10-2016

Need based aid from selective universities and the achievement gap between rich and poor

Sunha MYONG
Singapore Management University, sunhamyong@smu.edu.sg

Need-Based Aid from Selective Universities and the Achievement Gap between Rich and Poor

Sunha Myong*†

October 31, 2016

Abstract

I study the role of need-based aid from selective universities in closing the achievement gap between rich and poor high school students. I focus on the incentive aspect of need-based aid that can change high school student's effort choices. The impact of increasing need-based aid depends on the extent of borrowing constraints and how competition affects the relative performance of low- and high-income students. 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. I control for other types of barriers for low-income students such as a lack of information or low high school quality. I use a geographic variation in costs of attending home-state nonselective universities to control selection biases driven by an unobservable characteristic correlated with family income. I find that 6.9% of high-ability low-income students do not apply to selective universities because of borrowing constraints. If selective universities double the amount of grants per attending student from the bottom quintile of the income distribution, the effort gap, as measured by the number of Advanced Placement (AP) classes taken, decreases by 33.4%, the achievement gap, as measured by the SAT score, by 20.2%, and the wage gap by 10.2% among students with the initial test scores in the top 20th percentile in 10th grade. The aggregate achievement score also increases because elevated competition raises the effort level of high ability applicants from all income backgrounds. Doubling need-based aid can close the achievement gap better than merit-based aid that requires the same budget by 21%, while they have similar impacts on the aggregate achievement level.

*School of Economics, Singapore Management University; Email: sunhamyong@smu.edu.sg

†I am deeply grateful to B. Ravikumar, Limor Golan, and Carl Sanders for their guidance and support. I also thank Goerge-Levi Gayle, Juan Pantano, Yongseok Shin, John Nachbar, Bruce Petersen, Jungho Lee, and JungJae Park for valuable discussions and suggestions.

1 Introduction

Based on a nationally representative sample of high school students in the United States, I observe a widening achievement gap between rich and poor students in test scores during high school. For example, looking at students within the top 10th percentile of standardized scores in 10th grade, students from the top quintile of income distribution obtain SAT scores 95 points higher than those from the bottom quintile of income distribution.[1] Previous studies usually focused on different ability, parental investment, and high school quality between rich and poor students to explain the achievement gap.[2] However, I find that students' own effort choice is also a significant determinant of the widening achievement gap during high school. For example, the regression result suggests that different effort choices, as measured by the number of Advanced Placement classes students take, explain 28% of the conditional achievement gap related to family income. This finding is relevant to understanding a source of persisting income inequality across generations because the achievement score of students accounts for substantial variations in labor earnings within the same education group.[3]

What can be a cost effective way to close the achievement gap between rich and poor high school students? I focus on need-based financial aid from selective universities.[4] Because of the competitive admission processes for the limited capacity, students who *apply* to selective universities, almost half of four-year college attendees, generally put more efforts into their studies such as taking Advanced Placement (AP) classes. This, in turn, leads to higher achievement scores. However, there exists a large disparity between rich and poor students in terms of college application behaviors, in particular, whether to apply for selective universities or not.[5] Although recent studies focused on information disparity regarding college opportunity[6], I focus on need-based aid for following reasons. First, the direct cost of college education is still a significant determinant of the application decision for selective universities. In particular, $10,000 increase in tuition of home-state nonselective universities increases the application rate for selective universities by 5.0%. More importantly, low-income students are more sensitive to the direct cost; they increase the application rate by 1.6% more than the average students if attending nonselective universities becomes more expensive by $10,000. Second, although selective universities have increased financial aid per student by more than 60% over the last decade, more selective universities started to award grants

[1]Data comes from the Education Longitudinal Study of 2002. It is the sum of the verbal and math SAT scores. The pattern holds for other percentiles of the standardized test scores in 10th grade.

[2]For example, see Todd and Wolpin (2007) and Rivkin, Hanushek, and Kain (2005).

[3]For example, see Hanushek and Weissman (2008) and Neal and Johnson (1996).

[4]Selective universities refer to the four-year colleges in the United States that belong to the top two categories among the seven categories of the NCES-Barron's Admissions Competitiveness Index Data Files 2004. According to this classification, selective universities account for about 20% of four-year enrollment.

[5]Focusing on students within the top 20th percentile of the initial test scores, students from the top quintile of income distribution apply 13% more to selective universities than those from the bottom quintile income distribution.

[6]For example, see Hoxby and Turner (2013).

to middle- and high-income students by redefining need-based aid or increasing merit-based aid.[7]

In this paper, I quantify to what extent additional need-based aid from selective universities can close the achievement gap between rich and poor students. In particular, I focus on the incentive aspect of need-based aid from selective universities that can change high school students' effort choices. In doing so, I examine how tuition subsidies from selective universities affect high school students' forward looking behaviors when they make choices about the number of AP classes to take, private high school attendance, and college applications. This is an empirical question because the size of impact depends on the extent of borrowing constraints, the importance of other types of barriers such as a lack of information, and how elevated competition affects student's choices conditional on family income and student ability.

My paper is the first to quantify the impact of financial aid on high school students' academic achievement when there is competition among students for the limited capacity in selective universities. Unlike need-based aid from nonselective universities or state-funded merit based aid such as Georgia's HOPE Scholarship,[8] the chance of receiving financial aid from selective universities depends on student's *relative* academic performance compared to all other applicants. Also, this paper is the first to evaluate how borrowing constraints in college financing affect high school student's effort choices and test scores. Most previous studies take high school test score as a fixed ability measure.

To quantify the impact of additional need-based aid from selective universities, I develop a structural model of student's learning, application, and admission processes. In the model, attending selective universities provides pecuniary and nonpecuniary benefits. However, attending selective universities is much more expensive than 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 AP classes to take during high school, and (iii) whether to apply for selective universities. After the admission result is realized, the student takes out student loans to finance the net cost of college education. Borrowing constraints could limit student's borrowing capacity as a fraction of her future annual earnings. The admission probability is determined in an equilibrium such that given the admission cutoff value in test scores, the number of seats in selective universities is equal to the number of attendees.

I estimate the model based on three data sets: (i) the Education Longitudinal Study of 2002 (ELS2002), (ii) the Integrated Postsecondary Education System Data (IPEDS 2004), and (iii) the NCES-Barron's Admissions Competitiveness Index Data Files 2004. The main data set is the

[7] Table 11 and Table12.

[8] The HOPE Scholarship (Helping Outstanding Pupils Educationally) is originated in Georgia which has awarded grants to students in state public and private colleges from 1993. Funded by the state lottery, $3 billion in grants were awarded to 900,000 Georgia students from 1993 to 2006. In the original program, to become eligible a student has to have B average GPA.

ELS2002, which provides a wide range of students' demographic characteristics, high school curriculum choices and test scores, the college application and admission results, and the hourly wage rate. I use the IPEDS for tuition information and the NCES-Barron's Admissions Competitiveness Index to define college selectivity. Four particular variables are useful in identifying the model. First, the geographic variations of tuition of home-state nonselective universities and the cost for the room and books are important exogenous variations in identifying students' responses to financial incentives. Second, I use the total amount of student loans to identify the intertemporal preference over the consumption between college and the work period. Then the different choices of student loans between rich and poor students help to identify the extent of borrowing constraints. Third, the number of information sources about the college application process helps to control the 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 impact of additional need-based aid from selective universities on the achievement gap between rich and poor is not obvious for the following reasons. First, low-income students receive more than twice the amount of aid from selective universities as high-income students. Therefore, it is not clear whether low-income students at the margin of the application are borrowing constrained. If the main reason that those marginal students did not apply for selective universities was the low admission probability, tuition subsidies might not affect their behaviors. Second, other types of barriers, for example, the information barrier regarding the college application process or the disparity in the number of AP classes available in high school, can be too high for low-income students to respond to the financial incentive. Third, elevated competition driven by an increased number of applicants could affect the behavior of high-income students. If high-income students also increase their effort levels to remain competitive, and can further increase test scores by spending more money, for example, by attending private high school, the achievement gap could increase.

The estimated model provides four sets of quantitative results. First, although need-based aid from selective universities is already extensive, 6.9% of low-income students do not apply for selective universities because of borrowing constraints. Focusing on students within the top quintile of initial test score distribution, borrowing constraints increase the effort gap by 14.0%, the achievement gap by 8.4%, and the wage gap by 4.7%. Second, further increases in need-based aid from selective universities can not only reduce the achievement gap between rich and poor students, but also increase the aggregate effort level. In particular, if selective universities double the amount of grants per attending student from the bottom quintile of the income distribution, which corresponds to \$12,000 more annual grants per low-income attendee, the average effort level of those low-income high school students increases by 13.4%. This decreases the effort gap by 33.4%, the achievement gap by 20.2%, and the wage gap by 10.2% among students in the top 20th percentile of standardized scores in 10th grade. By relaxing borrowing constraints facing low-income high ability students whose effort choice would be most elastic to tuition subsidies, need-based aid can effectively reduce the achievement gap between rich and poor students. Also, increasing need-based aid increases the aggregate effort level by 0.7%. Focusing on those who actually increase the effort level, low-income students increase by 52.1% and high-income students increase by 12.0%. Because of the tournament effect–the additional effort applicants put forth to increase the probability of admission–is more elastic to student ability than to family income, elevated competition increases the effort level of all high ability applicants. Third, as a counterfactual policy analysis, I compare the impact of doubling need-based aid to merit-based aid from selective universities. I find that need-based aid can close the achievement gap better then merit-based aid by 21.1%, while keeping the aggregate achievement level almost the same. Merit-based aid is not significantly better at providing incentives to students. Because merit-based aid also increases the test score of high-income students, some low-income high-ability students are discouraged from application. Finally, for a comparison of the existing policies, I examine the impact of changing admission criteria similar to the Texas Top 10 Law—where only the high school GPA is taken into account in the admission process. The Texas Top 10 Law[9] can substantially reduce the achievement gap among students in the top quintile of the initial test scores, as measured by SAT scores, by 39.5%. However, the aggregate effort level decreases by 49.2%, which corresponds to a $1,300 dollar reduction in annual average income of low-income high-ability students.

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. Looking at the composition rate of applicants and attendees in selective universities conditional on the 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 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 describes the estimation method. Section 7 documents the estimation result. Section 8 presents the counterfactual analysis. Section 9 shows an out-of-sample prediction. Section 10 concludes.

2 Related Literature

First, this paper quantifies how financial aid for college education can affect the high school achievement score. This complements the existing literature which mainly discusses the impact of tuition subsidies on college attainment results (Cameron and Taber (2004), Cameron and Heckman (1998),

[9]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.

and Lochner and Monge (2002)). Also, this paper relates to Becker and Tomes (1994) which shows how borrowing constraints can explain the different educational investments between rich and poor students. As an extension, this paper studies the impact of increasing need-based aid from selective universities on high school students' effort choices, private high school attendance, and test scores. Kinsler and Pavan (2011) shows that family income is significant determinant of the quality of higher education especially among high-ability students. This paper extends their analysis by investigating how different college application behavior between rich and poor students are related to the achievement gap, and by accounting for different structural channels—initial academic achievement, high school type, unobservable characteristic, and borrowing constraints for college financing—through which family income affect choices of high school students.

Second, this paper estimates the quantitative importance of the competitive college admission processes on student's effort choices at the high school level. This relates to Hickman (2013) which studies the impact of race-based affirmative action in the college admission process on students' effort choices based on the auction theoretical framework. The main difference is that this paper focuses on the interaction between borrowing constraints and competitive admission processes, whereas the college financing is not considered in Hickman (2013). By differentiating family income from ability, this paper isolates the impact of financial incentives on student's effort choices in the presence of a competitive admission process.

Third, this paper contributes to the literature regarding the impact of the college admission processes on the student-college assignment. Previous studies focus on race-based affirmative action (Arcidiacono (2005) and Hickman (2013)) or the Texas Top 10 Law (Kapor, 2014). This paper focuses on need-based aid from selective universities, a major policy instrument of selective universities to recruit high ability low-income students.[10] This paper also relates to Hoxby and Turner (2013) which shows the importance of information barriers facing low-income students to explain different application behaviors for selective universities between rich and poor students. By controlling the number of information sources about college application processes, this paper evaluates the impact of need-based aid in the presence of information barriers.

Finally, the structural model allows me to compare the counterfactual impact of need-based aid with other policies such as merit-based aid or the Texas Top 10 Law on the achievement gap and the aggregate achievement level. Dynarsky (2010) and Kane (2003) evaluate the treatment effect of state-funded merit-based aid on student's enrollment decisions.[11] Based on reduced form analyses,

[10] On the other hand, some studies 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, as a result of collecting student information from the Free Application Federal Student Aid (FAFSA). Different from those studies that take students' academic achievement at the moment of college application are taken as given, 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.

[11] Van De Klaauw (2002) shows that merit-based aid at an institutional level can effectively raise enrollment rates

they find that those merit-based scholarships significantly increases the enrollment rates into state universities[12]. They find that merit-based aid from state universities benefits mostly the middle- and high-income students and it widens the achievement gap between rich and poor students. This paper contributes to the literature by focusing on need-based aid from selective universities and showing that need-based aid from selective universities can reduce the achievement gap better than merit-based aid with the same amount of budget.[13]

3 Data

3.1 Sample Description

I use the Education Longitudinal Study of 2002 (ELS2002), the Integrated Postsecondary Education Data System (IPEDS), and the NCES-Barron's Admissions Competitiveness Index 2004 to estimate the model. The ELS2002 is a nationally representative sample of United States high school students, following 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 the students' academic achievements as measured by standardized test scores, school characteristics, family background, college attendance/graduation, and the wage rate. However, the information to which university the student applied, admitted 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 to which the student applied, admitted, 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 10th graders in 2002. This paper only focuses on four-year college attendees. I dropped data with missing information such as the initial math score, parents' educational attainment, family income, high school type, SAT, GPA, and AP score, or the history of college application, admission and attendance, and the wage rate. The final sample size is 3,080.[14]

The transcript data in the ELS2002 allows me to observe the entire history of a student's high school curriculum choices. In particular, I track the total number of AP classes taken by students during high school years. 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

in the presence of competition among colleges for students by using the Regression Discontinuity approach.

[12] 3-7% increase in the enrollment rate into the state universities if the amount of grants increases by $1,000.

[13] The total amount of grants from selective universities per year, most of which are awarded based on students' financial need, is $3 billion (IPEDS 2012). This is equivalent to the amount of grants from the Georgia's HOPE Scholarship program during 1993-2006.

[14] Wage rates are observed in the 3rd follow up and the attrition rate is high.

school achievement, I use three scores: SAT, GPA, and the AP score.[15] Observing students' initial academic achievement as measured by the standardized math score in 10th grade[16] is useful to control for different achievement levels at the earlier period of high school. Having early academic achievement helps to quantify how student's own effort choice and high school type contribute to achieving high test scores.

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 that students receive at the first-attended postsecondary institution. I use 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.[17] This helps to control for information disparity between the rich and the poor, an issue which has received increasing public and policy interest.[18]. The ELS2002 also has data reported in 10th grade on students' preference for certain features of college education. For example, students ranked the importance of the college's reputation on a scale 1 to 3. To account for the pure preference difference in application behaviors across students I include student's reported preference for the college's reputation, location, and whether her parents attended the same college.

The IPDES has two pieces of important information. First, I use the posted tuition and fee of each university along with the cost of room, board, and books to account for the direct costs of college education. I assume that tuition of nonselective universities is the average tuition of nonselective universities in student's home state. I also account for home state discounts as reported by each institution in the IPEDS (Figure 3- 4).[19] Second, I include the number of selective universities in students' home states. Including exogenous geographic variations, presumably independent of student's unobservable characteristic, is useful in controlling the selection bias driven by the potential correlation between family income and unobservable characteristic.

I use NCES-Barron's Admissions Competitiveness Index 2004, to define college selectivity. It

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

[16]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 1988. 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. The t-test shows that the growth rate does not based on whether the student attends private high schools or not.

[17]The questionnaire asks 10th grade high school students about whether they obtain information about college application processes from parents, high school counselor, teachers, coaches, relatives, etc.

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

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

covers all four-year colleges across the United States and provides an admission-competitive index for each university. There are seven categories of selectivity level in NCES-Barron's Index, but in this paper, selective universities are in the top two categories of the index in 2004.[20] This leads to 103 selective universities out of 1,247 post secondary institutions in 2004. In my sample 20% of all four-year college enrollees in my sample attend selective universities.

3.2 Descriptive Analysis

Increasing Achievement Gap between Rich and Poor Students

If I look at the SAT score (the sum of verbal and math SAT scores) of students within the top 10th percentile of the standardized math score in 10th grade, students from the top quintile of income distribution receive 95 higher score than students from the bottom quintile of income distribution. If I look at students within the top 20th percentile of the standardized math score in 10th grade, the corresponding number is 75 SAT points. Although students had the same test scores in 10th grade, the achievement level, as measured by SAT score, at the moment of high school graduation is higher for rich students than poor students. Figure 2 describes the distribution of the SAT score of low- and high-income students who had the top quintile of initial test scores in 10th grade.

Table 1 shows the linear regression of the log SAT score on student's initial test scores, family income, and other demographic characteristics. The coefficient of the log family income before controlling for the number of AP classes taken is 0.037. This implies that 10% higher family income corresponds to 3.7% higher SAT score. If I control the number of AP classes taken, it decreases to 2.6%. Thus, the elasticity of the SAT score with respect to family income decreases by 28% if I control student's own effort choices.

High School Achievement and Wage Rate

Table 2 shows the linear regression of the wage rate on high school achievement, family backgrounds, and curriculum choice. Column (3) indicates that log SAT (with a coefficient of 0.242) and log GPA (0.207) scores significantly increase the log hourly wage, after controlling race, gender, high school curriculum choice, initial math score, family income, parents' education, and two personality traits: motivation and action control. [21] Table 3 shows that log SAT (0.166) and log GPA (0.125) remain significant after controlling for college major, dropout, and further education. These results suggest that high school achievements predict a significant portion of early period wage rate.

To see a long-run effect I use the other data set, the National Longitudinal Study of Youth 1979. Table 4 shows the wage regression result. It is a pooling estimation of the panel data that

[20] Wage premium from attending colleges with the top two categories are significantly different from that of attending the rest of four-year colleges.

[21] The motivation variable measures student's attitude toward financial success in life.

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 (with a coefficient of 0.63), while its impact on the growth rate (SAT$\times$AGE) is not significant.

Cost of Attending Selective Universities and Composition Rate by Income

Table 8 documents tuition, cost of room, board and books 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.

Figure 5 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(a) and 9(b) document the attendance rate (attendees/all), application rate (applicants/all), and admission probability (attendees/applicants) of students by income quintile. Figure 9(a) shows the measures of students at all achievement levels, whereas Figure 9(b) shows the measures for high-achieving students within the top 20th math scores in 10th grade. In aggregate, family income is positively correlated with all three measures. For high-ability students, Figure shows an inverted U-shape. The admission probability upon application is actually higher for students from the lowest income quintile distribution than those from the highest income quintile distribution. However, still the application rate of high-ability students from the bottom quintile of income distribution is 13% lower than that of high-ability students from the top quintile of income distribution.

Financial Aid from Selective and Nonselective Universities

Figure 10 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. [22] At nonselective universities, students from the richest families receive more aid from because nonselective universities also have merit-based aid. Table 9 shows the regression result of the amount of grants students received in ELS2002 sample. The amount of aid from selective universities decreases substantially as family income increases. Thus, although the sticker price of attending selective universities is much more

[22] The income classification is based on the data in the IPEDS 2004.

expensive, attending selective universities can be less expensive for low-income students than high-income students.

Figure 11 shows the trend in institutional grants from selective and nonselective universities from 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 minor increase in institutional grants offered by nonselective universities. Table 11 describes the recent introduction of the 'no-loan policy' by top-ranked universities that replaces loans with grants from the financial aid package especially for low-income students. For students with family income less than $60,000 attending those top-universities would now receive a full tuition discount. Figure 12 shows the trend in the average Pell Grant per student at selective and nonselective universities over the same period. There is no significant difference in the trend 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.

However, although selective universities have increased average financial aid, recently more aid is awarded to middle- and high-income students. As shown in Table 11 the 'no-loan policy' includes increasing income level for qualifying for need-based aid. Table 12 shows a couple of examples of recent changes in the average merit based aid from some selective universities. Some selective universities increased merit-based aid, whereas others decreased it. within selective universities. This might provide motivation to compare the impact of need-based aid from merit-based aid from selective universities. [23]

Not only need-based aid from selective universities is extensive, but also the size is quite large. In 2012-2013 academic year, the total amount of grants awarded to the first-year students in selective universities was $3 billion among which $2.5 billion was accounted for institutional grants. This is almost 40% of entire Pell Grants provided by the federal government during the same academic year.[24]

Direct Costs, Information Barrier, and Application Behaviors

Table 15 shows the probit regression of application decision for selective universities on student's demographics, family income, test scores, the average cost of attending home-state nonselective universities, and the number of information sources about college application processes. It suggests

[23] Because this information is not available in the IPEDS, I refer to the New York Times article that analyzes Recent College Board data from more than 600 nonprofit colleges. The analysis suggests that half of selective universities have increased the average amount of merit-based aid, whereas the other half have decreased after 2007. `http://www.nytimes.com/interactive/2012/07/08/education/edlife/8edlife_chart.html`

[24] Data comes from IPEDS2012. The corresponding numbers for nonselective universities were $14.9 billion and $6.7 billion. On the other hand, the amount of Pell Grants of which 90% recipients come from families with less than $40,000 annual income given to the first-year college attendees was $0.2 billion in selective universities and $5.8 billion in non-selective universities. Considering extensive need-based aid in selective universities, the proportion of public expenditure that goes to low-income college attendees is much smaller in selective universities than nonselective universities.

that the direct cost as measured by the cost of attending home-state nonselective universities explains more the application behavior of students than the number of information sources available to students. In particular, \$10,000 increase in the cost of attending home-state nonselective universities raises the application rate by 5.0%. Importantly, low-income students are more sensitive to the direct cost than students from other income levels. In particular, \$10,000 increase in the direct cost would raise the application rate of low-income students additionally by 1.6%. On the other hand, additional information source increases the application rate only by 1.1%.[25] Although Hoxby and Turner (2013) show the importance of information barrier facing low-income students, the regression result suggests that the direct cost of college education still matters to understand the application gap between rich and poor students with similar academic qualification.

Advanced Placement Classes

Figure 9 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 the high school of students who do not take any AP classes. Of those nonparticipants, more than 25% of the students from the bottom quintile of income distribution do not have AP classes available in high school, whereas the corresponding number for the highest income quintile group is 15%.[26]

Table 13 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 14 and Table 2 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

[25]If I compare the average number of information sources available to students, it is 3.4 for students from the top quintile of income distribution and 2.9 for the bottom quintile of income distribution. Table 17 shows the regression result of the number of information sources available to students about the college application processes. Family income and attending private high school significantly increase the number of information sources available to students.

[26]This motivated the government to increase policy interventions that aim to increase AP participation by low-income students. For example, during 2002-2011, the Federal government granted \$22 million in Advanced Placement Incentive Program grants to 20-30 high schools in certain districts (Advanced Placement Incentive Program Grants).

wage return. Table 5 is the regression result of SAT score including other types of effort choices during high school. It suggests that AP class explains SAT score substantially more than other types of effort choices such as the time spending on homework or the unit of total math credits.

4 The Model

I consider an individual who is about to enter a high school and live three periods. Everyone goes to a four-year college after high school graduation, and there are two types of colleges: selective and nonselective. A competitive admissions process exists for the limited seats in selective universities, and that process focuses mainly on students' academic achievements. In the first period, she engages in high school education. At the beginning of the first period, she chooses whether to attend private or public high school, how many AP classes to take, and whether to apply for selective universities. Test scores, college admission result, and the amount of financial aid are realized at the end of the high school. In the second period, she attends a college. She might dropout depending on her academic preparation level, college selectivity, and a random shock. She finances the cost of college education and takes student loans if necessary. In the third period, the individual becomes a full time worker and repays student loans. Attending selective universities provides pecuniary and nonpecuniary benefits. Its direct costs are much greater than those of attending nonselective universities. The individual consumes during the last two periods. In what follows, first, I describe the general description of the structural model, then I 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 entering high school. A is the initial observable ability, θ is an unobservable characteristic, 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 consumption 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. Let $\epsilon = \{\epsilon_H, \epsilon_p, \epsilon_{BA}, \epsilon_w, \epsilon_{AP}\}$ be i.i.d. random shocks associated with $Y = \{H, I_{sel}, I_{BA}, M_2\}$ and the effort cost U_{AP}. (ϵ_H, ϵ_p) are realized at the end of the first period, ϵ_{BA} is realized at the beginning of the second period, and ϵ_w is realized at the beginning of the third period. Attending a private high school costs tuition t

which follows a normal distribution of $N(\mu_t, \sigma_t^2)$. Students observe t before they decide whether to attend private high school.

The test score H is determined following a learning technology, $H = F_H(x; A, \theta, \epsilon_H)$ where $x = \{I_{private}, N_{AP}\}$ are two inputs that affect the final test score. Note that the learning efficiency is affected by initial observable ability A and unobservable characteristic θ. On the other hand, ϵ_H is realized at the end of the high school period. The admission criteria and financial aid policy are exogenously given. There are stochastic components that affect (i) the admission into selective universities (I_{ad}), (ii) the college graduation (I_{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 ability 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 in the test score 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 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).[27] 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 realized net tuition.

At the moment of college entrance, the student knows $(I_{sel}, I_{BA}, \overline{T})$. She has correct information about the expected labor market earning $\overline{M_2} = E_{\epsilon_w}[M_2] = F_M(\theta, H, I_{sel}, I_{BA}, \epsilon_w)$. 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 a government loan and a 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 borrowing constraints. In particular, I assume that the student can borrow only up to a certain fraction of her future expected income, $L \leq \lambda \overline{M_2}$.

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 characteristic 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 financing channel, all other correlations between family income, choices of students, and economic outcomes are attributed to θ. The student's problem at the moment of high school entrance can be written as

[27] Data shows that test scores do not affect the amount of financial aid from selective universities but affect the amount of aid from nonselective universities.

$$\max_{\{N_{AP}, I_{private}, I_{apply}, L, C_1, C_2\}} \int \Bigg\{ P_{ad}\Big[u(C_1^s) + \beta u(C_2^s) + U_{sel}\Big] + \\ (1 - P_{ad})\Big[u(C_1^n)) + \beta u(C_2^n))\Big] - U_{AP} - U_{apply} \Bigg\} dF(S') \tag{4.1}$$

subject to

$$\begin{aligned} &C_1^k(S') + T^k(S') \leq M_1 + L^k(S') - tI_{private} \\ &C_2^k(S') \leq M_2^k(S') - (1+r)L^k(S') \\ &L^k(S') \leq \lambda E_{\epsilon_w}[M_2^k(S')] \end{aligned} \tag{4.2}$$

for all $S' = \{H, I_{BA}, \epsilon_w\}$ and where $k \in \{n, s\}$. 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 in terms of the choice about admission criteria, financial aid and tuition policy. Now I describe the empirical specification of each component of the structural model.

4.1 Preference

I assume a log utility from consumption. β is the discount factor between the college and the working period. Denoting 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 the student's initial observable ability (A), unobservable characteristic (θ), the number of selective universities in student's home state ($N_{SelHome}$), and the student's preference for certain features of college education such as location, reputation, and whether they want to attend the same college as their parents attended (RP_m). Ability measures in the nonpecuniary benefit are expected 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 expected 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.[28]

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

$$U_{sel} = \nu_{10} + \nu_{11} \ln(A) + \nu_{12}\theta + \nu_{13} N_{SelHome} + \sum_{m=1}^{3} \nu_{RP_m} RP_m$$

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

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

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

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

4.2 Unobservable Characteristic

There is one-dimensional unobservable characteristic across individuals denoted by θ. It might capture the time management skill or parents' guidance that help students to obtain better outcomes. θ is allowed to affect each part of the structural component: the wage rate, learning efficiency, the 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 the student's family income (M_1) and the parent's education (Edu_p).

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

, where $\epsilon_\theta \sim N(0, \sigma_\theta^2)$. Although the mean of θ depends on family background, it is possible that low- and high-income students have the same θ depending individual specific realization of ϵ_θ.

4.3 Learning Technology

I assume that $H = (H_{SAT}, H_{AP}, H_{GPA})$: there are three test scores - SAT, AP, and GPA, relevant in the admission process, the wage rate, and graduation rate from the college. The test scores depend on how many AP classes were taken and whether the student attends private high school. Also, the learning efficiency is affected by the 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 such as SAT and AP and the GPA which is affected by the achievement level of peers. In particular, if the average ability level of peers in a private high school is higher than that of public high school, students who attend private high school may obtain a lower GPA. For $j = \{1, 2, 3\}$,

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

, 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 are exogenously given and students are fully aware of the rule. Selective universities rank applicants based on a measure for student's merit, $g(D)$, which is a function of $D = \{race, H, \theta\}$. Let h^* be the cutoff value in $g(D)$ that determines the admission result. The admission criteria $g(D)$ is specified as

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

Let $P_{ad}(D)$ be the admission probability. Then $P_{ad}(D) = P(g(D) > h^*) = \Phi(g(D) - h^*)$.

4.5 Financial Aid

Let G_k be grants from college with selectivity $k \in \{sel, non\}$. Assume that the amount of financial aid from selective universities is completely need-based, and it only depends on the quintile of the student's family income. 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 quintile and the academic achievement scores (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

After entering the college, the student faces a stochastic process determining whether she can graduate from the college or drop out. 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} and a random shock η_{BA}. Let P_{BA} be the probability of obtaining BA degree and it can be written as follows [29]

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

4.7 Wage Rate

I assume that every student becomes a full time worker in the third period. Let $I_{drop} = 1 - I_{BA}$ be the dummy variable regarding the college dropout. The wage rate is determined by student's high school achievement (H), unobservable characteristic (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} captures the loss of returns from a college degree net of returns from actual labor market experience.

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

, where $Z_w = \{Black, Asian, Hispanic, Female\}$ are the demographic characteristics that affect 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 Equilibrium

Let Q_{sel} be the number of seats in selective universities. The college determines the admission cutoff value h^* that satisfies

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

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

The cutoff value of admission is determined to equalize the number of seats in selective universities and the number of admitted students.[30]

Let Π be the set of structural parameters including the wage rate Γ_0. Define $X^i = \{N^i_{AP}, I^i_{apply}, I^i_{private}, C^i_1, C^i_2, L^i\}$ and $Y^i = \{H^i, I^i_{sel}, I^i_{BA}\}$ for each $i \in N$.

Suppose that Π is fixed. Let $\{X^i, Y^i\}_{i=1}^N$, $\{M^i_1\}_{i=1}^N$ and h^* constitute an equilibrium if for each $i \in N$, X_i solves the following problem taking Π as given

$$\max_{X^i} E\left[u\big(C^i_1(S'^i)\big) + \beta u\big(C^i_2(S'^i)\big) + U_{sel}(X^i)\right] - U_{AP}(X^i) - U_{apply}(X^i)$$

subject to

$$\begin{aligned}
&C^i_1(S'^i) + T^i(S'^i) \leq M^i_1 + L^i(S'^i) - tI^i_{private} \\
&C^i_2(S'^i) \leq M^i_1 - (1+r)L^i(S'^i) \\
&L^i(S'^i) \leq \lambda\overline{M_2}(S'^i)
\end{aligned}$$

and satisfies

$$\sum_{i=1}^{N} I^i_{sel}(g(D_i) > h^*) = Q_{sel}$$

,where Q_{sel} is the number of seats in selective universities.[31]

4.9 Characterization

4.9.1 Family Income and Application for Selective Universities

Suppose that everyone graduates from the college. Denote M^{s*}_2 and M^{n*}_2 to be the labor market income when graduating from selective and nonselective universities respectively. For simplicity of exposition, I abstract from effort choice and private high school attendance decision. Suppose that the student who attends $j \in \{sel, non\}$ type of college is not borrowing constrained. Then the optimal choices for the consumption and student loan are $C^*_1 = \frac{1}{1+\beta}\Big[M_1 + \frac{1}{1+r}M^{j*}_2 - T^j\Big] = \frac{1}{\beta(1+r)}C^*_2$,

[30] I assume that there is no asymmetric information regarding θ between students and admission committees. Or it can be considered as (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) $g(H,\theta) = g(H, f^{-1}_\theta \theta)$ is known to everyone. Thus, the student can infer her admission probability correctly.

[31] Given Π, I can show that the equilibrium cutoff value h^* is unique. The key assumption is that the wage rate $\Gamma_0 \subset \Pi$ is fixed. The student at the margin of application has to equalize the marginal benefit of effort choices regardless of the admission probability to the marginal cost. This pins down the unique cutoff value of h^* given the set of structural parameter Π.

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 the borrowing 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* $K = \frac{1}{1+\beta}\left(\frac{U_{apply}}{P_{ad}} - U_{sel}\right)$.

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 in a dormitory at a prestigious campus values more than living in a dormitory at 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 terms of the utility. For expositional simplicity, I assume that there is no effort choice in the following discussion and abstract from characteristic difference in demographics, preferences, and geographic location of residence. 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 the probability of admission P_{ad}. Thus it is possible that two students make different application decisions 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 same for the rich and the 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* 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 spend want to attend selective universities and enjoy the nonpecuniary benefit of prestigious campus life. [32]

Proposition. *The student decides not to apply for selective universities because of the borrowing*

[32] 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_2. This is the typical case discussed in the previous literature such as Cameron and Taber (2004) and Cameron and Heckman (1998).

constraint although the optimal choice is to apply for selective universities iff (i) $M_1 < \Big[\frac{1}{\beta(1+r)} - \lambda\frac{1+\beta}{\beta}\Big]M_2^{s*} + \beta T^s$*, and (ii)* $J_1 \frac{\left(\left[M_1+\lambda M_2^{s*}-T^s\right]\left[(1-\lambda(1+r))M_2^{s*}\right]^{\beta}\right)^{\frac{1}{1+\beta}}}{\left[M_1+\frac{1}{1+r}M_{2*}^{n}-T^n\right]} < exp(K) < \Big[\frac{M_1+\frac{1}{1+r}M_2^{s*}-T^s}{M_1+\frac{1}{1+r}M_2^{n*}-T^n}\Big].$

, where $J_1 = exp[\beta ln\big(\beta(1+r)\big)]^{-\frac{1}{1+\beta}}$. The first condition implies that low income students are more likely to be constrained. The second inequality of the second condition implies that applying for selective universities is optimal without borrowing constraints. The first inequality of the second condition implies that if the cost of not being able to smooth consumption between two periods is too high, it is better not to apply for selective universities. Denoting $KK = (\big[M_1 + \lambda M_2^{s*} - T^s\big]\big[(1-\lambda(1+r))M_2^{s*}\big]^{\beta}\Big)$, in most cases $\frac{\partial ln(KK)}{\partial\lambda} = \frac{M_2^{s*}}{M_1+\lambda M_2^{s*}-T^s} - \beta\frac{1+r}{1-\lambda(1+r)} > 0$. Therefore as the extent of borrowing constraints becomes more stringent (smaller λ) , more constrained students would change their application decision and make suboptimal application choices.

4.9.2 Tournament Effect

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 a tuition subsidy would affect the intensive margin of the choices of students, the effort choices, 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

$$\underbrace{\frac{\partial P_{ad}}{\partial N_{AP}}(V^s - V^n)}_{(1)} + \underbrace{P_{ad}\frac{\partial(V^s - V^n)}{\partial N_{AP}}}_{(2)} + \underbrace{\frac{\partial V^n}{\partial N_{AP}}}_{(3)} = \frac{\partial U_{AP}}{\partial N_{AP}} \tag{4.3}$$

For students who do not apply for selective universities, the marginal benefit of effort is just Part (3). Other things being equal, the marginal benefit decreases as family income increases. If I focus on students who do not apply for selective universities, they would take fewer AP classes than applicants, and rich students would take fewer AP classes than comparable poor students because of the income effect.

Part (1) captures how additional AP classes 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 $j, k \in \{SAT, GPA, AP\}$, Part (1) can be rewritten as

$$\frac{\partial P_{ad}}{\partial N_{AP}}(V^s - V^n) = \underbrace{J\phi(g(D) - h^*)}_{\text{(I)}} \frac{1}{N_{AP}} \underbrace{\left[ln\Big(\frac{M_1 + \frac{1}{1+r}M_2^s - T^s}{M_1 + \frac{1}{1+r}M_2^n - T^n}\Big) + U_{sel} \right]}_{\text{(II)}} \tag{4.4}$$

Part (I) of equation (4.4) 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 (4.4) 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 borrowing constraints, 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 4.4 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 (4.4) 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 (4.4) 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-ability students, whereas it has the opposite impact on low-ability students.

On the other hand, changing 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 in more efforts, thus some of less academically able high-income students could remain competitive in the admission process.

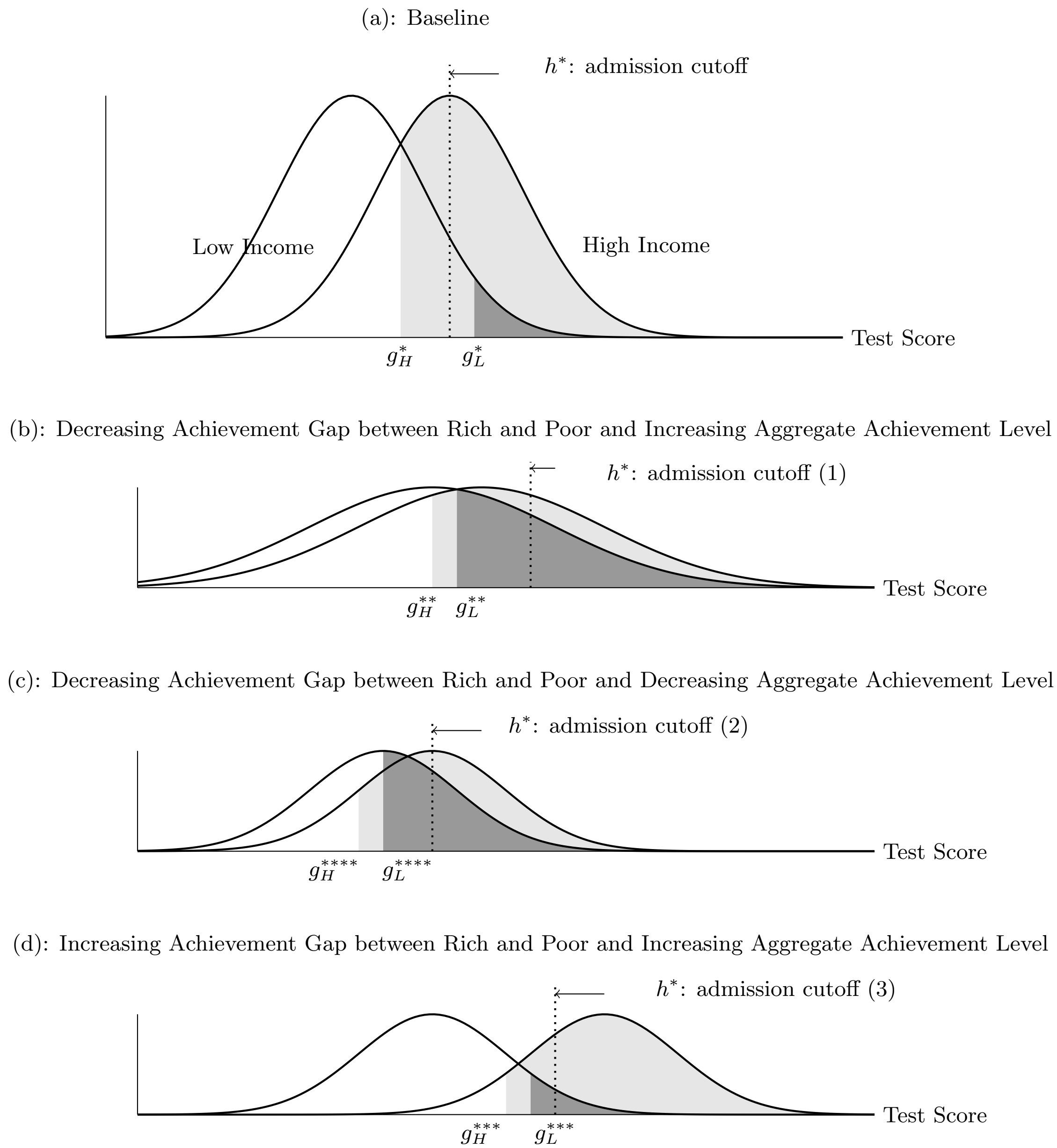


Figure 1: Increasing Need-Based Aid and the Distribution of the High School Achievement Score

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. If the tournament effect is more sensitive to ability than to family income, increasing need-based aid can reduce the achievement gap by raising the achievement of low-income high-ability students (b). If need-based aid attracts many low-ability low-income students and makes the admission process noisier, the achievement gap can decrease as high-ability students put less efforts (c). Elevated competition can make more high-income students attend private high school and get higher test scores, which can widen the achievement gap (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 $lnWage = F_M(\theta, H, I_{sel}, I_{BA}, \epsilon_w)$, and the admission probability $P_{ad} = \Phi(D)$. 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}), and the utility cost of taking AP classes (U_{AP}). To identify those preference components, I use the variation in students' choices: the amount of student loan, application decision, the number of AP classes taken by students, and the decision to attend private high school or not.

To quantify the impact of financial aid, it is essential to control the selection bias driven by unobservable characteristic 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 wage. Or when it comes to academic achievement, this characteristic might capture the time management skill or parents' guidance. This characteristic seems to have a strong correlation with family background because high-income students tend to obtain better educational and economic outcome than observationally equivalent low-income students in data.

To address this issue, 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. There are two ways to identify parameters associated with the distribution of θ. First, I use the geographic variations in tuition levels of attending home-state nonselective universities as an instrumental variable (Figure 3). In particular, I assume the following identification assumptions: (i) unobservable characteristic θ is not correlated with student's geographic location of residence conditional on family income and parents' educational attainment, (ii) students who do not attend selective universities attend home-state nonselective universities, (iii) labor market is perfectly mobile so that the wage rate is determined at the national level and tuition does not directly affect student's productivity. Under those assumptions I can compare two groups of students from two different states with the same observable characteristics. Although I do not observe θ at the individual level, in aggregate at the state level two groups would have the same expected value of θ. Because two groups are identical except for the direct cost of attending home-state nonselective universities, different application behaviors and other choices between two groups are supposed to be driven by differences in direct cost, not θ. Second, the structural model generates choices and outcomes of students conditional on all other structural parameters. If I can identify other parameters from other moment conditions, then the conditional mean of student's choices and outcomes can be used to identify the parameters associated with the distribution of θ.

To identify the intertemporal preference in U_c, I use three variables in the data: family income, labor earning, and the amount of the student's loan. In the model, all students have the same intertemporal preference. I identify intertemporal preference from the choices of high-income students who are not likely to be borrowing constrained. Next, 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 the student's home state and the 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 the 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 borrowing constraints in the financial market. First, borrowing constraints would decrease low-income students' consumption at college period. Without the 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 borrowing constraints λ. Second, borrowing constraints can also decrease the application rate of low-income students. If high ability low-income students did not apply for selective universities due to borrowing constraints, there are disproportionately more high ability students from low-income families among nonapplicants. Outcome measures, such as wage rate, and test scores can be used to determine 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 receive when attending selective universities and nonselective universities conditional on family income and test scores, (ii) the number of AP classes offered by their high school, (iii) the number of information sources about college application processes. I assume that those components are exogenous processes from the 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 of 7%.[33]

[33]The interest rate for an unsubsidized Federal Stafford Loan is 6.8% as of 2015 (6.83% for a PLUS loan). Although the interest rate of the subsidized loan is 3.4%, since private loan would cost more

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. For example, I simulate the application decision and effort choices for 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 moments are 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. There are 79 parameters and 101 moment conditions.

7 Results

Figure 15-20 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 observable ability. The estimated model tracks the observed pattern in data fairly well, especially the outcome variations across income quintile groups.[34] The model slightly understates the number of AP classes taken by students from the bottom quintile of income distribution. One potential reason is the impact of policy interruptions regarding AP program participation that might not be fully captured by controlling the number of AP classes offered by high schools.

7.1 Distribution of Unobservable Ability

Table 29 documents the estimation result of the distribution of unobservable learning ability. The mean of unobservable ability increases by 0.03 for a unit increase in the log family income and increases by 0.06 for a unit increase in the log of 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

(http://www.finaid.org/loans/privatestudentloans.phtml), I pick the interest rate for unsubsidized loan. Also, I do not consider repayment during college education.

[34] Because data on family income has categorical values, the number of students from the third quintile of income distribution is larger than other quintile groups.

increases the log wage rate by one-to-one. On average, students with a family income of $10,000 with college graduated parents get 3.7% higher wage rate than students with a family income of $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. θ also affects test scores, wage rate, admission probability, application cost, and nonpecuniary benefit.

7.2 The Benefit of Attending Selective Universities

Table 23 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 for SAT, AP score, GPA, and the unobservable characteristic of those students. The model assumes that the wage premium from attending selective universities is the same for every attendee.[35] For a full time worker who works 2,000 hours per year earns about $2,670 more if she attended selective universities on average.

Note that students repay the loan over 10 years after college graduation which is actually the standard plan chosen by majority students. To calculate the life-time pecuniary benefit of attending selective universities, I calculate the years to make a break-even point for each group. Table 33 shows the result. It takes 13.6 years for students from the top quintile of initial ability and from the bottom quintile of income distribution, whereas it takes 15.7 years for students with the same initial ability and from the top quintile of income distribution. This is calculated under the assumption that the wage premium does not change over time. Thus if the premium also increases over the life-cycle, the number of years to meet the break-even point would be shorter. The estimation result in this paper suggests that the pecuniary benefit accounting for expensive tuition is quite moderate. Interestingly, it takes longer for high-ability students to meet the break-even point than low-ability students because she has to give up receiving merit-based aid from nonselective universities.

Table 27 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 the student's home state and whether the student values the reputation of the college are significant factors. This result seems consistent with Fu (2014) which suggests that students with similar academic achievement scores have very different preferences regarding different type of colleges.

[35] Table 19 and Table 20 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.

7.3 The Extent of Borrowing Constraints

Table 29 shows the estimation result of the extent of liquidity constraint $\hat{\lambda} = 0.76$; students cannot borrow more than 76% of their expected annual income in each year of their college education. Note that I assume that students take student loans for four years of college education then they repay the total amount of loan over 10 years. Therefore, $\hat{\lambda} = 0.76$ actually implies that students need to repay 30% of their annual income over 10 years after college education.

Figure 28 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 than those from the second highest quintile of income distribution. The 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. Figure 22 shows the loan to expected income ratio the model predicts of students from the top and bottom quintile of income distribution. It suggests that a lot more students from low-income family borrow very close to the amount of borrowing limits. This result suggests that even after taking into account generous financial aid, difficulty in financing college cost can be an important problem for low-income students.

8 Counterfactual Analysis

First, I discuss the impact of borrowing constraints on the choices of low- and high-income high school students. Second, I discuss the quantitative importance of students' initial characteristics to explain the choice differences between low- and high-income students. Third, I discuss the respective roles of features of selective universities on the choice difference between rich and poor students. Fourth, I discuss the impact of doubling need-based aid from selective universities for attendees from the bottom quintile of distribution on the choices of students, the achievement gap, and aggregate effort level. Finally, I compare the impact of increasing need-based aid from selective universities to alternative policies such as merit-based aid, changes in the admission criteria (the Texas Top Ten Law), increasing need-based aid from nonselective universities, and income-based affirmative action (income quota). In doing so, I simulate the model over 100 times so that under the new cut-off value of the admission criteria the number of attendees is equalized to the number of seats in selective universities on average.

8.1 The Impact of Borrowing Constraints on Students' Choices

Table 36 and 37 summarize the impact of borrowing constraints on the choice and outcome of low-income students and the gap between rich and poor students. The impact of borrowing constraints on high ability, low-income students is substantial. First, borrowing constraints decreases the number of low-income applicants by 6.9%. Focusing on students within the top quintile of initial test score distribution, borrowing constraints increase the effort gap by 14.0%, the achievement gap by 8.4%, and the wage gap by 4.7%. If the financial aid from selective universities were randomly allocated across attendees, it could have increased the effort gap by 22%, the gap in the number of applicants by 3%, the gap in SAT score by 13%, and the gap in wage rate by 7%. Therefore, need-based aid from selective universities play an important role in reducing effort gap, but still there are constrained low-income high ability students.

8.2 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 38 and Figure 27 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, the difference in the initial math score is the most important indication that why rich students take more AP classes and apply more often 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.

8.3 Selective Universities and Choices of Students

Selective universities have several features that can affect students' choices: pecuniary/nonpecuniary benefit, application cost, a competitive admission process based on test scores, need-based financial aid, and expensive tuition. Table 40 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 of high-income students by 4%. This suggests that some of the less academically able high-income students benefit from the 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%.

Figure 23 and Figure 24 show how student's effort level changes as the admission probability among applicants decreases as the capacity in selective universities decreases. The tournament effect—additional effort applicants put to increase admission probability—increases as it becomes more difficult to get an admission from selective universities. The average effort level shows a non-monotonic relationship with the admission probability because if the capacity is too small, more students are discouraged from applying for selective universities.

Figure 25 and Figure 26 describe how student's own characteristics and environmental factors such as tuition and test-based admission process account for the choice difference between rich and poor students. In this counterfactual analysis, I consider two students from the top and bottom income distribution and have the same population average observable characteristics except for family income. In the baseline, I assume that tuition is free, the admission process is random, there is borrowing constraints, and no heterogeneity in the unobservable characteristic θ. Even in this case, high-income students would apply more for selective universities because they can afford private high school which reduces application costs. Current tuition does not affect application behavior of high-income students, while it reduces the application rate of low-income students slightly. Test-based competition reduces the application rate of high-income students by 9% because it requires extra efforts for higher admission probability and making additional effort is costly. Introducing borrowing constraints widens the gap in application rate by 3%, whereas heterogeneity in θ further increases the gap by 6%. This suggests that although borrowing constraints contribute to the different application behavior between rich and poor students, heterogeneity in the unobservable characteristic θ is quantitatively more important to explain the disparity. Figure 26 shows the similar patterns for the effort gap between rich and poor students. Test-based competition in the admission process is important to make high-income students put more efforts. Similar to the application rate, heterogeneous θ is quantitatively much more important than borrowing constraints to explain the effort gap between rich and poor students. This analysis suggests that increasing need-based aid could decrease the achievement gap because it can relax borrowing constraints, however, its impact would be limited because the unobservable characteristic plays an important role in shaping student's behaviors.

8.4 Increasing Need-Based Aid from Selective Universities

If selective universities double the amount of grants per attending student from the bottom quintile of the income distribution, the average effort level of those low-income high school students increases by 13.4%. the effort gap decreases by 33.4%, the achievement gap by 20.2%, and the wage gap by 10.2% of students with initial test scores in top 20th percentile in 10th grade. The aggregate achievement score also increases because elevated competition raises the effort level of high ability applicants from all income backgrounds. It increases the aggregate effort level by 0.7%. Elevated competition increases the effort level of high ability students from all income background. In particular, the effort level of low-income applicants increases by 2.6% and the effort level of applicants from other income backgrounds by 0.5%. Focusing on those who actually increase the effort level, low-income students increases by 52% and high-income students increase by 12%. Because low-income high-ability students respond more sensitively to the change, doubling need-based aid can reduce the achievement gap.

Increasing need-based aid can also increase the application rate of students from the bottom quintile income distribution by 14.1%. Among those low-income students who changed their application decision, 53% attend selective universities. Relaxing borrowing contraints explain half of the increase in the application rate of low-income students. The remaining half is explained by the income effect—because the pecuniary benefit is too small to compensate for the expensive tuition, only those who already have a lot of money to spend want to attend selective universities and obtain the nonpecuniary benefit.

Table 39 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 a change in the academic qualification of the applicant pool. The effort reduction is much larger for students with a low math score (8%) compared to students with a 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 their application decision, increased need-based aid from selective universities would also improve the academic quality of applicants. This 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 talented high school students have a better opportunity for attending selective universities but also it can increase the aggregate achievement level by raising effort choices of high ability applicants from all income background.

8.5 Alternative Policies

I compare a couple of policies in terms of their impacts on low-income student's academic achievement. First, doubling need-based aid can close the achievement gap better than merit-based aid, while keeping the aggregate achievement level as almost the same. Second, changing admission criteria similar to the Texas Top 10 Law, using GPA only to determine the admission result, is much more effective to reduce the achievement gap than doubling need-based aid from selective universities. However, the Texas Top 10 Law decreases the overall effort level of students by half, which corresponds to a 33 point reduction in SAT the score. This suggests the importance of student's effort choice to evaluate the impact of college admission criteria on the aggregate achievement level. Third, increasing need-based aid from selective universities is more effective per dollar than the same policy implemented from nonselective universities. This is because need-based aid from selective universities targets high-ability low-income students whose effort choices would be the most elastic to tuition subsidies. By targeting those students, increasing need-based aid from selective universities can be much more effective per dollar than other policy that target all low-income students. Table 42 and 43 summarize the outcomes of low-income students and the gap in outcome between students from the top and the bottom quintile of income distribution under different counterfactual experiments.

8.5.1 Merit-Based Aid

Selective universities have awarded financial aid mostly based on student's need. However, recently, more selective universities started to award aid for middle- and high-income students by redefining need-based aid or increasing merit-based aid. Because merit-based aid is more frequently discussed as a way of providing incentives to students for higher academic achievement, comparing the impact of need-based aid from merit-based aid can be informative to understand the incentive aspect of need-based aid from selective universities.

I consider the counterfactual financial aid policy such that selective universities provide only merit-based aid. I use the merit-based aid component from nonselective universities as a benchmark to determine how aid is awarded based on student's test scores. Also, I keep the budget spending on merit-based aid as the same as the budget required to doulbe need-based aid.

The result suggests that doubling need-based aid can close the achievement gap better than merit-based aid by 21.1% among high-ability students. The aggregate achievement level is higher under merit-based aid case but the difference is less than 1%. Merit-based aid increases the effort level of low-income high-ability students by 1.5%, whereas doubling need-based aid can increases the effort level of low-income high-ability students by 11.3%. Merit-based aid increases the effort level of high-income high-ability students by 1.1%. As a result, merit-based aid results in 33.4% larger effort gap and 10.7% larger wage gap between rich and poor students with high initial ability compared to doubling need-based aid.

8.5.2 The Texas Top 10 Law

As an alternative to the race-based affirmative action in the college admission process in 1997 Texas introduced the Texas Top 10 Law. This law guaranteed an automatic admission to flagship universities in Texas for students who ranked in the top 10% of their classes in a Texas high school. Motivated by this policy change, I evaluate the impact of making admission criteria similar to the Texas Top 10 Law such that only the GPA is taken into account in the admission process. This change does not need extra resources to implement.

The Texas Top 10 Law can substantially reduce the achievement gap among students in the top quintile of the initial test scores, as measured by SAT scores, by 39.5%. However, the aggregate effort level decreases by 49.2%, which corresponds to a 33 point reduction in the average SAT score and $1,300 dollar decrease in annual average income of low-income high-ability students. Because taking AP classes is less effective in obtaining higher GPA, even high ability students take many fewer AP classes during high school. This highlights the importance of students' effort choices when it comes to evaluating changing college admission criteria.

8.5.3 Increasing Need-Based Aid from Nonselective Universities

I consider the case in which nonselective universities raise need-based aid by $12,000 for students from the bottom quintile of income distribution. This decreases the number of low-income applicants by 7% and the number of AP classes low-income students take by 11%. Focusing on low-income students within the 20th percentile of initial test score distribution, increasing need-based aid from nonselective universities decreases the number of applicants by 6% and the number of AP classes taken by students by 8%. It also increases the effort gap between rich and poor students with top quintile distribution of the initial ability by 24%, the gap in the SAT score by 15%, the gap in wage rate by 7%. Although this policy might target students at the margin of college attainment decision, it could have unintended results on high ability low-income students by leading them to apply less frequently to selective universities.

8.5.4 Income Quota

As an alternative to race-based affirmative action in the college admission process, income-based affirmative action is often discussed.[36] Motivated by these discussions, I consider following the counterfactual experiment; the income quota system. The system gives additional points to the lowest income quintile students in the admission process so that the number of lowest income quintile students doubles.

Table 42-43 suggest that the income-based affirmative action, the income quota system, can increase the number of low-income applicants by 59%. However, students with low initial test score respond more sensitively such that the number of low-income high ability applicants increases by 27%. It also increase the number of AP classes taken by low-income by 14%, but it slightly reduces the effort level of low-income high ability students. The gap between rich and poor students decreased more than increasing need-based aid from selective universities. For example, it decreases the effort gap by 7%, the gap in the number of applicants by the 12%, the gap in SAT score by 20%, the wage gap by 23%. Similar to the Texas Top 10 Law, income quota system would be more effective to reduce the achievement gap between rich and poor students, but it comes with a large reduction in the aggregate achievement level.

8.5.5 Providing More AP Classes

There has been increasing policy interests in expanding AP program to low-income students. To compare its impact with increasing need based financial aid from selective universities, I consider the following counterfactual. First, I compute the total budget required to implement increasing need based financial aid from selective universities. Then, based on the estimated start up cost for average AP class documented by College Board [37], I compute how many AP classes can be newly offered by high schools in which students from lowest income quintile family attend. It turns out that with the budget required to increase financial aid for lowest income quintile students who attend selective universities, 1.70 AP classes can be offered to every student from the bottom quintile of income distribution.

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

[37] 4,343 dollar per 25 size class

First, low-income high-ability students take 6% more AP classes if more AP classes are offered by the high school. It does not affect choices of students from other income quintile groups. Second, it is less effective compared to doubling need-based aid from selective universities in increasing the effort level of low-income students. This is because expanding AP classes targets all low-income students, whereas need-based aid from selective universities focuses on low-income high-ability.

9 Out-Of-Sample Prediction

To examine the validity of the structural model, I consider an out-of-sample prediction. The National Education Longitudinal Study of 1988 is a nationally representative sample of 8th graders. 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' points of view, 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 closely the predicted outcome tracks the actual data observed in the earlier cohort, the NELS1988.

Figure 29-30 show the result of the 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 ability 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 in selective universities 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 predicts much smaller disparity between students with high- and low-ability. In particular, the data shows a stark difference in the composition rates between students from the first and the second highest quintile of initial ability distribution. A more difficult application process and a much lower number of available AP classes in the NELS1988 cohort may explain why the selection into selective universities in the earlier cohort was more strongly driven by the student's initial ability.

10 Conclusion

In this paper, I examine whether increasing need-based aid from selective universities can reduce the achievement gap between rich and poor students conditional on the initial ability. I find that although low-income students receive more than twice the amount of aid from selective universities as high-income students, borrowing constraints bind for low-income high ability students. These constraints lead to an increasing achievement gap between rich and poor students with the same initial ability. Additional need-based aid from selective universities for low-income students can not only reduce the achievement gap between rich and poor, but also increase the achievement level of all high-ability students. I find that need-based aid can close the achievement gap better than merit-based aid, while keeping the aggregate achievement level of the entire population almost the same.

Important things remain for 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 change their financial aid policies. Thus, incorporating competition between colleges could be a 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 as well.

References

[1] Abbott, B., Gallipoli, G., Meghir, C., and L.Violante, G. Education policy and intergenerational transfer in equilibrium. *working paper* (2013).

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

[3] Arcidiacono, P., Aucejo, E., Coate, P., and Hotz, J. Affirmative action and university fit: Evidence from proposition 209. *working paper* (2014).

[4] Autor, D., and Scarborough, D. Does job testing harm minority workers? evidence from retail establishments. *Quarterly Journal of Economics 123(1)* (2008), 219–277.

[5] Avery, C., Hoxby, C., Jackson, C., Burek, K., Poppe, G., and Raman, M. Cost should be no barrier: An evaluation of the first year of harvard's financial aid initiative. *working paper* (2006).

[6] Cameron, S., and Heckman, J. Life cycle schooling and daynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy 106* (1998), 262–333.

[7] Cameron, S., and Heckman, J. The dynamics of educational attainment for black, hispanic, and white males. *Journal of Political Economy 109* (2001), 455–499.

[8] Cameron, S., and Taber, C. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy 112* (2004), 132–182.

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

[10] Chatterjee, S., and Ionescu, F. Financial aid policy: Lessons from research. *working paper* (2010).

[11] Cunha, F., and Heckman, J. The technology of skill formation. *working paper* (2007).

[12] Dale, S., and Krueger, A. Estimating the payoff to attending a more selective college: An applicaion of selection on observables and unobservables. *Quarterly Journal of Economics 117(4)* (2002), 1491–1527.

[13] Dynarski, S. Hope for whom? financial aid for the middle class and its impact on college attendance. *National Tax Journal September* (2010), 629–661.

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

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

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

[17] Hawkins, D., and Clinedinst, M. State of college admission. *National Association for College Admission Counseling* (2006).

[18] Heathcote, J., Storesletten, K., and Violante, G. L. Optimal tax progressivity: An analytic framework. *working paper* (2014).

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

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

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

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

[23] Hoxby, C., and Avery, C. The missing one-offs: The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution 46 (Spring)* (2013), 1–65.

[24] Hoxby, C., and Turner, S. Expanding college opportunities for high-achieving, low-income students. *working paper* (2013).

[25] Hoxby, C., and Turner, S. Expanding college opportunities for high achieving, low income students. *working paper* (2013).

[26] Kane, T. A quasi-experimental estimate of the impact of financial aid on college-going. *working paper* (2003).

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

[28] Kinsler, J., and Pavan, R. Family income and higher education choices: The importance of accounting for college quality. *Journal of Human Capital 5(4)* (2011), 453–477.

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

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

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

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

[33] Lucca, David, T. N., and Shen, K. 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).

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

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

[36] Rivkin, S., Hanushek, E., and Kain, J. Teachers, schools, and academic achievement. *Econometrica 73* (2005), 417–458.

[37] Rosen, S., and Willis, R. Education and self-selection. *Journal of Political Economy 87(5)* (1979), 7–36.

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

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

[40] Stinerbrickener, R., and Stinerbrickner, T. 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.

[41] Stinerbrickener, R., and Stinerbrickner, T. Learning about academic ability and the college drop-out decision. *working paper* (2009).

[42] Todd, P., and Wolpin, K. The production of cognitive achievement in children: Home, school, and racial test score gpas. *Journal of Human Capital 1* (2007), 91–136.

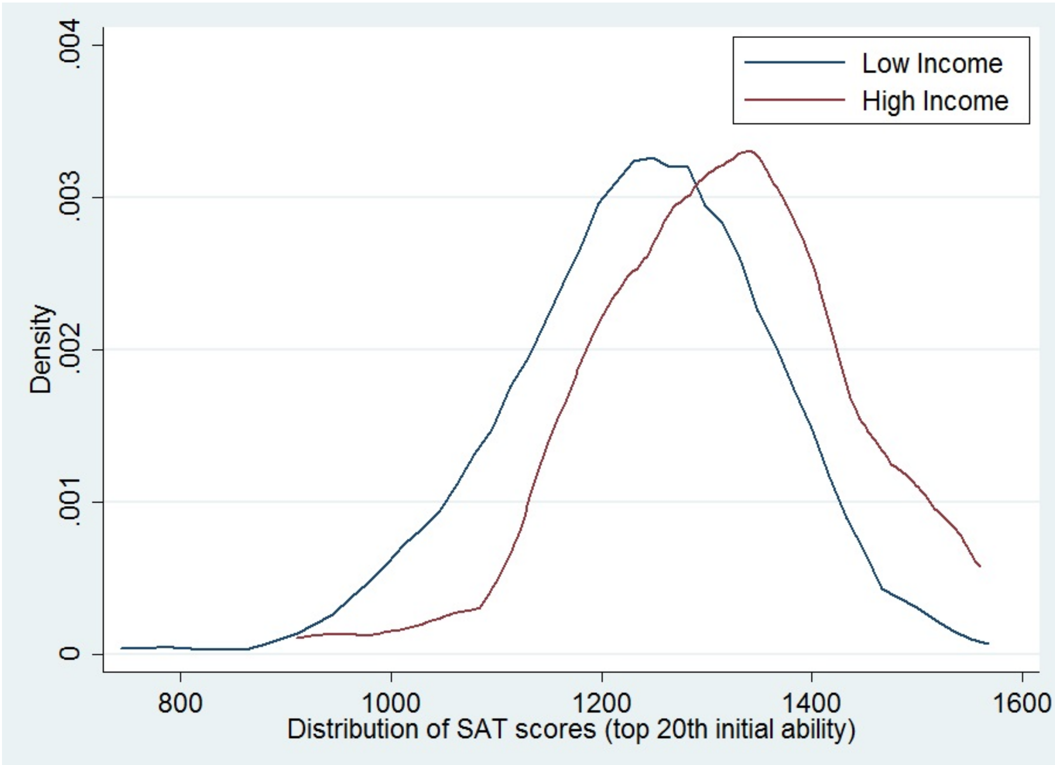
[43] van der Klaauw, W. Estimating the effect of financial aid offers on college enrollment: A regression-discountinuity approach. *International Economic Review 43* (2002), 1249–1287.

[44] Woo, J., and Soldner, M. 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).

Figures and Tables

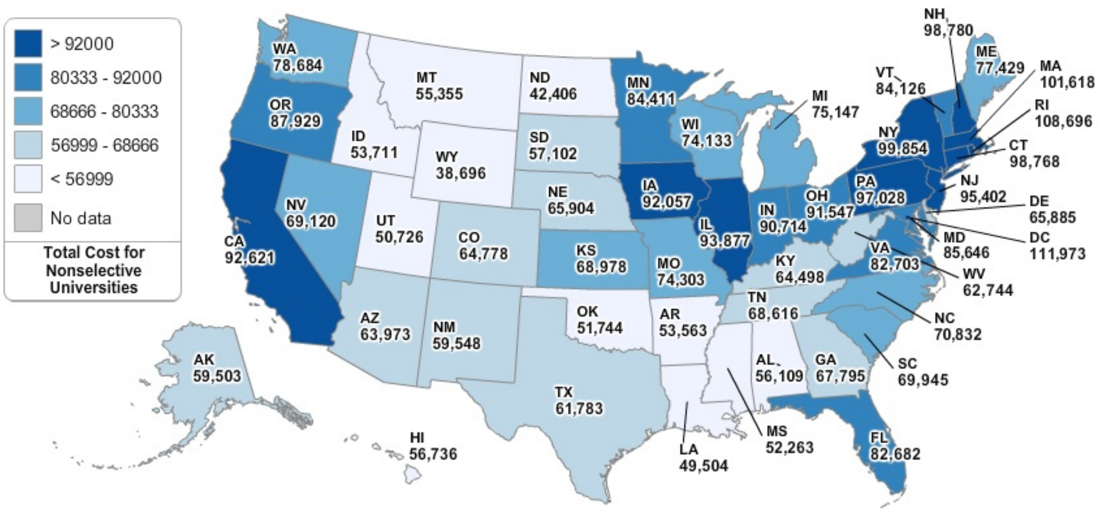
Motivating Facts (Figures)

Figure 2: SAT Score Distribution of Low- and High-Income Students within the Top 20th Initial Ability



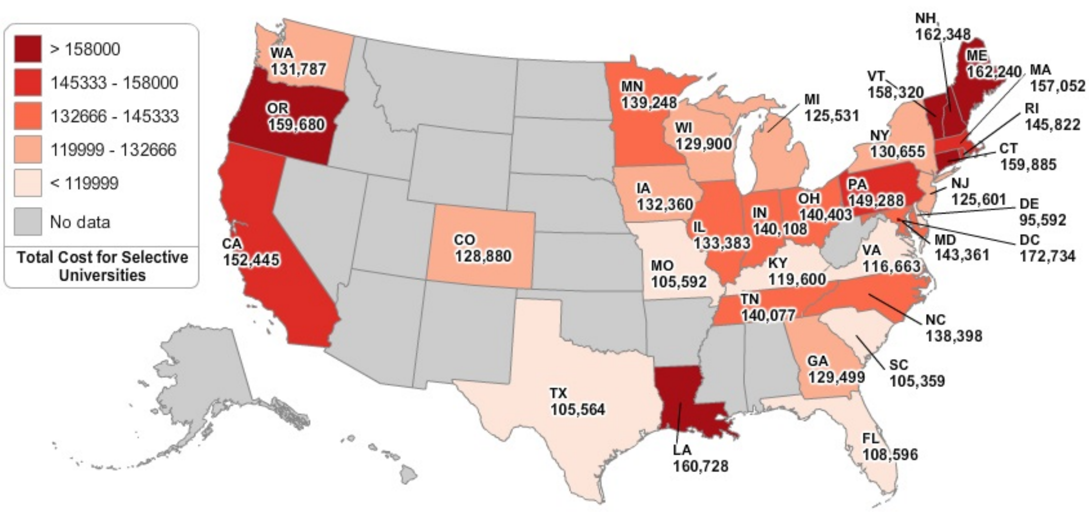
Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). The graph shows the distribution of SAT score of low- and high-income students who belong to the top 20th percentile of the initial ability distribution. The low- and high-income students refer to students from the bottom and top income distribution. Initial ability is measured by the standardized math test score in 10th grade.

Figure 3: Total Direct Cost of Attending Nonselective Universities in Home States for Four Years



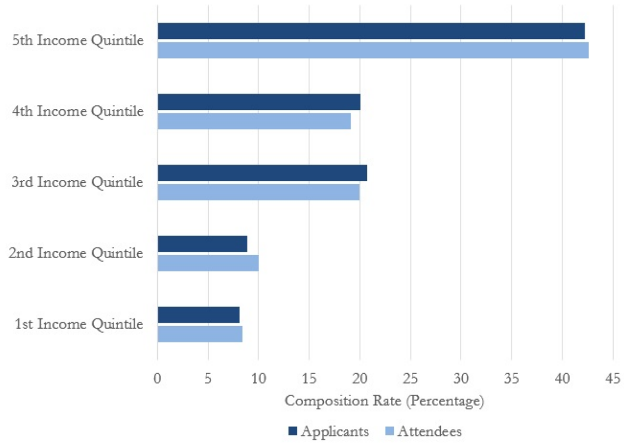
Note. Data comes from the Integrated Postsecondary Education Data System 2004. The total cost includes tuition/fees and the cost for room, board, and book. Regarding the cost of room and board, I use the cost of living in campus instead of the cots of living with family.

Figure 4: Total Direct Cost of Attending Selective Universities in Out-of-Home States for Four Years



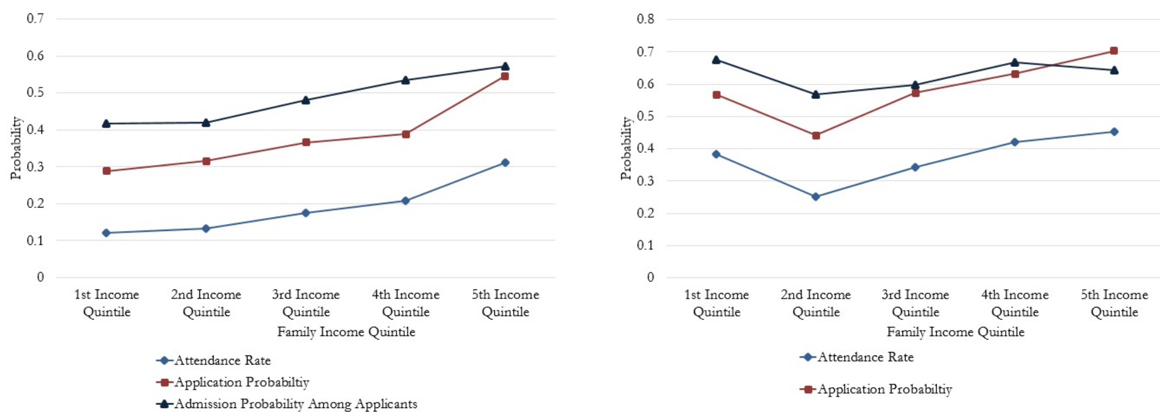
Note. Data comes from the Integrated Postsecondary Education Data System 2004 (IPEDS 2004). The total cost includes tuition/fees and the cost for room, board, and book. Regarding the cost of room and board, I use the cost of living in campus.

Figure 5: Composition Rate of Attendees and Applicants in Selective Universities By Income Quintiles



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

Figure 8: Attendance Rate, Application Rate, and Admission Rate upon Application

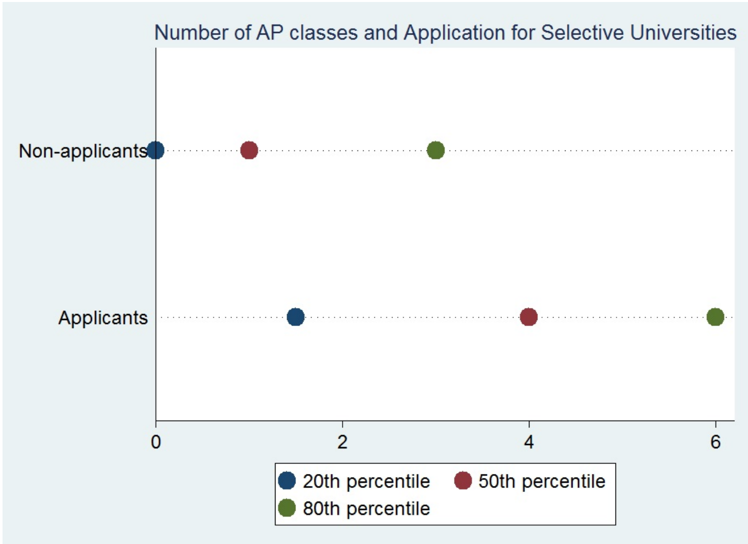


(a) All Students

(b) High-Achieving Students with Top 20th Initial Ability

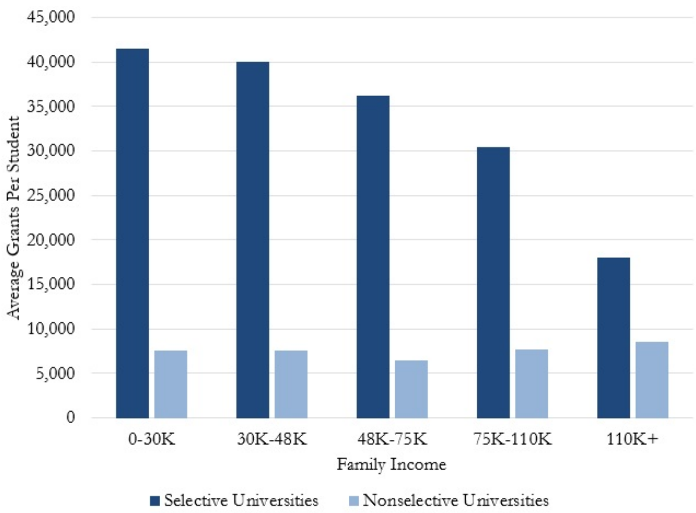
Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). The unit is percent. Figure 9(a) is based on the entire sample, and Figure 9(b) is based on high-achieving students who belongs to top 20 math score at the sophomore year of the high school. Attendance rate (application probability) 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 for each income quintile group.

Figure 9: Number of AP Classes Taken by Applicants and Nonapplicants



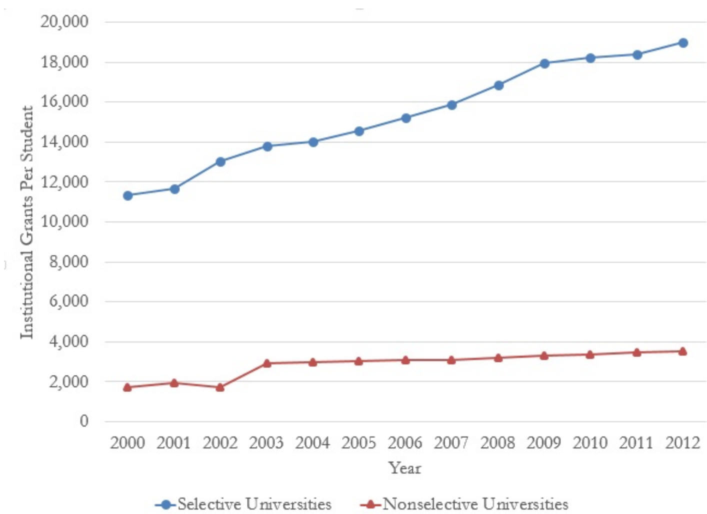
Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). I plot the 20th, 50th, and 80th percentile of the distribution of the number of AP classes taken by applicants and non-applicants. The median number of AP classes taken by applicants is more than three times larger than that of nonapplicants.

Figure 10: Average Grants Given to Attendees by Family Income



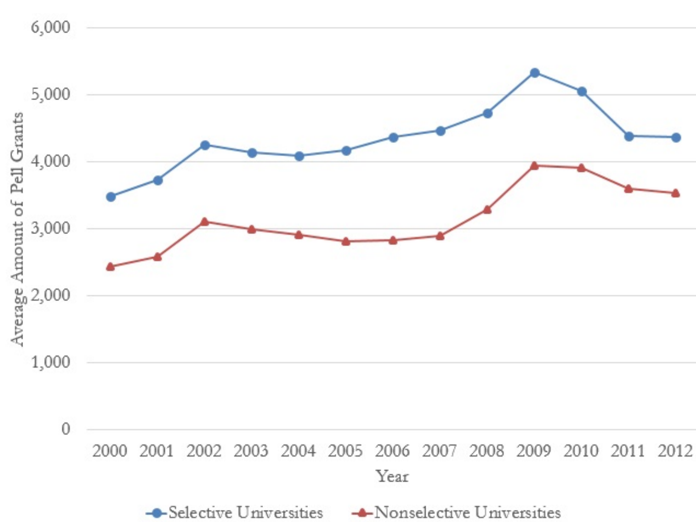
Note. Data comes from the IPEDS of 2012-2013 academic year. The unit is 2012 dollar. The total grant includes the Federal grants, state grants, and institutional grants. The income categorization follows that of IPEDS data. Selective universities are four-year colleges in the U.S. that belong to the top two categories of the Barron's Index 2004, whereas nonselective universities account for the rest of the U.S. four-year colleges. I use the IPEDS of 2012-2013 academic year because the IPEDS of 2002-2003 academic year does not have such information, the amount of average grant given to students reported by each institution.

Figure 11: Trend in the Amount of the Institutional Grants Per Student from Selective and Nonselective Universities



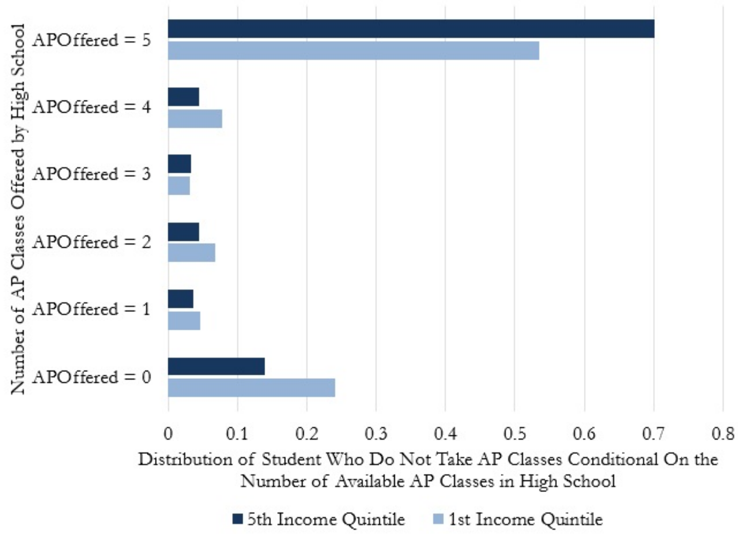
Note. Data comes from the Integrated Postsecondary Education System (IPEDS) 2000-2012. I report the average amount of grants offered by the institution the student attends. The unit is 2004 Dollar. Selective universities are 4-year colleges with top 2 categories of NCES-Barron's Admission Competitive Index, which account for about 20% of entire enrollees of 4-year colleges.

Figure 12: Trend in the Amount of the Average Federal Grants



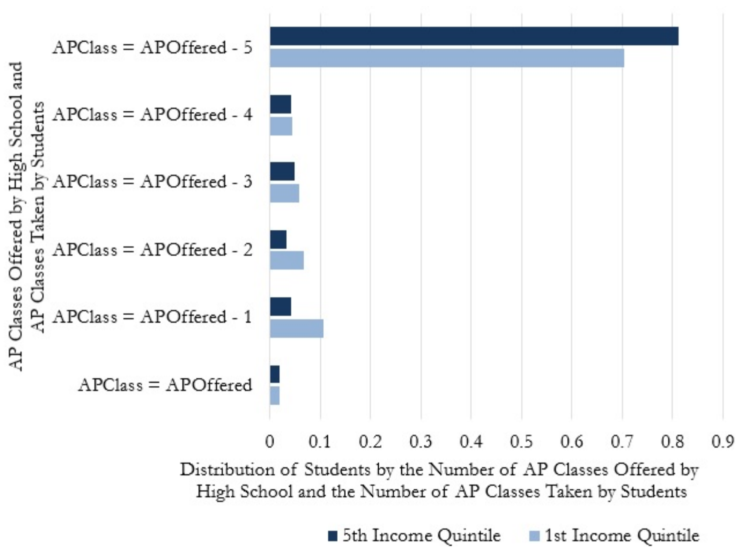
Note. Data comes from the Integrated Postsecondary Education System (IPEDS) 2000-2012. The unit is 2004 Dollar. Federal grants includes Pell Grants, Federal Supplemental Educational Opportunity Grants (FSEOG), Teacher Education Assistance for College and Higher Education (TEACH) Grants, Iraq and Afghanistan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

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



Note. Data comes from the High School Transcript of the Education Longitudinal Study of 2002 (ELS2002). The first income quintile indicates the bottom quintile of income distribution. Focusing on students who do not take AP classes, this graph shows the number of AP classes offered by high school in which those students attend.

Figure 14: Number of AP Classes offered by High School and Number of AP Classes Taken By Students



Note. Data comes from the High School Transcript of the Education Longitudinal Study of 2002 (ELS2002). I focus on students who take at least one AP Class. The graph shows the number of AP classes offered by high school in which those students attend. The first income quintile indicates the bottom quintile of income distribution.

Motivation (Tables)

Table 1: Achievement Gap conditional on Initial Test Scores

Variable	lnSAT	lnSAT
Black	-0.082*** (0.008)	-0.077*** (0.007)
Asian	0.047*** (0.007)	-0.011* (0.006)
Hispanic	-0.036*** (0.008)	-0.056*** (0.007)
Female	-0.005 (0.004)	-0.014*** (0.004)
ln(A)	0.092*** (0.002)	0.073*** (0.002)
lnfincome	0.037*** (0.003)	0.026*** (0.002)
ln(APclass+1)		0.096*** (0.003)
Const	6.243*** (0.031)	6.374*** (0.028)
Num Obs.	4,080	4,080
R-square	0.385	0.523

Note. The sample consists of students who attend four-year colleges. This is the OLS estimation for the log SAT score (sum of verbal and math SAT score). 'ln(A)' indicates the standardized math score in grade 10th. 'lnfincome' indicates the log of family income. Since almost half of students did not take AP tests, I use $ln(AP+1)$ instead of $ln(AP)$. Standard errors are in the parentheses. Without student's own effort choice, as measured by the number of AP classes, the R-square decreases by 28%.

Table 2: Log Hourly Wage I

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***
ln(A)	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*
ln(APclass+1)	0.069***	0.057***	0.026
ln(A)	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 comes from the Education Longitudinal Study of 2002 (ELS2002). This is a OLS estimation of the log of the hourly wage rate on students' demographic, family background, initial math score (ln(A)), 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 (that measures how much the student values monetary return in the future).

Table 3: Log Hourly Wage II

Parameter	(1)
Black	-0.070***
Asian	0.010
Hispanic	0.007
Female	-0.043***
lnFincome	0.039***
$\ln(P_{Edu})$	-0.086
lnSAT	0.166***
lnGPA	0.125**
ln(APscore+1)	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***
Dropout BA	-0.189***
Graduate School	0.087***
Constant	1.302***

Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). This is a OLS estimation of the log of hourly wage rate. I further control major choice, college dropout, and whether to attend graduate school or not. 'P_{Edu}' indicates the average years of schooling of parents. 'Dropout BA' indicates whether the student drops out of college, whereas 'Graduate School' indicates whether the student attends graduate schools.

Table 4: Log Hourly Wage III - 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 comes from the National Longitudinal Study of Youth 1979 (NLSY79). This is a pooling OLS estimation of the log hourly. I divide the annual income by the annual working hours to get the hourly wage rate. Column (4) suggests that even after controlling for the AFQT score and its interaction with the age, the SAT score still has a significantly positive correlation with the hourly wage rate. However, the SAT score does not have significant age effects on the wage rate after controlling for the AFQT score and its interaction with the age.

Table 5: Different Effort Measures and SAT Scores

	(1)	(2)	(3)
Black	-44.72***	-43.08***	-42.68***
Asian	-17.15***	22.08***	29.97***
Hispanic	-39.94***	-17.15**	-14.72*
Female	-0.90	3.40	8.14*
ln(family income)	6.72**	10.68***	10.56***
ln(Parent's education)	138.35***	185.24***	195.60***
ln(Initial Score)	112.45***	132.51***	134.85***
public	-34.01***	-20.44***	-23.51***
Unit Math	5.31**	13.56***	13.98***
Unit English	-1.04	1.81	1.11
Extracurricular	-0.31	-0.27	0.04
Homework	0.82***	2.48***	
AP class	26.57***		
Const	172.67***	-115.30**	-130.07***
R-square	0.603	0.523	0.517

Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). The dependent variable is the sum of math and verbal SAT score. I consider 5 types of effort choices. 'Unit Math' and 'Unit English' are the Carnegie units of the total math and English classes students take during high school. 'Extracurricular' is the average weekly hours students spend on extracurricular activities. 'Homework' is the average weekly hours spending on homework. 'AP class' is the number of total AP/IB classes students take during high school. Once I take into account the number of AP classes, the coefficient of math classes and that of homework hours decreased almost by 60%.

Table 6: Student Composition in Selective Universities

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

Note. Data comes 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

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

Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). 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: Average Tuition by College Selectivity

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	24,824(7,497)	22,749 (10,140)	12,492 (7,131)	35,241
Nonselective	15,061(5,098)	12,492 (7,131)	7,180 (1,651)	22,241

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

Table 9: Grants as Proportion 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 comes 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 10: Changes in Financial Aid Policy of 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 11: Changes in Financial Aid Policy of 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	All Families	2008-2009
	(\$4,500 contribution by earning from work)		
	No parent contribution,	Less than \$60,000	2008-2009
	No tuition/room/board		
	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 comes from `http://www.finaid.org/questions/noloansforlowincome.phtml`

Table 12: Recent Changes in Merit-Based Aid

Colleges	Ave Merit-Aid	Change from 2007-8
Stanford Univ.	\$5,085	31%
Vanderbilt Univ.	\$24,505	78%
Boston Univ.	\$19,960	39%
Carnegie Mellon Univ.	\$8,293	-20%
Univ. of Chicago	\$11,636	-19%
Washington Univ. in St. Louis	\$8,803	-13%
George Washington Univ.	\$18,495	-9%

Note. Data comes from
`http://www.nytimes.com/interactive/2012/07/08/education/edlife/8edlife_chart.html`

Table 13: 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
ln(A)	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 (ln(A)), and curriculum choice (lnUNITMath: log of the number of total math credit (Carnegie unit), and lnAPclass: log number of AP classes students take).

Table 14: College Drop Out

	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 comes from the Education Longitudinal Study of 2002 (ELS2002). This is a probit estimation of the college drop out decision on students' demographic, family income, test scores, and whether the student attended selective universities ($Dummy_{Sel}$).

Table 15: Application Rate, Home State Tuition, and Information Sources

	Probit	ME
Black	0.547***	0.164***
Asian	0.653***	0.196***
Hispanic	0.726***	0.218***
Female	-0.007	-0.002
lnFincome	0.207***	0.062***
lnParEdu	1.255***	0.377***
lnSAT	2.514***	0.755***
lnGPA	0.021	0.006
lnAPscore	0.409***	0.123***
HomeTuition(10K)	0.184***	0.055***
HomeTuition× DummyLowIncome	0.043**	0.013***
Information Sources	0.036***	0.011***
public	-0.118**	-0.036***
Action Control	0.067**	0.020**
Importance of Reputation	0.195***	0.059***
Const	-25.612***	

Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). This is a probit estimation of applying for selective universities on students' demographic, family income, test scores, and the average cost of home-state nonselective universities, and the number of information sources about college application processes.

Table 16: Application, Attendance and Admission Rate

	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 comes 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 17: Number of Information Sources Available Regarding College Application Processes

Variable	(1)	(2)
Black	0.356***(0.111)	0.370*** (0.111)
Asian	0.255***(0.100)	0.296*** (0.101)
Hispanic	-0.090(0.122)	-0.092*** (0.122)
Female	-0.027(0.061)	-0.014 (0.061)
lnfincome	0.135***(0.038)	0.111*** (0.039)
public		-0.181*** (0.065)
Const	2.596***(0.434)	2.964*** (0.453)
Num Obs.	3,442	3,442
R-square	0.007	0.008

Note. Data comes from the Education Longitudinal Study of 2002 (ELS2002). The sample consists of students who attend four-year colleges. The dependent variable is the number of information sources available to students during high school. 'lnfincome' indicates the log of family income. Since almost half of students did not take AP tests, I use $ln(AP+1)$ instead of $ln(AP)$. Standard errors are in the parentheses.

Table 18: 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 comes 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 19: Potential Heterogeneity in Returns from Attending Selective Universities by Family Income

	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 comes from the Education Longitudinal Study of 2002 (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 20: Matching Efficiency

	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 comes from the Education Longitudinal Study of 2002 (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.

First Stage Estimation

Table 21: First Stage Estimates—Amount of Financial Aid

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 comes from the Education Longitudinal Study of 2002 (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 grants only provide need-based aid, whereas nonselective universities provide both need-based and merit-based aid.

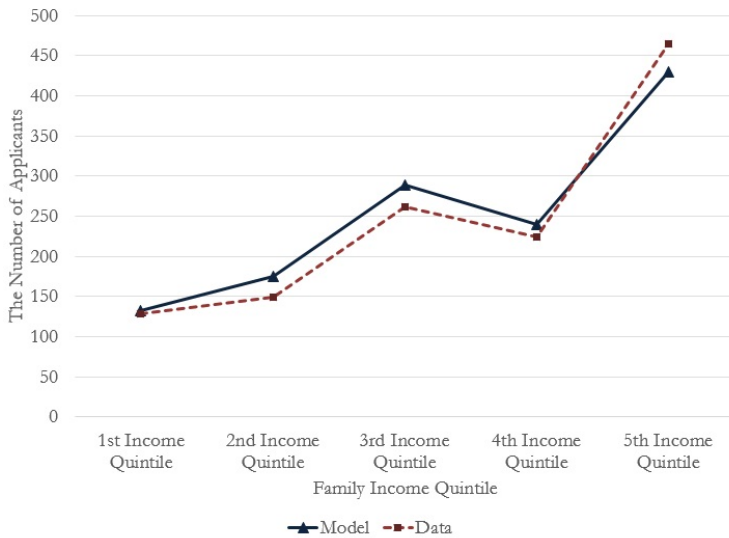
Table 22: First Stage Estimates—Number of AP Classes Offered by High School/Number of Information Sources Available to Students

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 comes from the Education Longitudinal Study of 2002 (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.

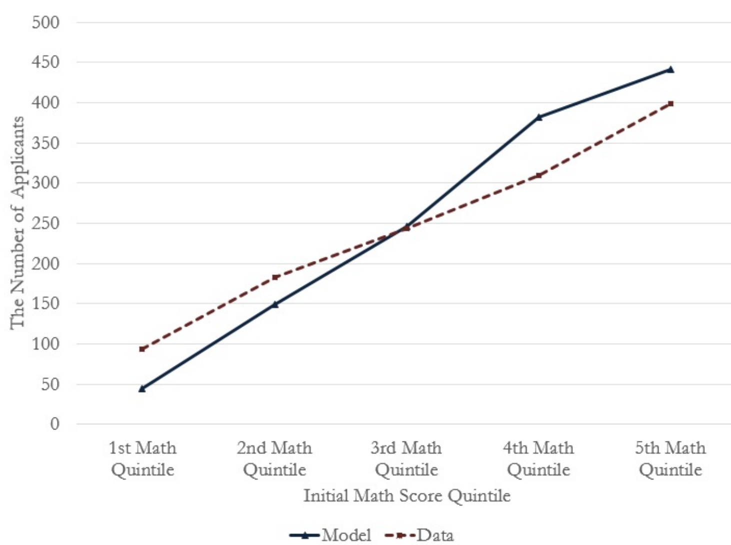
Estimation Results (Figures)

Figure 15: Model Fit: Number of Applicants for Selective Universities by Family Income Quintile



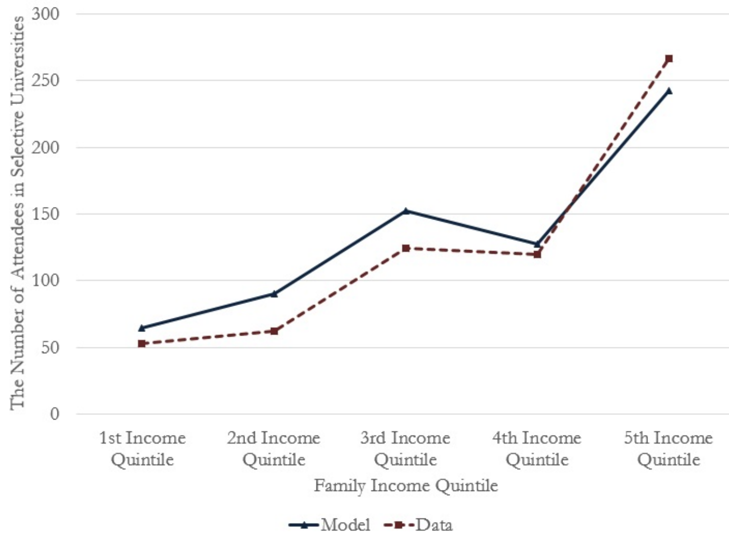
Note. The graph shows the predicted number of applicants by family income quintile from the estimated model the observed number of applicants by family income quintile from the data.

Figure 16: Model Fit: Number of Applicants for Selective Universities by Initial Ability Quintile



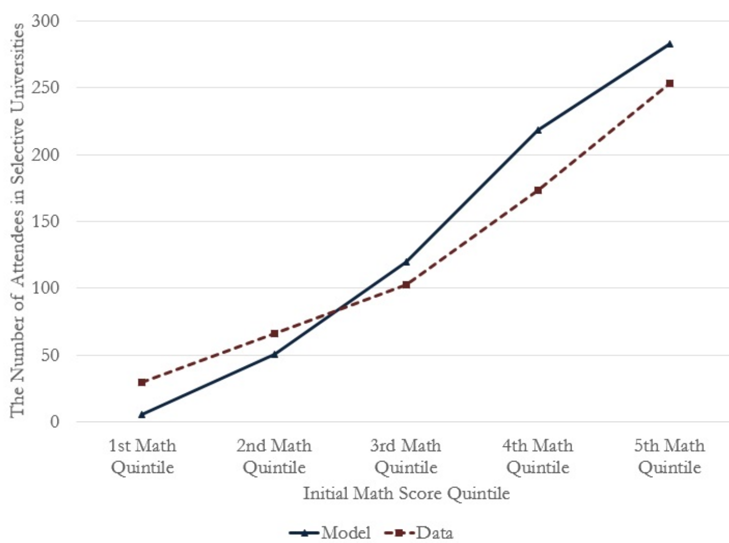
Note. The graph shows the predicted number of applicants by ability quintile from the estimated model the observed number of applicants by ability quintile from the data.

Figure 17: Model Fit: Number of Attendees in Selective Universities by Family Income Quintile



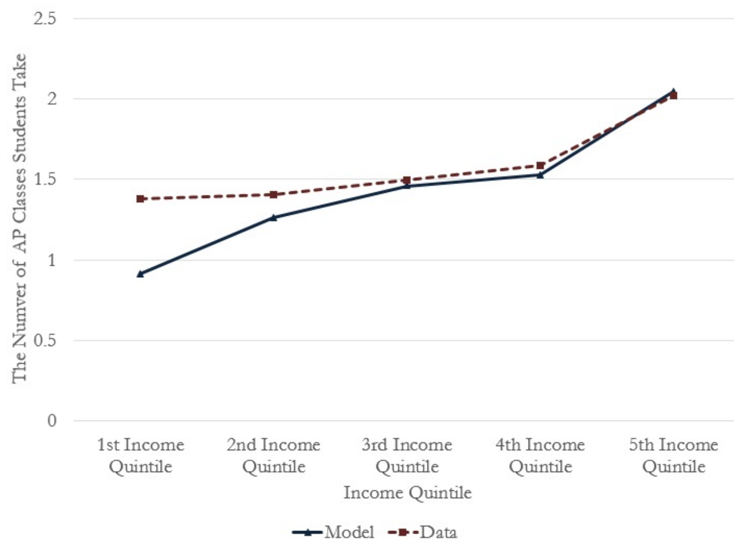
Note. The graph shows the predicted number of attendees by family income quintile from the estimated model the observed number of attendees by family income quintile from the data.

Figure 18: Model Fit: Number of Attendees in Selective Universities by Initial Ability Quintile



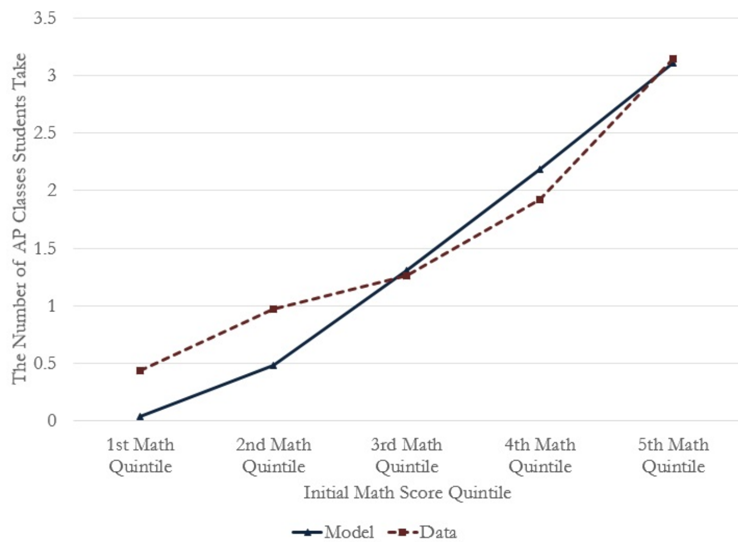
Note. The graph shows the predicted number of attendees by ability quintile from the estimated model the observed number of attendees by ability quintile from the data.

Figure 19: Model Fit: Number of AP Classes Students Take by Family Income Quintile



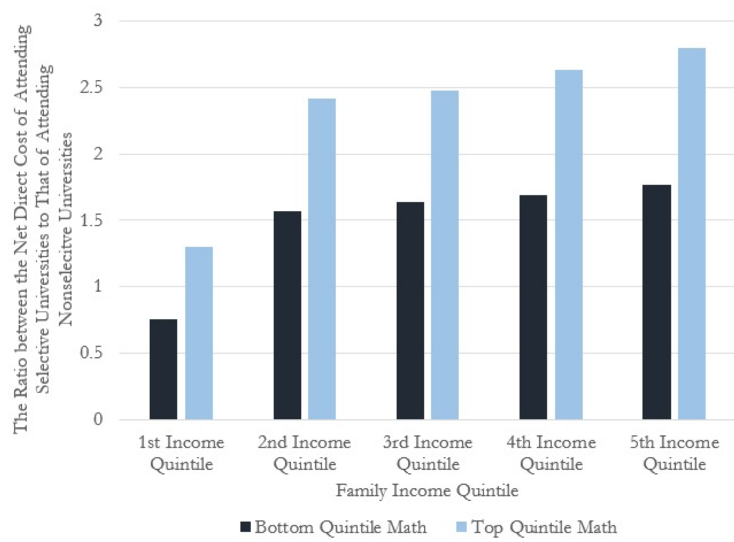
Note. The graph shows the predicted number of AP classes students take by family income quintile from the estimated model the observed number of AP classes students take by family income quintile from the data.

Figure 20: Model Fit: Number of Applicants for Selective Universities by Initial Ability Quintile



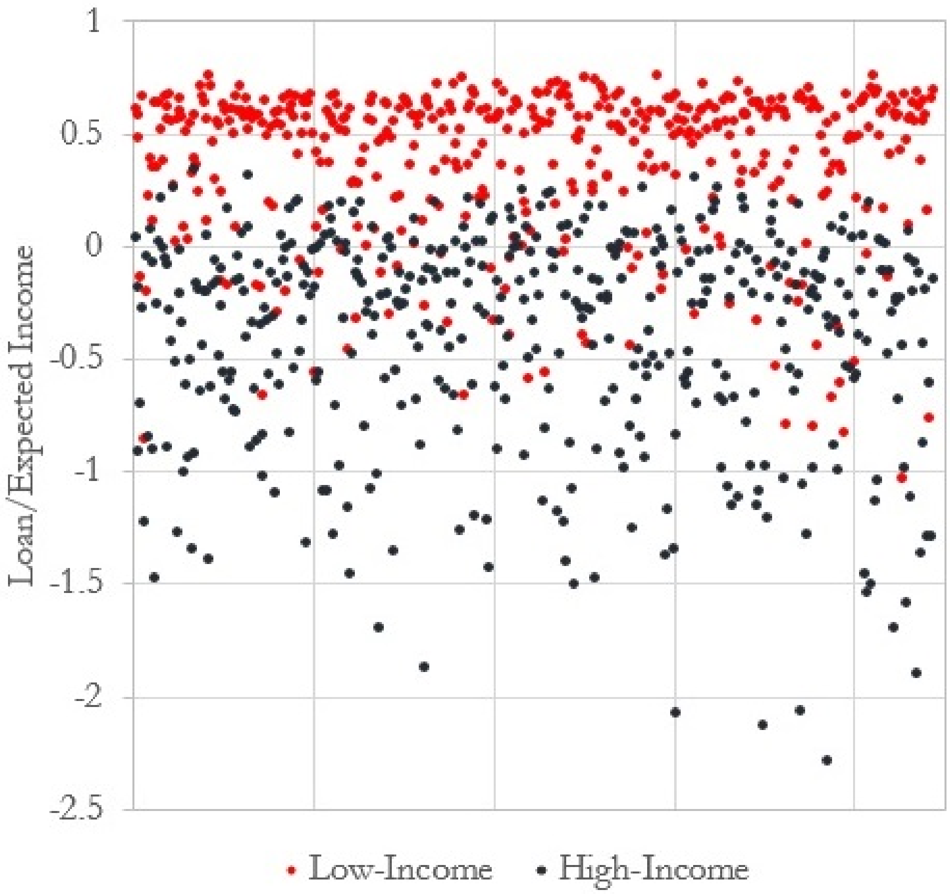
Note. The graph shows the predicted number of AP classes students take by initial ability quintile from the estimated model the observed number of AP classes students take by initial ability quintile from the data.

Figure 21: 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 22: Loan to Expected Income Ratio



Note. Estimated result about the loan to expected income ratio of low- and high-income students.

Parameter Estimates

Table 23: Estimation Result: 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 29.

Table 24: Estimation Result: 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 25: Estimation Result: 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 26: Estimation Result: 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 27: Estimation Result: 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 28: Estimation Result: 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 29: Estimation Result: 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.

Estimation Result

Table 30: Estimated Income of Graduates from Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	29,063	30,095	31,464	31,187	33,348
Ability2Q	31,644	32,235	33,560	33,512	34,550
Ability3Q	33,536	33,129	33,901	34,159	36,050
Ability4Q	33,357	34,333	35,463	35,201	36,859
Ability5Q	34,479	36,355	37,185	36,934	38,032

Note. Predicted labor income if students attend nonselective universities

Table 31: Estimated Income of Graduates from Selective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	31,354	32,269	33,946	33,647	35,978
Ability2Q	34,140	34,777	36,206	36,155	37,274
Ability3Q	36,180	35,741	36,574	36,853	38,893
Ability4Q	35,988	37,040	38,260	37,977	39,765
Ability5Q	37,198	39,222	40,117	39,847	41,031

Note. Predicted labor income if students attend selective universities

Table 32: Cost Difference between Graduates of Selective Universities and Graduates of Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	5,014	4,790	5,670	7,393	8,077
Ability2Q	7,118	6,691	6,818	9,535	9,247
Ability3Q	6,080	8,976	7,347	9,617	10,088
Ability4Q	8,577	8,620	8,783	9,211	11,557
Ability5Q	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 33: Number of years until making break-even point for attending selective universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	9.4	9.4	9.8	12.9	13.1
Ability2Q	12.2	11.3	11.0	15.4	14.5
Ability3Q	9.8	14.7	11.8	15.3	15.2
Ability4Q	14.0	13.6	13.4	14.2	17.0
Ability5Q	13.6	12.7	12.7	16.0	15.7

Note. The model assumes a repayment plan such that students repay the loan over 10 years after college graduation. I do not consider the wage growth over the life-cycle to find the break-even point.

Table 34: Amount of Student Loan When Attending Selective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	23,434	18,393	15,472	9,667	-8,670
Ability2Q	24,143	19,708	16,174	11,245	-6,045
Ability3Q	23,908	20,515	17,157	12,391	-5,967
Ability4Q	26,001	22,292	17,509	14,090	-5,207
Ability5Q	25,614	23,270	18,236	13,671	-5,923

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

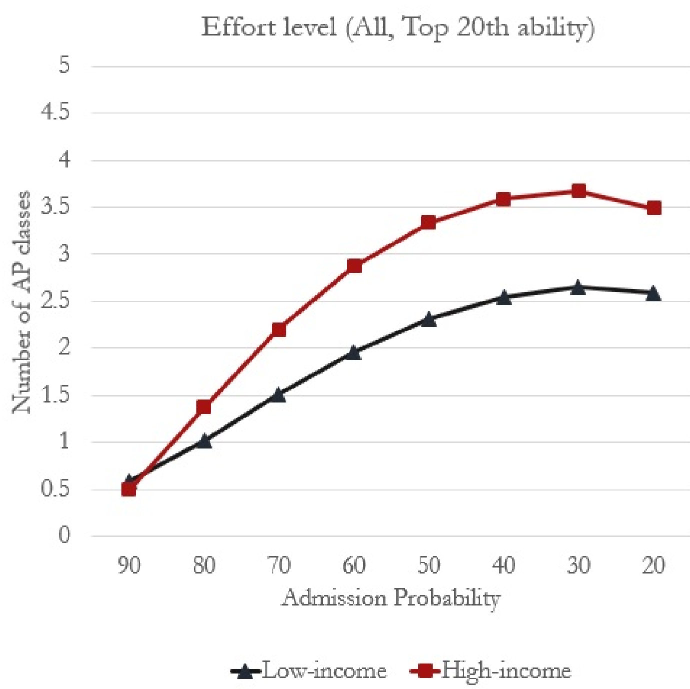
Table 35: Amount of Student Loan When Attending Nonselective Universities

	Income1Q	Income2Q	Income3Q	Income4Q	Income5Q
Ability1Q	18,493	15,161	12,118	6,767	-12,657
Ability2Q	17,974	16,734	12,314	7,339	-10,114
Ability3Q	19,481	17,092	13,747	8,706	-10,432
Ability4Q	20,551	17,924	13,853	9,904	-9,572
Ability5Q	22,136	19,268	14,344	9,386	-10,617

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

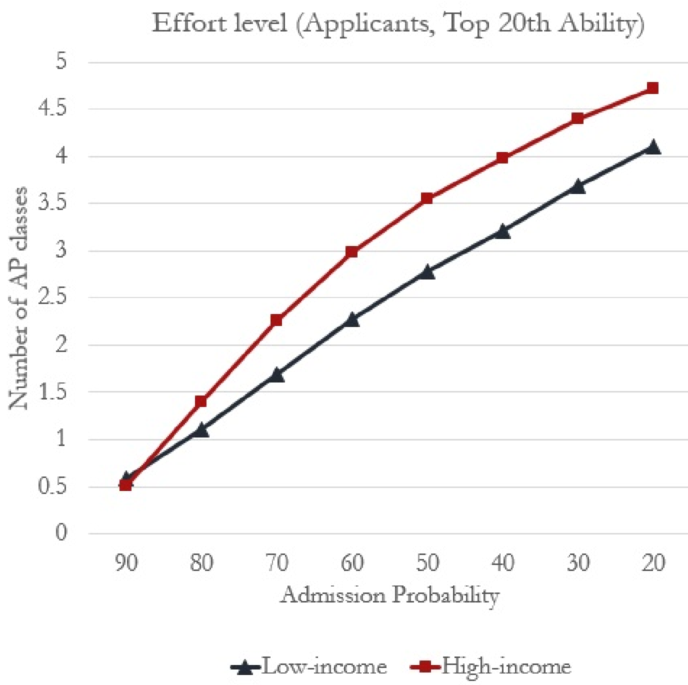
Appendix G: Counterfactual Analysis

Figure 23: Effort Choices as the Capacity in Selective Universities: Both Applicants and Nonapplicants



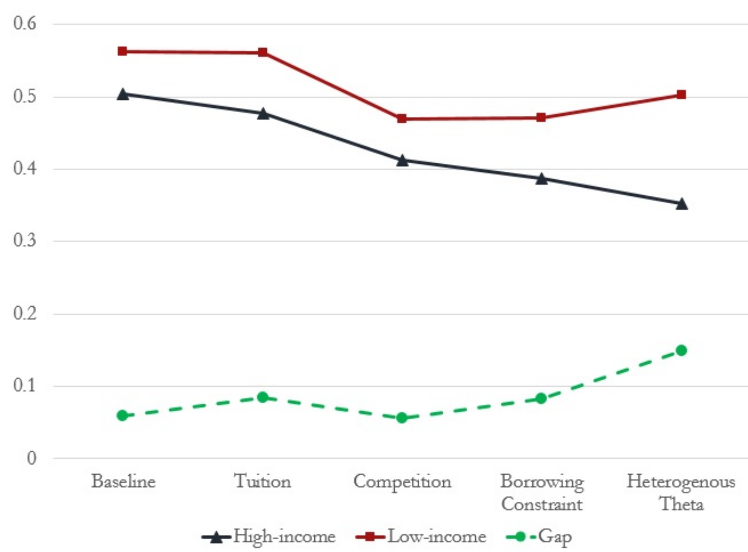
Note. I change the admission probability of applicants from 90% to 20% and see how student's effort choices change as competition is elevated. This is the average effort choice of low- and high-income students (the top and bottom quintile of income distribution) including both applicants and non-applicants.

Figure 24: Effort Choices as the Capacity in Selective Universities: Applicants



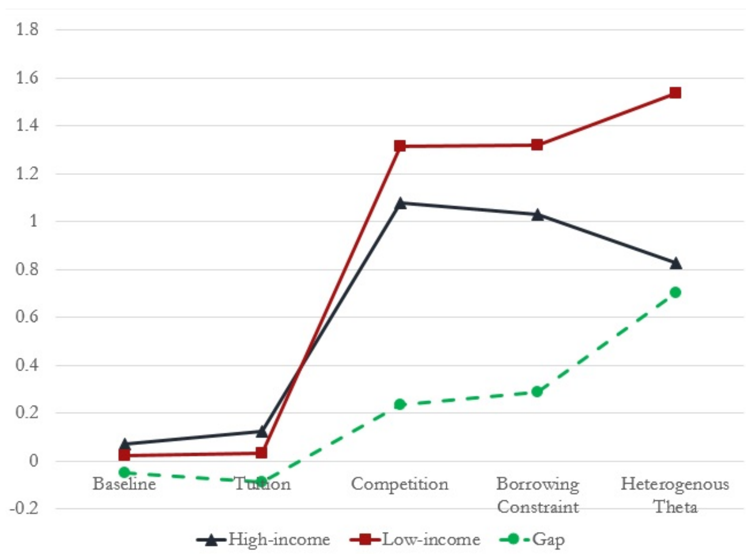
Note. I change the admission probability of applicants from 90% to 20% and see how student's effort choices change as competition is elevated. This is the average effort choice of low- and high-income applicants (the top and bottom quintile of income distribution).

Figure 25: What Explains the Application Gap



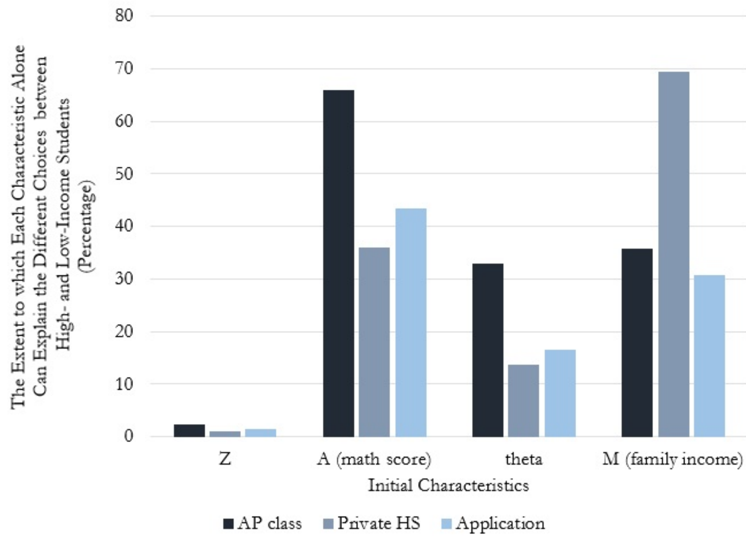
Note. To evaluate the importance of student's own characteristics and environmental factors in explaining the different application behaviors between rich and poor, I consider two representative students from the top and bottom quintile of income distribution. They have population average characteristics but family income. The baseline is the case when tuition is free, random admission into selective universities given the current capacity limit, and no heterogeneity in θ between rich and poor students. Then I compare how the application behaviors of rich and poor students changes as I introduce current tuition, test-based admission process, borrowing constraints, and heterogeneous θ.

Figure 26: What Explains the Effort Gap



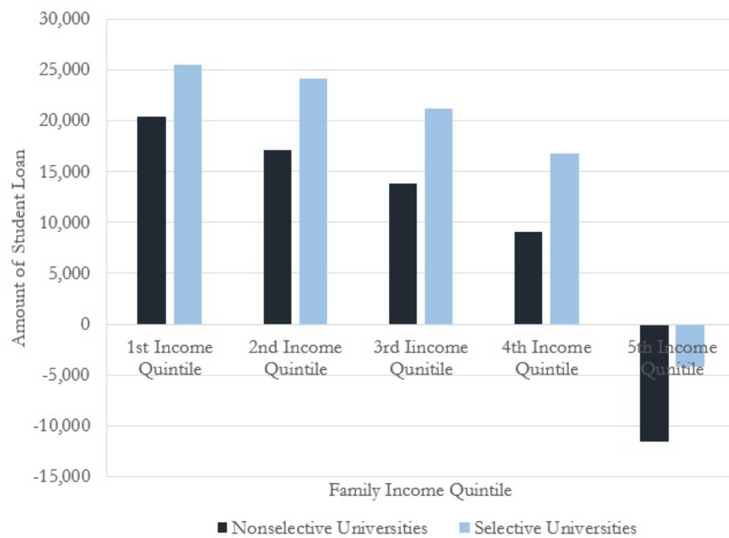
Note. To evaluate the importance of student's own characteristics and environmental factors in explaining the effort gap between rich and poor, I consider two representative students from the top and bottom quintile of income distribution. They have population average characteristics but family income. The baseline is the case when tuition is free, random admission into selective universities given the current capacity limit, and no heterogeneity in θ between rich and poor students. Then I compare how the application behaviors of rich and poor students changes as I introduce current tuition, test-based admission process, borrowing constraints, and heterogeneous θ.

Figure 27: Percentage Explained By Initial Characteristics Regarding Choice Difference Between Rich and Poor



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. By the rich and poor, I refer to those from the top and bottom quintile of income distribution. I obtain the average characteristics of students from the highest quintile and the bottom quintile of income distribution. In the counterfactual exercise, I compare the representative student from the high-income family and a counterfactual student who is different from the representative high-income student in one type of characteristic. 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 28: Amount of Student Loan by Family Income



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.

Table 36: Borrowing Constraints, Number Applicants, and the AP Classes among Students from the Bottom Quintile Income Distribution

Number of low-income applicants	Baseline	No Liquidity Constraint	Random Financial Aid
All Initial Scores	131	139	120
Top Quintile Initial Score	36	38	33
Effort level (AP class)			
All Initial Scores	0.9	1.0	0.8
Top Quintile Initial Score	2.6	2.7	2.4

Note. I round up the number for the report after simulating 100 times of the model. The total number of students from the bottom quintile of income distribution is about 450.

Table 37: Borrowing Constraints and the Gap between Rich and Poor Among Top Quintile Initial Score Students

	Baseline	No Liquidity Constraint	Random Financial Aid
Difference in AP classes	0.91	0.78 (86%)	1.10 (122%)
Difference in the number of applicants	135	132 (97%)	139 (103%)
Difference in SAT score	71	65 (92%)	80 (113%)
Difference in wage rates	1.83	1.75(95%)	1.95 (107%)

Note. I round up the number for the report after simulating 100 times of the model. The number is the difference between students within the top 20th percentile of initial test scores and from the top and the bottom quintile of income distribution. The total number of students from the bottom quintile of income distribution is about 450.

Table 38: Choice Difference between the Highest and the Lowest Income Quintile Students Explained by Different Initial Characteristics

Percentage Change	AP classes	Private HS	Application Rate
$\Delta(Z)$	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 39: Percentage Changes in the Number of AP Classes Conditional on Family Income and Initial Ability 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 40: 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 41: 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 42: Counterfactual Analysis: Number of Low-Income Applicants

Number of low-income applicants	Baseline	(1)	(2)	(3)
All Initial Ability	131	149 (114%)	126 (96%)	181 (38%)
Top Quintile Initial Ability	36	41 (113%)	36 (99%)	32 (89%)
Effort level (AP class)				
All Initial Ability	0.9	1.03 (113%)	0.91 (100%)	0.47 (52%)
Top Quintile Initial Ability	2.58	2.88 (102%)	2.62 (102%)	1.46 (57%)
Number of low-income applicants		(4)	(5)	(6)
All Initial Ability		122 (93%)	209 (159%)	134 (102%)
Top Quintile Initial Ability		34 (94%)	45 (127%)	37 (102%)
Effort level (AP class)				
All Initial Ability		0.81 (89%)	1.03 (114%)	0.99 (109%)
Top Quintile Initial Ability		2.38 (92%)	2.57 (99%)	2.75 (106%)

Note. I round up the number for the report after simulating 100 times of the model. (1): doubling need-based aid from selective universities ($12,000 more grants) for the students from the bottom quintile of income distribution, (2): merit-based aid with the same budget spending as (1), (3): the Texas Top 10 Law (only GPA is taken into account in the admission process), (4): increasing need-based aid from nonselective universities by $12,000 for the low-income students, (5): income quota system that gives bonus points in the admission process to double the number of low-income attendees in selective universities, (6): offering more AP classes to low-income students with the same extra budget needed to implement increasing need-based aid from selective universities by $12,000.

Table 43: Counterfactual Analysis: Gaps in the Outcome of the Low-income and High-income High-Ability Students

	Baseline	(1)	(2)	(3)
Difference in AP classes	0.91	0.60 (67%)	0.90 (100%)	0.43 (47%)
Difference in the number of applicants	135	130 (96%)	136 (101%)	118 (87%)
Difference in SAT score	71	57 (80%)	72 (101%)	43 (61%)
Difference in wage rates	1.83	1.65 (90%)	1.84 (100%)	1.65 (90%)
		(4)	(5)	(6)
Difference in AP classes		1.12 (124%)	0.84 (93%)	0.74 (82%)
Difference in the number of applicants		138 (102%)	119 (88%)	134 (99%)
Difference in SAT score		82 (115%)	57 (80%)	64 (89%)
Difference in wage rates		1.91 (107%)	1.41 (77%)	1.77 (96%)

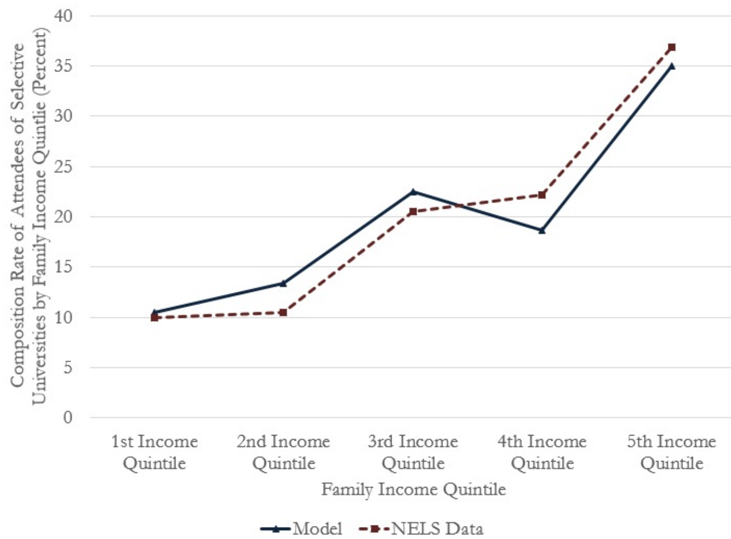
Note. I round up the number for the report after simulating 100 times of the model. (1): doubling need-based aid from selective universities ($12,000 more grants) for the students from the bottom quintile of income distribution, (2): merit-based aid with the same budget spending as (1), (3): the Texas Top 10 Law (only GPA is taken into account in the admission process), (4): increasing need-based aid from nonselective universities by $12,000 for the low-income students, (5): income quota system that gives bonus points in the admission process to double the number of low-income attendees in selective universities, (6): offering more AP classes to low-income students with the same extra budget needed to implement increasing need-based aid from selective universities by $12,000.

Out-of-Sample Prediction

Figure 29: 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 30: 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.

Additional Tables

Table 44: Moment Conditions for Outcome Variables

	Observed	Simulated	
K_i	Z_i	$\frac{1}{n}\sum_{i=1}^{n} K_i Z_i$	$\frac{1}{n}\sum_{i=1}^{n} \tilde{K}_i Z_i$
$ln(Wage)$	1	2.8229	2.8109
$ln(SAT)$	1	6.9931	6.9774
$ln(GPA+1)$	1	1.1771	1.1591
$ln(APscore+1)$	1	0.4562	04618
I_{sel}	1	0.2027	0.2089
I_{BA}	1	0.7363	0.7596

Note. There are four outcome variables in the model: wage rates, test scores (SAT and AP score) and GPA, admission results, college graduation rates. I also include the interaction terms of each outcome variable with a set of instrumental variables in the moment condition.

Table 45: Moment Conditions for Choice Variables

	Observed	Simulated	
K_i	Z_i	$\frac{1}{n}\sum_{i=1}^{n} K_i Z_i$	$\frac{1}{n}\sum_{i=1}^{n} \tilde{K}_i Z_i$
N_{AP}	1	2.8229	2.8109
$I_{private}$	1	6.9931	6.9774
I_{apply}	1	1.1771	1.1591
$L \times I_{BA} \times I_{sel}$	1	0.1037	0.1070
$L \times I_{BA} \times (1 - I_{sel})$	1	0.4750	0.4380

Note. There are six choice variables in the model: the number of AP classes students take ($N_{APclass}$), whether to attend private high school ($I_{private}$), whether to apply for selective universities (I_{sel}), consumption during college period (C_1), consumption when working (C_2), and the amount of student loan (L). Because I do not directly observe consumption levels, I do not include consumptions in the moment condition. The model predicts the optimal consumption level based on family income, labor earning, and student loan. I also include the interaction terms of ($N_{AP}, I_{private}, I_{application}, L$) with a set of instrumental variables such as family income, initial math score, race, sex, tuition level, reported preference for college characteristics.