselective universities

Table 34: Net Pecuniary Benefit from Attending Selective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	-2,723	-2,416	-3,188	-4,933	-5,447
Math2Q	-4,622	-4,149	-4,172	-6,892	-6,523
Math3Q	-3,436	-6,364	-4,674	-6,923	-7,245
Math4Q	-5,946	-5,913	-5,986	-6,435	-8,651
Math5Q	-5,948	-5,654	-5,801	-7,953	-8,009

Note. I calculate the pecuniary benefit of attending each type of college by subtracting annual cost from annual labor income. Then I subtract the pecuniary benefit of attending nonselective universities from that of selective universities.

Table 35: Amount of Student Loan When Attending Selective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	23,434	18,393	15,472	9,667	-8,670
Math2Q	24,143	19,708	16,174	11,245	-6,045
Math3Q	23,908	20,515	17,157	12,391	-5,967
Math4Q	26,001	22,292	17,509	14,090	-5,207
Math5Q	25,614	23,270	18,236	13,671	-5,923

Note. Predicted loan amount if the student attends selective universities.

Table 36: Amount of Student Loan When Attending Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	18,493	15,161	12,118	6,767	-12,657
Math2Q	17,974	16,734	12,314	7,339	-10,114
Math3Q	19,481	17,092	13,747	8,706	-10,432
Math4Q	20,551	17,924	13,853	9,904	-9,572
Math5Q	22,136	19,268	14,344	9,386	-10,617

Note. Predicted loan amount if the student attends nonselective universities.

Appendix G: Counterfactual Analysis

Table 37: Counterfactual Analysis (1): Liquidity Constraint and Application Decision

True \No Constraint	NonSel,Drop	NonSel,BA	Sel,Drop	Sel,BA
NonSel,Drop	629	0	3	1
NonSel,BA	4	1766	0	11
Sel,Drop	0	0	114	1
Sel,BA	0	1	1	552

Note. I consider four discrete choices of college education. I simulate the model with and without liquidity constraint, ant the number in element (i, j) indicates the number of students who chose i in the equilibrium, whereas the optimal choice without liquidity constraint is j.

Table 38: Counterfactual Analysis (2): Liquidity Constraint and Application Decision when Financial Aid from Selective Universities are Random

True No Constraint	NonSel,Drop	NonSel,BA	Sel,Drop	Sel,BA
NonSel,Drop	630	0	5	1
NonSel,BA	4	1749	0	22
Sel,Drop	0	0	112	1
Sel,BA	0	3	0	556

Note. I consider four discrete choices of college education. To evaluate the impact of current need-based aid from selective universities, I assume that the financial aid policy from selective universities is randomly allocated. Then I simulate the model with and without liquidity constraint. The number in element (i, j) indicates the number of students who chose i in the equilibrium, whereas the optimal choice without liquidity constraint is j.

Table 39: Counterfactual Analysis (3): Liquidity Constraint and the Number of Applicants

	Baseline		Without Need Based Financial Aid	
	True	No Constraint	True	No Constraint
1st Income Quintile	122 (88%)	138	106 (80%)	132
2nd Income Quintile	160 (95%)	169	160 (96%)	167
3rd Income Quintile	299 (99%)	303	304 (98%)	309
4th Income Quintile	234 (101%)	232	237 (99%)	239
5th Income Quintile	464 (102%)	457	476 (100%)	474

Note. I simulate the model with and without liquidity constraint and compare the number of applicants from each quintile of income distribution. Then I assume that the financial aid policy from selective universities is completely random, and compare how the number of applicants changes with and without liquidity constraint.

Table 40: Counterfactual Analysis (4): Liquidity Constraint and AP Classes (Baseline vs. Random Financial Aid)

	Baseline		Without Need Based Financial Aid	
	True	No Constraint	True	No Constraint
Income1Q	1.128 (96%)	1.180	1.000 (92%)	1.088
Income2Q	1.578 (100%)	1.571	1.557 (101%)	1.544
Income3Q	2.052 (100%)	2.058	2.050 (99%)	2.067
Income4Q	2.183 (100%)	2.180	2.183 (99%)	2.202
Income5Q	3.089 (100%)	3.085	3.134 (100%)	3.123

Note. I simulate the model with and without liquidity constraint and compare the number of AP classes students take from each quintile of income distribution. Then I assume that the financial aid policy from selective universities is completely random, and compare how the number of AP classes taken by students changes with and without liquidity constraint.

Table 41: Counterfactual Analysis (5): Liquidity Constraint and Probability of Attending Private High School (Baseline vs. Random Financial Aid)

	Baseline		Without Need Based Financial Aid	
	True	No Constraint	Random Aid	Random Aid without Constraint
Income1Q	15.8 (52%)	30.2	13.0 (45%)	28.8
Income2Q	18.7 (61%)	30.7	23.9 (67%)	35.7
Income3Q	27.0 (77%)	34.9	32.1 (86%)	37.2
Income4Q	33.3 (84%)	39.6	40.6 (89%)	45.7
Income5Q	54.5 (95%)	57.4	48.4 (94%)	51.3

Note. I simulate the model with and without liquidity constraint and compare the probability of attending private high school from each quintile of income distribution. Then I assume that the financial aid policy from selective universities is completely random, and compare how the probability of attending private high school changes with and without liquidity constraint.

Table 42: Choice Difference between the Highest and the Lowest Income Quintile Groups Explained by Group Different in Initial Characteristic

Percentage Change	AP classes	Private HS	Application Rate
$\Delta(\tilde{Z})$	2.2	0.9	1.4
$\Delta(A)$	65.9	35.9	43.4
$\Delta(\theta)$	32.9	13.6	16.6
$\Delta(M)$	35.8	69.5	30.8

Note. I classify students' characteristics into 4 categories. A: initial math score, θ: unobservable ability, M:family income, Z: all other characteristics including race, sex, parents' education, home state, preference for college characteristics. I obtain the average characteristics of students from the highest quintile and the bottom quintile of income distribution. For each exercise, I consider a student who has the exactly same characteristics as students from the highest income quintile distribution, then replace one characteristic with the one of the bottom quintile of income distribution. The y-axis implies to what extent difference in one type of characteristics between rich and poor students can explain the estimated true choice difference between rich and poor students (percent).

Table 43: Percentage Changes in the Number of AP Classes Conditional on Family Income and Initial Math Score Quintile

Number of AP Classes	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	7.9	12.7	10.7	9.8	6.3
Ability2Q	6.3	5.5	5.6	6.5	6.1
Ability3Q	3.4	3.8	3.7	3.7	3.7
Ability4Q	2.2	2.3	2.2	2.6	2.1
Ability5Q	0.9	1.1	1.2	1.4	0.1

Note. The unit is percentage change in the main counterfactual analysis compared to the baseline model. The main counterfactual analysis is the one that increases need-based aid from selective universities by $10,000 without changing tuition and aid to other income groups. The admission cutoff is determined in equilibrium to equalize the number of attendees and the number of seats available in selective universities.

Table 44: Counterfactual Analysis: Features of Selective Universities and Choice Difference between Rich and Poor

Lowest Income Quintile	AP class	Application	Private High School
Baseline	0.36 (100%)	13.43 (100%)	12.03 (100%)
Random Admission	0.05 (12.9%)	40.42 (301.0%)	31.92 (265.3%)
Tuition as Nonselective Univ	0.41 (116.3%)	15.44 (115.0%)	12.42 (103.2%)
Random Financial Aid	0.33 (91.9%)	12.20 (90.8%)	11.42 (94.9%)
No Nonpecuniary Benefit	0.05 (15.0%)	4.80 (35.7%)	11.39 (94.7%)
Highest Income Quintile	AP class	Application	Private High School
Baseline	2.07 (100%)	42.10 (100%)	44.70 (100%)
Random Admission	0.08 (3.8%)	59.13 (140.5%)	61.40 (137.4%)
Tuition as Nonselective Univ	2.06 (99.7%)	40.77 (96.6%)	43.17 (99.7%)
Random Financial Aid	2.13 (102.7%)	43.66 (103.7%)	46.22 (103.4%)
No Nonpecuniary Benefit	0.09 (4.4%)	6.0 (14.3%)	12.49 (27.9%)
Difference between Two Groups	AP class	Application	Private High School
Baseline	1.71 (100%)	28.67 (100%)	32.67 (100%)
Random Admission	0.03 (1.9%)	18.71 (65.3%)	29.48 (90.2%)
Tuition as Nonselective Univ	1.65 (96.2%)	25.33 (88.4%)	30.75 (94.1%)
Random Financial Aid	1.80 (104.9%)	31.46 (109.7%)	34.80 (106.5%)
No Nonpecuniary Benefit	0.04 (2.3%)	1.20 (4.2%)	1.10 (3.4%)

Note. I document how choices of students change if a certain feature of selective universities changes. In particular, I compare the choices of baseline model with those in each counterfactual exercise: (i) replacing the admission criteria with a random one, (ii) replacing the tuition of selective universities with those of nonselective universities, (iii) replacing need-based aid from selective universities into a random allocation, and (iv) if there is no nonpecuniary benefit from attending selective universities.

Table 45: Counterfactual Analysis: Features of Selective Universities and Outcome Difference between Rich and Poor

Lowest Income Quintile	Attendance Rate	SAT	lnWage	Graduation Rate
Baseline	3.60 (100%)	937.42 (100%)	2.73 (100%)	62.41 (100%)
Random Admission	16.77 (465.8%)	936.02 (99.8%)	2.73 (100%)	61.86 (101.5%)
Tuition as Nonselective Univ	4.18 (115.0%)	939.21 (100.2%)	2.73 (100%)	62.37 (100.1%)
Random Financial Aid	3.34 (92.8%)	936.77 (99.9%)	2.73 (100%)	62.28 (100.3%)
No Nonpecuniary Benefit	4.80 (133.3%)	929.20 (99.1%)	2.73 (100%)	61.99 (99.3%)
Highest Income Quintile	Attendance Rate	SAT	lnWage	Graduation Rate
Baseline	18.85 (100%)	1119.8 (100%)	2.86 (100%)	69.61 (100%)
Random Admission	24.16 (128.2%)	1064.3 (95.04%)	2.83 (99.1%)	67.95 (97.6%)
Tuition as Nonselective Univ	17.48 (92.7%)	1118.12 (99.9%)	2.86 (100%)	69.64 (100%)
Random Financial Aid	20.08 (106.5%)	1122.61 (100.3%)	2.86 (101.8%)	69.67 (102.6%)
No Nonpecuniary Benefit	6.00 (31.8%)	1045.91 (93.4%)	2.83 (98.9%)	68.67 (98.6%)
Difference between Two Groups	Attendance Rate	SAT	lnWage	Graduation Rate
Baseline	15.25 (100%)	182.0 (100%)	0.125 (100%)	7.20 (100%)
Random Admission	7.39(48.5%)	128.27 (70.5%)	0.096 (77.1%)	6.09 (84.6%)
Tuition as Nonselective Univ	13.30(87.2%)	178.92 (98.3%)	0.12 (98.2%)	7.27 (101.0%)
Random Financial Aid	16.74(109.8%)	185.84 (102.1%)	0.13 (101.8%)	7.39 (102.6%)
No Nonpecuniary Benefit	1.20 (7.9%)	116.71 (64.1%)	0.10 (0.78%)	6.68 (92.8%)

Note. I document how academic achievement, graduation rate, and wage rate change if a certain feature of selective universities changes. In particular, I compare the choices of baseline model with those in each counterfactual exercise: (i) replacing the admission criteria with a random one, (ii) replacing the tuition of selective universities with those of nonselective universities, (iii) replacing need-based aid from selective universities into a random allocation, and (iv) if there is no nonpecuniary benefit from attending selective universities.

Table 46: Counterfactual Analysis: The Impact of Redistributive Policies

Additional Aid	Baseline	\$10,000	\$15,000	\$20,000	\$25,000	\$30,000	\$35,000
Tuition Increase	Baseline	\$750	\$800	\$1,300	\$1,375	\$1,600	\$1,700
Bottom Quintile							
Composition Rate (FQ1)	9.54	10.76	11.15	11.42	11.59	11.71	11.79
AP class (FQ1)	1.18	11.24%	15.30%	18.00%	19.83%	20.87%	21.62%
AP Attendees (FQ1)	5.19	0.06%	1.19%	1.52%	1.75%	1.80%	1.85%
Other Income Groups							
AP class (FQ2-FQ5)	2.23	-0.71%	-0.96%	-1.11%	-1.26%	-1.38%	-1.43%
AP Attendees (FQ2-FQ5)	6.28	0.33%	0.43%	0.47%	0.50%	0.56%	0.56%
All							
AP class (FQ1-FQ5)	2.07	0.27%	0.38%	0.43%	0.47%	0.45%	0.45%
AP Attendees (FQ1-FQ5)	6.17	0.14%	0.22%	0.24%	0.26%	0.30%	0.29%

Note. In this counterfactual analysis, I increase need-based aid from selective universities to attending students from the bottom quintile of income distribution from \$10,000 to \$35,000, while raising tuition accordingly in order to keep the total tuition revenue in selective universities the same.

Table 47: The number of AP classes

Baseline	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0.0	0	0.04	0	0.14
Math2Q	0.56	0.58	0.75	0.87	1.03
Math3Q	1.21	1.13	1.31	1.58	2.58
Math4Q	2.49	2.37	3.05	3.04	4.13
Math5Q	2.80	3.91	4.15	4.41	4.95
Need Based Aid from Sel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0.0	0	0.04 (100%)	0	0.14 (98%)
Math2Q	0.63 (113%)	0.58 (100%)	0.70 (94%)	0.87 (100%)	1.01 (100%)
Math3Q	1.47 (121%)	1.12 (99%)	1.31 (100%)	1.59 (100%)	2.58 (100%)
Math4Q	2.74 (110%)	2.35 (99%)	3.04 (100%)	2.98 (98%)	4.11 (100%)
Math5Q	3.22 (115%)	3.91 (100%)	4.15 (100%)	4.42 (100%)	4.94 (100%)
Need Based Aid from NonSel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0	0	0.04 (100%)	0	0.15 (109%)
Math2Q	0.43 (77%)	0.58 (100%)	0.75 (101%)	0.88 (101%)	1.09 (141%)
Math3Q	1.12 (93%)	1.22 (108%)	1.32 (101%)	1.58 (100%)	2.58 (106%)
Math4Q	2.18 (88%)	2.36 (100%)	3.09 (101%)	3.02 (100%)	4.12 (118%)
Math5Q	2.62 (93%)	3.93 (100%)	4.14 (100%)	4.43 (100%)	5.02 (112%)
Income Quota	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0.01	0	0.03 (75%)	0	0.14 (100%)
Math2Q	0.71 (127%)	0.51 (87.5%)	0.70 (94%)	0.82 (94%)	0.84 (82%)
Math3Q	1.48 (122%)	1.14 (101%)	1.25 (95%)	1.43 (91%)	2.42 (94%)
Math4Q	2.25 (90%)	2.25 (95.0%)	2.95 (97%)	2.84 (94%)	4.03 (98%)
Math5Q	2.63 (94%)	4.00 (102.2%)	4.15 (100%)	4.38 (99%)	4.87 (98%)
Texas Top 10	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0	0	0	0.01	0.04 (27%)
Math2Q	0.32 (56%)	0.23 (39%)	0.30 (40%)	0.57 (66%)	0.39 (38%)
Math3Q	0.59 (49%)	0.34 (30%)	0.41 (31%)	0.57 (36%)	1.15 (45%)
Math4Q	1.12 (45%)	1.12 (47%)	1.43 (47%)	1.40 (46%)	1.70 (41%)
Math5Q	1.55 (55%)	2.43 (62%)	2.01 (48%)	2.59 (59%)	2.47 (50%)
Offering more AP classes	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	0	0	0.04	0	0.14
Math2Q	0.63 (113%)	0.58 (100%)	0.75 (100%)	0.87 (100%)	1.03 (100%)
Math3Q	1.32 (109%)	1.13 (100%)	1.31 (100%)	1.58 (100%)	2.58 (100%)
Math4Q	2.63 (106%)	2.37 (100%)	3.05 (100%)	3.03 (100%)	4.12 (100%)
Math5Q	2.93 (105%)	3.91 (100%)	4.15 (100%)	4.41 (100%)	4.95 (100%)

Note. The table shows changes in the number of AP classes students take in each counterfactual policy change. I consider (i) increasing need-based aid by $10,000 per attending students from selective universities without changing tuition and financial aid policy for students from other income backgrounds, (ii) the same policy implemented by *nonselective universities*, (iii) affirmative action based on income (income quota), (iv) Texas Top 10 Law, and (v) offering more AP classes to low-income students.

Table 48: The Application Rate

Baseline	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	5.08	5.41	7.78	5.62	12.35
Math2Q	19.82	23.39	28.57	25.00	30.26
Math3Q	26.67	34.88	38.19	39.06	48.13
Math4Q	56.92	43.64	53.19	50.35	69.90
Math5Q	55.00	59.76	65.03	65.35	74.68
Need Based Aid from Sel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	5.08 (100%)	5.41 (100%)	7.78 (100%)	5.62 (100%)	12.35 (100%)
Math2Q	21.62 (109%)	23.39 (100%)	27.92 (98%)	25.00 (100%)	29.61 (98%)
Math3Q	32.22 (121%)	33.72 (97%)	38.19 (100%)	39.06 (100%)	47.59 (99%)
Math4Q	58.46 (103%)	42.73 (98%)	52.66 (99%)	48.94 (97%)	69.42 (99%)
Math5Q	66.67 (121%)	59.76 (100%)	65.03 (100%)	65.35 (100%)	74.25 (99%)
Need Based Aid from NonSel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	5.08 (100%)	5.41 (100%)	7.78 (100%)	5.62 (100%)	13.58 (110%)
Math2Q	17.12 (86%)	23.39 (100%)	28.57 (100%)	25.00 (100%)	30.92 (102%)
Math3Q	26.67 (100%)	36.05 (102%)	38.89 (102%)	39.06 (100%)	48.13 (100%)
Math4Q	53.85 (95%)	43.64 (104%)	55.32 (104%)	50.35 (100%)	70.87 (101%)
Math5Q	53.33 (97%)	59.76 (101%)	65.73 (101%)	66.14 (101%)	75.97 (102%)
Income Quota	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	11.86 (233%)	4.05 (75%)	5.56 (71%)	5.62 (100%)	11.11 (90%)
Math2Q	42.34 (214%)	20.97 (90%)	25.97 (91%)	22.00 (88%)	23.68 (78%)
Math3Q	51.11 (192%)	31.40 (90%)	34.72 (90%)	35.94 (92%)	43.32 (90%)
Math4Q	67.69 (119%)	40.00 (92%)	49.47 (93%)	43.97 (87%)	65.53 (94%)
Math5Q	70.00 (127%)	59.76 (100%)	62.94 (97%)	61.42 (94%)	71.24 (95%)
Texas Top 10	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	25.42 (500%)	25.68 (475%)	34.44 (443%)	29.21 (520%)	37.04 (300%)
Math2Q	41.44 (209%)	45.97 (197%)	44.16 (155%)	48.00 (192%)	44.74 (148%)
Math3Q	34.44 (129%)	44.19 (127%)	45.14 (118%)	45.31 (116%)	55.61 (116%)
Math4Q	50.77 (89%)	49.09 (112%)	54.79 (103%)	49.65 (99%)	64.08 (92%)
Math5Q	51.67 (94%)	57.32 (96%)	58.27 (90%)	58.27 (89%)	65.67 (88%)
Offering more AP classes	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	5.08 (100%)	5.41 (100%)	7.78 (100%)	5.62 (100%)	12.35 (100%)
Math2Q	20.72 (105%)	23.39 (100%)	28.57 (100%)	25.00 (100%)	30.26(100%)
Math3Q	26.67 (100%)	34.88 (100%)	38.19 (100%)	39.06 (100%)	48.13 (100%)
Math4Q	56.92 (100%)	43.64 (100%)	53.19 (100%)	50.35 (100%)	69.90 (100%)
Math5Q	53.33 (97%)	59.76 (100%)	65.03 (100%)	65.35 (100%)	74.68 (100%)

Note. The table shows changes in the application rate (the number of applicants/all students conditional on income and math score quintile) in each counterfactual policy change. I consider (i) increasing need-based aid by $10,000 per attending students from selective universities without changing tuition and financial aid policy for students from other income backgrounds, (ii) the same policy implemented by *nonselective universities*, (iii) affirmative action based on income (income quota), (iv) Texas Top 10 Law, and (v) offering more AP classes to low-income students.

Table 49: The Attendance Rate into Private High School

Baseline	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	8.47	10.81	13.33	15.73	18.52
Math2Q	21.62	15.32	24.03	26.00	30.92
Math3Q	12.22	20.93	25.00	33.59	51.34
Math4Q	21.54	25.45	29.79	33.33	69.42
Math5Q	18.33	19.51	37.06	51.18	71.67
Need Based Aid from Sel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	8.47 (100%)	10.81 (100%)	13.33 (100%)	15.73 (100%)	18.52 (100%)
Math2Q	20.72 (96%)	15.32 (100%)	24.03 (100%)	26.00 (100%)	30.92 (110%)
Math3Q	12.22 (100%)	20.93 (100%)	25.00 (100%)	32.81 (98%)	50.80 (99%)
Math4Q	20.00 (93%)	25.45 (100%)	29.26 (98%)	32.62 (98%)	68.93 (102%)
Math5Q	18.33 (100%)	19.51 (100%)	37.06 (100%)	51.18 (100%)	71.24 (99%)
Need Based Aid from NonSel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	8.47 (100%)	10.81 (100%)	13.33 (100%)	15.73 (100%)	18.52 (100%)
Math2Q	21.62 (100%)	15.32 (100%)	24.03 (100%)	26.00 (100%)	30.92 (100%)
Math3Q	13.33 (109%)	20.93 (100%)	25.00 (100%)	33.59 (100%)	51.34 (97%)
Math4Q	23.08 (107%)	25.45 (100%)	31.91 (107%)	33.33 (100%)	70.39 (99%)
Math5Q	23.33 (127%)	18.29 (94%)	37.76 (102%)	51.97 (102%)	72.96 (93%)
Income Quota	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	10.17 (120%)	10.81 (100%)	12.22 (92%)	15.73 (100%)	17.28 (93%)
Math2Q	26.13 (121%)	16.13 (105%)	22.08 (92%)	24.00 (92%)	25.00 (81%)
Math3Q	18.89 (155%)	19.77 (94%)	23.61 (94%)	30.47 (91%)	45.99 (90%)
Math4Q	27.69 (129%)	23.64 (93%)	27.66 (93%)	28.37 (85%)	66.02 (95%)
Math5Q	21.67 (118%)	19.51 (100%)	19.51 (96%)	47.24 (92%)	68.24 (95%)
Texas Top 10	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	10.17 (120%)	12.16 (112%)	14.44 (108%)	21.35 (135%)	38.27 (207%)
Math2Q	18.02 (83%)	14.52 (95%)	23.38 (97%)	26.00 (100%)	30.92 (100%)
Math3Q	11.11 (91%)	17.44 (83%)	18.75 (75%)	21.88 (65%)	33.16 (65%)
Math4Q	13.85 (64%)	18.18 (71%)	18.62 (63%)	15.60 (47%)	37.38 (54%)
Math5Q	6.67 (36%)	9.76 (50%)	18.18 (49%)	15.75 (31%)	34.76 (49%)
Offering more AP classes	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	8.47 (100%)	10.81 (100%)	13.33 (100%)	15.73 (100%)	18.52 (100%)
Math2Q	22.52 (104%)	15.32 (100%)	24.03 (100%)	26.00 (100%)	30.92 (100%)
Math3Q	12.22 (100%)	20.93 (100%)	25.00 (100%)	33.59 (100%)	51.34 (100%)
Math4Q	23.08 (107%)	25.45 (100%)	29.79 (100%)	33.33 (100%)	69.42 (100%)
Math5Q	20.00 (109%)	19.51 (100%)	37.06 (100%)	51.18 (100%)	71.67 (100%)

Note. The table shows changes in the probability of attending private high school in each counterfactual policy change. I consider (i) increasing need-based aid by $10,000 per attending students from selective universities without changing tuition and financial aid policy for students from other income backgrounds, (ii) the same policy implemented by *nonselective universities*, (iii) affirmative action based on income (income quota), (iv) Texas Top 10 Law, and (v) offering more AP classes to low-income students.

Table 50: The Attendance Rate into Selective Universities

Baseline	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	2.54	0	1.11	0	0
Math2Q	9.01	7.26	9.09	8.00	10.53
Math3Q	13.33	15.12	17.36	21.09	24.06
Math4Q	35.38	25.45	33.51	24.82	36.41
Math5Q	30.00	42.68	43.36	34.65	44.21
Need Based Aid from Sel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	2.54 (100%)	0	1.11 (100%)	0	0
Math2Q	9.91 (110%)	7.26 (100%)	9.09 (100%)	8.00 (100%)	9.87 (94%)
Math3Q	14.44 (108%)	13.95 (92%)	17.36 (100%)	21.09 (100%)	24.06 (100%)
Math4Q	36.92 (104%)	25.45 (100%)	32.98 (98%)	24.82 (100%)	35.44 (97%)
Math5Q	41.67 (139%)	42.68 (100%)	41.96 (97%)	34.65 (100%)	43.35 (98%)
Need Based Aid from NonSel Univ	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	2.54 (100%)	0	1.11 (100%)	0	0
Math2Q	7.21 (80%)	7.26 (100%)	9.09 (100%)	8.00 (100%)	11.18 (106%)
Math3Q	13.33 (100%)	15.12 (100%)	18.06 (104%)	21.09 (100%)	24.06 (100%)
Math4Q	33.85 (96%)	25.45 (100%)	34.04 (102%)	24.82 (100%)	36.89 (101%)
Math5Q	30.00 (100%)	42.68 (100%)	44.06 (102%)	34.65 (100%)	45.06 (102%)
Income Quota	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	6.78 (267%)	0	0	0	0
Math2Q	24.32 (270%)	6.45 (89%)	8.44 (93%)	8.00 (100%)	7.24 (69%)
Math3Q	31.11 (233%)	12.79 (85%)	15.28 (88%)	18.75 (89%)	20.32 (84%)
Math4Q	50.77 (143%)	20.91 (82%)	31.91 (95%)	24.11 (97%)	33.01 (91%)
Math5Q	51.67 (172%)	41.46 (97%)	39.16 (90%)	31.50 (91%)	41.63 (94%)
Texas Top 10	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	14.41 (567%)	9.46	15.56 (1402%)	8.99	9.88
Math2Q	20.72 (230%)	23.39 (322%)	16.88 (186%)	19.00 (238%)	15.13 (144%)
Math3Q	14.44 (108%)	17.44 (115%)	20.83 (120%)	18.75 (89%)	25.13 (104%)
Math4Q	24.62 (70%)	20.91 (82%)	26.06 (77.8%)	24.82 (100%)	28.16 (77%)
Math5Q	25.00 (83%)	31.71 (74%)	31.47 (72.6%)	23.62 (68%)	30.47 (69%)
Offering more AP classes	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Math1Q	2.54 (100%)	0	1.11 (100%)	0	0
Math2Q	9.91 (110%)	7.26 (100%)	9.09 (100%)	8.00 (100%)	10.53 (100%)
Math3Q	13.33 (100%)	15.12 (100%)	17.36 (100%)	21.09 (100%)	24.06 (100%)
Math4Q	35.38 (100%)	25.45 (100%)	33.51 (100%)	24.82 (100%)	36.41 (100%)
Math5Q	31.67 (106%)	42.68 (100%)	43.36 (100%)	34.65 (100%)	44.21 (100%)

Note. The table shows changes in the attendance rate in selective universities in each counterfactual policy change. I consider (i) increasing need-based aid by $10,000 per attending students from selective universities without changing tuition and financial aid policy for students from other income backgrounds, (ii) the same policy implemented by *nonselective universities*, (iii) affirmative action based on income (income quota), (iv) Texas Top 10 Law, and (v) offering more AP classes to low-income students.

Appendix H: Discussion

Income Effect in Attendance Rate into Private High School

I discuss the characterization of attendance decision into private high school. For simplicity, let V^r and V^u be the value of attending private high school and that of attending public high school. It is better to choose private high school iff

$$V^r - V^u > 0$$

, where

$$V^r = P^r(V^{s,r} - V^{n,r}) + V^{n,r} - U^r_{APcost} - U^r_{apply}$$
$$V^u = P^u(V^{s,u} - V^{n,u}) + V^{n,u} - U^u_{APcost} - U^u_{apply}$$

, which implies that

$$P^r(V^{s,r} - V^{n,r}) - P^u(V^{s,u} - V^{n,u}) + V^{n,r} - V^{n,u} > (U^r_{APcost} - U^u_{APcost}) + U^r_{apply} - U^u_{apply}$$

Note that conditional on application and AP classes, the R.H.S is independent of the family income. Given the same effort level and characteristics, $P^r > P^u$ due to higher learning efficiency. One sufficient condition that LHS is increasing in M_0 is to show

$$\frac{\partial \left[P^r \left[(V^{s,r} - V^{n,r}) - (V^{s,u} - V^{n,u}) \right] + V^{n,r} - V^{n,u} \right]}{\partial M_0} > 0$$

If the individual is not borrowing constrained, the first part of the above equation is increasing in M_0 if the following does so.

$$G = \left[\frac{(M_0 + \frac{M_1^{s,r}}{1+r} - T^s - T^r)(M_0 + \frac{M_1^{n,u}}{1+r} - T^n)}{(M_0 + \frac{M_1^{n,r}}{1+r} - T^n - T^r)(M_0 + \frac{M_1^{s,u}}{1+r} - T^s)} \right]$$

Under the consumption that $M_1^{n,u} + M_1^{s,r} \approx M_1^{n,r} + M_1^{s,u}$ I can show that $\frac{\partial G}{\partial M_0} > 0$ iff

$$\frac{\partial G}{\partial M_0} = \frac{(2M_0 - (T^s + T^n + T^r))(T^r(T^s - T^n))}{(M_0^2 - (T^s + T^n + T^r)M_0 + T^r T^s)^2} > 0$$

Therefore, the positive income effect is stronger for those who apply for selective universities.

Discussion on the Estimate of the Extent of Liquidity Constraint

If I only consider students who have positive amount of student loan, the average is estimated as 14,057 dollar per year. This is larger than the sample average, 12,780 dollar. $\hat{\lambda}$ can be overestimated if the model can rationalize poor students' low consumption level during college periods and low application rate without liquidity constraint. There are four model components related to these margins: application cost (U_{apply}), admission probability (P_{ad}), nonpecuniary benefit (U_{sel}), and intertemporal preference (β). The first three components determine the application decision, whereas the last one decides the amount of student loan. For example, the model could understate the importance of family income in the first three components (i) if there is a private service that can assist application process and write an impressive essay (U_{apply}), (ii) if engaging in expensive extracurricular activities increases admission probability (P_{ad}), (iii) if some of wealthiest student can attend selective universities if parents donate sufficiently large amount of money (P_{ad}), and (iv) if students from low-income families may have emotional stress for not being able to spend extra money for social activities (U_{sel}).

However, the model already allows to reduce application cost by paying tuition and attending private high school, and legacy admission may not be relevant to middle or low-income students who are most likely to be liquidity constrained. Also, colleges also value students who overcome adverse growing up environment in admission process, which could partially counteract not being able to engage in expensive extracurricular activities as rich counterparts. Finally, the nonpecuniary benefit is not significantly affected by θ which is by construction positively correlated with family income.

Table 30 shows that β, the intertermporal preference of consumption is estimated to be 0.26. Small $\hat{\beta}$ can lead to higher $\hat{\lambda}$, because it would rationalize lower consumption value during the college period. One reason for small $\hat{\beta}$ is that all family income is entitled to the student in this paper. In data, although family income is more than twice as high as that of student's labor market earning, more than 70% students take student loans.

Competition Between Colleges and Need-Based Aid from Selective Universities

Consider two colleges $j = \{1, 2\}$ whose utilities depend on tuition revenue and the quality of attendees. Returns from attending college j is R_j. There are unit mass of high-income and low-income students with the distribution of academic qualification $q^L \sim F(\mu_L, \sigma_L^2)$ and $q^H \sim F(\mu_H, \sigma_H^2)$. Also, let $F(\mu, \sigma^2)$ be the pooling distribution of two distributions. Colleges choose a financial aid package to high-income and low-income attendees. I focus on a symmetric policy $(a_j, -a_j)$ such that low-income students get positive transfers and high-income students pay more. The tuition is normalized to zero. Let I_i^L and I_i^H be the index of student's family income, let $I_i^{attend} \in \{1, 2\}$ be the index of whether the student attends College 1 or 2. Colleges first post the financial aid package, then students decide which college to attend. College j's utility function is

$$U^{c_j} = \sum_{i=1}^{N} (I_i^{attend} = j) \cdot \Big[-a \cdot I_i^L + a \cdot I_i^H \Big] + E[q_i^L | I_i^{attend} = j] + E[q_i^H | I_i^{attend} = j] \tag{10.1}$$

Proposition. *If $R_1 = R_2$, $\forall\ a \in [0, R]$, $((a, -a), (a, -a))$ and students are randomly apply to each college is a Nash Equilibrium.*

Even though $-a + \mu_L < a + \mu_H$, college cannot attract only high-income students. Since students are randomly selected into each college, the equilibrium tuition revenue becomes 0 regardless of a. However, to make students attend either college, $R - a \geq 0$.

Next, consider two colleges with different returns $R_1 > R_2$. It can be seen as a selective and a nonselective university. Let F_L and F_H be the cumulative distribution function of q_i^L and q_i^H. Let $\tilde{a}_2$ be a_2 that satisfies

$$-\tilde{a}_2 F_L(h^*) + \tilde{a}_2 F_H(h^*) + E[q_L | q_L < h^*] + E[q_H | q_H < h^*] = E[q_{All}] \tag{10.2}$$

and denote h^* that satisfies $F^{-1}(h^*) = 0.5$. Then the Nash Equilibrium is unique if it exists, and all high-achieving students would attend College 1.

Proposition. *Suppose that $R_1 > R_2$. Then $a_1^* = R_1 - R_2 + \tilde{a}_2$, $a_2^* = \tilde{a}_2$, and $I_i^{attend} = 1$ iff $q^j > h^*$ for $j \in \{1, 2\}$ is a unique Nash Equilibrium.*

Proof. First, in order to have high-achieving students attend College 1, the below condition has to hold.

$$\begin{aligned} R_1 + a_1 &> R_2 + a_2 \\ R_1 - a_1 &> R_2 - a_2 \end{aligned} \tag{10.3}$$

Suppose that the proportion of high-income students is larger than that of low-income students among all high-achieving students $q_j > h^*$; $F_H(-h^*) - F_L(-h^*) > 0$. Note that College 1 can always get the pooling equilibrium utility. If $(a_1, -a_1)$ is a best response of College 1 to $(a_2, -a_2)$, the following condition has to hold.

$$-a_1 F_L(-h^*) + a_1 F_H(-h^*) + E[q_L | q_L > h^*] + E[q_H | q_H > h^*] \geq \mu \tag{10.4}$$

To have $(\tilde{a}_2, -\tilde{a}_2)$ be the best response of College 2, the tuition benefit has to be large enough to compensate for quality loss compared to the pooling equilibrium.

$$-a_2 F_L(h^*) + a_2 F_H(h^*) + E[q_L | q_L < h^*] + E[q_H | q_H < h^*] \geq \mu \tag{10.5}$$

Therefore, if $\tilde{a}_2$ exists, then $a_1^* = R_1 - R_2 + \tilde{a}_2$, $a_2^* = \tilde{a}_2$, and $I_i^{attend} = 1$ iff $q^j > h^*$ for $j \in \{1, 2\}$ is a unique Nash Equilibrium.

Under what condition, competition between college would affect the need-based aid from selective universities? Note that in above discussion, I only allow colleges to offer need-based aid, whereas the extent of redistribution was the choice variable. Suppose that College 2 can now choose $a_2 < 0$ (similar to merit-based aid). In order for College 2 to attract all high-income students, the following condition has to hold.

$$\begin{aligned} R_1 + a_1 &> R_2 - a_2 \\ R_1 - a_1 &> R_2 + a_2 \end{aligned} \tag{10.6}$$

However, unless $-a_1 + \mu_L > \mu$, College 1 would not choose this action and it will not be supported in equilibrium. (Usually, $\mu > \mu_L$). Note that as long as $R_1 > R_2 + a_2$, College 1 can choose a_1 to make high-income students indifferent between College 1 and College 2, $a_1 = R_1 - R_2 - a_2$. If not, College 1 will choose the same aid profile as College 2 because pooling equilibrium is better. Therefore, if the difference in return is not large enough, competition between selective and nonselective universities can limit the need-based aid from selective universities. If there is complementarity between student's quality and college's quality then pooling equilibrium leads to inefficiency. Thus, increasing merit-based aid from nonselective universities might reduce the efficiency.

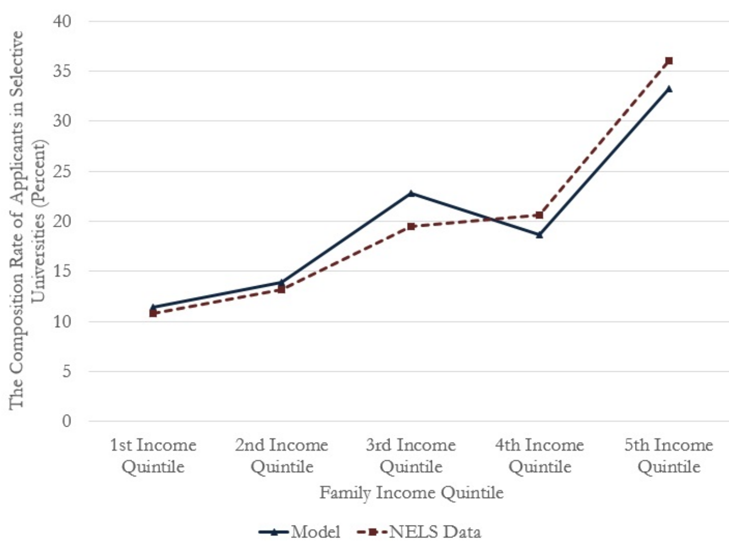
Out-of-Sample Prediction

To examine the validity of structural model, I consider an out-of-sample prediction. The National Education Longitudinal Study of 1988 is a nationally representative sample of 8th graders and it has four follow-up surveys in 1990, 1992, 1994, and 2000. The ELS2002 and the NELS1988 have almost identical survey instruments. Assuming that changes in financial aid policy from selective universities are exogenous from students' point of views, I consider the following exercise. First, I estimate the financial aid policy from selective universities in the NELS1988 sample. Second, I substitute the financial aid policy of the ELS2002 cohort with that of the NELS1988 cohort. Based on the estimated structural model and by using the data of the ELS2002, I can predict students' application decisions, admission results conditional on family income and initial math scores. Third, I look at the raw data of the NELS1988 and calculate the joint distribution of family income, the initial math score, the application, and the admission result. Then I compare how much the predicted outcome tracks the actual data observed in the earlier cohort, the NELS1988.

Figure 31-34 show the result of out-of-sample prediction. Because the aggregate application rate increased from 18% to 42% during this period, I focus on the composition rates of applicants and attendees in selective universities conditional on family income and initial math score rather than focusing on the application and admission rate of each group. Also, I do not consider the number of AP classes taken by students because the number of AP classes offered by the high school increased drastically during this period.

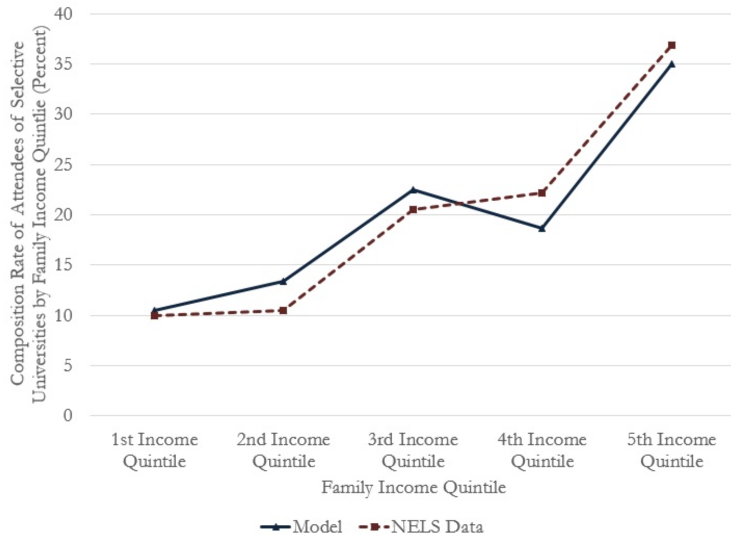
The estimated model predicts the NELS1988 data fairly well in terms of the composition rate by family income quintile. The disparity between the predicted model and the NELS1988 data is less than 3% for each income quintile group. However, the model predict much smaller disparity between students with high and low math scores. In particular, the data show a stark difference in the composition rates between students from the first and the second highest quintile of math score distribution. More difficult application process and much fewer number of available AP classes in the NELS1988 cohort may explain why the selection into selective universities in earlier cohort was more strongly driven by student's initial math score than the sorting pattern predicted based on the ELS2002 sample.

Figure 31: Out-of-Sample Prediction (Composition Rate of Applicants by Family Income Quintile)



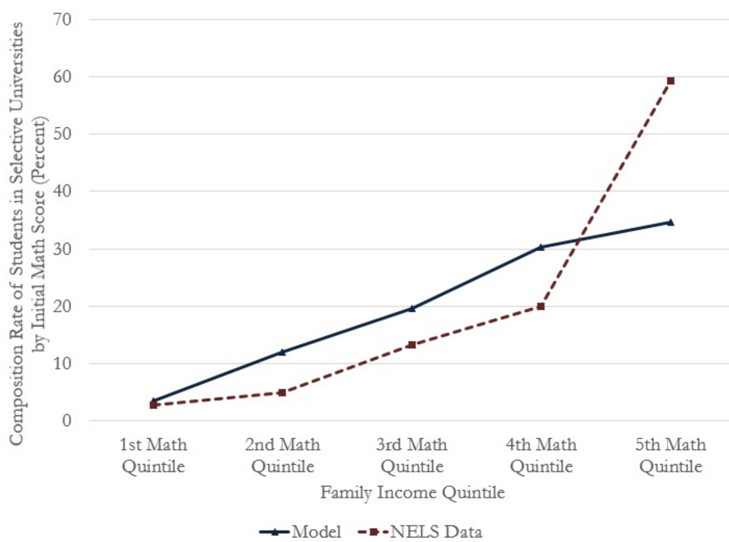
Note. To assess the external validity of the model, I consider a out-of-sample prediction using the National Education Longitudinal Study of 1988. First, I estimate the financial aid policy of the NELS1988 cohort. Second, I substitute the financial aid policy of the ELS2002 sample with that of the NELS1988, then I obtain the predicted counterfactual outcomes based on the estimated structural model. Third, I compare this predicted outcome with the observed data in the NELS1988. During the periods, the aggregate application rate increased more than 24% (from 18% to 42%). Thus I focus on the composition rate than the application rate.

Figure 32: Out-of-Sample Prediction (Composition Rate of Attendees by Family Income Quintile)



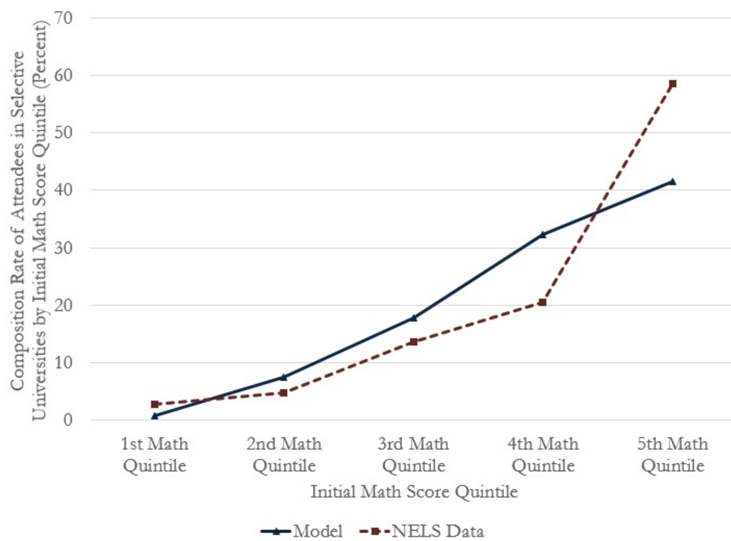
Note. To assess the external validity of the model, I consider a out-of-sample prediction using the National Education Longitudinal Study of 1988. First, I estimate the financial aid policy of the NELS1988 cohort. Second, I substitute the financial aid policy of the ELS2002 sample with that of the NELS1988, then I obtain the predicted counterfactual outcomes based on the estimated structural model. Third, I compare this predicted outcome with the observed data in the NELS1988. During the periods, the aggregate application rate increased more than 24% (from 18% to 42%). Thus I focus on the composition rate rather than the attendance rate.

Figure 33: Out-of-Sample Prediction (Composition Rate of Applicants by Initial Math Score Quintile)



Note. To assess the external validity of the model, I consider a out-of-sample prediction using the National Education Longitudinal Study of 1988. First, I estimate the financial aid policy of the NELS1988 cohort. Second, I substitute the financial aid policy of the ELS2002 sample with that of the NELS1988, then I obtain the predicted counterfactual outcomes based on the estimated structural model. Third, I compare this predicted outcome with the observed data in the NELS1988. During the periods, the aggregate application rate increased more than 24% (from 18% to 42%). Thus I focus on the composition rate rather than the application rate.

Figure 34: Out-of-Sample Prediction (Composition Rate of Attendees by Initial Math Score Quintile)



Note. To assess the external validity of the model, I consider a out-of-sample prediction using the National Education Longitudinal Study of 1988. First, I estimate the financial aid policy of the NELS1988 cohort. Second, I substitute the financial aid policy of the ELS2002 sample with that of the NELS1988, then I obtain the predicted counterfactual outcomes based on the estimated structural model. Third, I compare this predicted outcome with the observed data in the NELS1988. During the periods, the aggregate application rate increased more than 24% (from 18% to 42%). Thus I focus on the composition rate rather than the attendance rate.