

Agent-based Simulation Models of the College Sorting Process*

Sean F. Reardon[^]
Matt Kasman[^]
Daniel Klasik⁺
Rachel Baker[^]

[^] Stanford University

⁺ University of Maryland, College Park

* Corresponding Author: Sean F. Reardon, Stanford University, Center for Education Policy Analysis, 520 Galvez Mall, #526, Stanford, CA 94305, sean.reardon@stanford.edu.

Acknowledgements: This research has benefited from the thoughtful feedback of members of the Network on Inequality, Complexity, and Health (NICH), particularly Elizabeth Bruch and Rucker Johnson. This work was partially funded through a grant from the Office of Behavioral and Social Sciences Research of the National Institutes for Health (Contract OBSSR HSN2762008000136). In addition, Rachel Baker and Matt Kasman received support from the Stanford Predoctoral Training Program in Quantitative Education Policy Research, funded by the Institute of Education Sciences (IES Award R305B090016). The opinions expressed here do not represent those of the funding agencies. All errors are our responsibility.

Agent-based Simulation Models of the College Sorting Process

Abstract

We explore how dynamic processes related to socioeconomic inequality operate to sort students into, and create stratification among, colleges. We use an agent-based model to simulate a stylized version of this sorting processes in order to explore how factors related to family resources might influence college application choices and college enrollment. We include two types of “agents”—students and colleges—to simulate a two-way matching process that iterates through three stages: application, admission, and enrollment. Within this model, we examine how five mechanisms linking students’ socioeconomic background to college sorting might influence socioeconomic stratification between colleges including relationships between student resources and: achievement; the quality of information used in the college selection process; the number of applications students submit; how students value college quality; and the students’ ability to enhance their apparent caliber. We find that the resources-achievement relationship explains much of the student sorting by resources but that other factors also have non-trivial influences.

Keywords: agent-based models, socioeconomic inequality, college sorting, college admission, college enrollment

Agent-based Simulations of the College Sorting Process

Students with higher socioeconomic status are much more likely to enroll in college than their lower status peers (see, for example, Cabrera & La Nasa 2000; Bailey & Dynarski 2011). In addition, family resources are also strongly associated with the selectivity and quality of the particular college in which a student enrolls: students from families in the top income decile are 8 times more likely to enroll in top-tier colleges than students in the lowest decile, a gap that has been growing over time (Alon 2009; Bastedo & Jaquette 2011; Reardon, Baker, & Klasik 2012). Taken together, these two phenomena generate substantial socioeconomic stratification in the American postsecondary educational system.

We do not know specifically why socioeconomic status is so instrumental in determining whether and where students attend college. This is primarily because college attendance is determined by a complex, two-sided matching process. Students have an enormous array of colleges from which to choose when they submit their college applications. From the applications they receive, colleges then have discretion regarding whom to admit. Finally, students choose where to enroll from among the colleges that have admitted them. Socioeconomic background need not explicitly enter into any stage of this process, though there are a number of mechanisms that might affect the degree to which a student's socioeconomic background is associated with enrollment at highly-selective colleges.

Perhaps most significantly, academic achievement is strongly associated with family income and socioeconomic status: high-income students have much higher scores on standardized tests (including the SAT and ACT) than middle- and low-income students, and this

gap has been growing over time (Reardon 2011). Because academic achievement is a key criterion for admission to selective schools, it is not surprising that high-income students are more likely to be admitted to such schools.

The college admissions process, however, is complex and the relationship between income and achievement may be only one part of the explanation for the apparent income advantage in college enrollment. Many other mechanisms may play important roles. High-resource students may engage in activities that make them more attractive to colleges, such as using admissions consultants or spending more time pursuing extracurricular interests. These students may also have more knowledge of the postsecondary market on which to base their college decisions, may tend to submit a greater number of applications, or may evaluate the benefits of college attendance differently. Colleges themselves may also play a role by using recruitment or admissions strategies based on non-academic factors that might be related to socioeconomic status (e.g. giving preference to legacy admissions, or, conversely, giving admissions priority to qualified low-income students). Of the many ways through which socioeconomic status might affect college attendance, the relative importance and specific role each plays is unclear.

Our goal in this paper is to build intuition about the relative strength of some of these mechanisms in shaping the distribution of students, by socioeconomic background, among more- and less-selective colleges and universities. We focus on a subset of the possible mechanisms that may drive this sorting. In particular, we concentrate on those associated with student characteristics and behaviors: academic achievement differences, application behavior differences, and application enhancement differences among students of different

socioeconomic backgrounds.

To explore these issues, we use a two-sided agent-based model in which student agents make decisions about what colleges to apply to; colleges make decisions about which applicants to accept; and students make decisions about which admission offer to accept. By altering the distribution of student characteristics and the factors that govern their application behaviors, we use the model to explore the relative effects of various mechanisms on college enrollment patterns. These simulations are not intended to fully explain existing patterns of college enrollment, but rather to provide some insight into the ways in which socioeconomic status influences who attends college and where they enroll.

Background

College Sorting and Resources

In this article, we are primarily concerned with what we refer to as *college sorting*. We conceive of college sorting as the two-sided process in which students and colleges interact through the application, admission, and enrollment processes to determine the particular colleges in which students enroll. The attributes, constraints, and preferences of both students and colleges jointly determine the final distribution of students among schools.

While the field of higher education has historically focused on the question of whether students attend college at all, some scholars argue that increasing rates of college degree aspiration among high school students have made this question less important than studying how students choose *where* to attend college—how they sort into colleges (Hoxby 2004). This shift is the result of a growing body of research that shows that the benefits a student receives

from college can differ according to the quality or selectivity of the institution from which the student graduates. Students who attend elite, highly-selective schools enjoy larger tuition subsidies, disproportionately extensive resources, and more focused faculty attention (Hoxby 2009). Further, elite graduate schools and top financial, consulting, and law firms recruit almost exclusively at highly selective colleges; likewise, evaluators at elite firms routinely use school prestige as a key factor when screening resumes (Rivera 2009; Wecker 2012). These differences appear to be reflected in higher lifetime earnings for those students graduating from more selective colleges (Black & Smith 2004; Hoekstra 2009; Long 2008), though these higher earnings may be limited to minority students and those whose parents had low levels of education (Dale & Krueger 2011).

Given the advantages conveyed by graduation from a selective college, these institutions have the ability to play a large role in facilitating social mobility in the United States. But students, particularly ones from lower-income families, do not always appear to make application and enrollment decisions that would maximize these benefits (Bowen, Chingos, and McPherson 2009; Hoxby and Avery 2012; Roderick, Nagaoka, Coca, and Moeller 2008; Roderick, Nagaoka, Coca, & Moeller 2009). Hoxby and Avery (2012) show that as many as 92 percent of high-achieving low-income students often apply to much less selective colleges than similarly high-achieving high-income students, even when the availability of financial aid would mean they could afford to attend such schools.

Perhaps as a result of these application choices, students from low-income families are much less likely than high-income students to enroll in selective colleges. Figure 1 demonstrates this disparity by showing that students whose family income fell in the 80th percentile nationally

were four times more likely to enroll in one of these schools than a student in the 20th percentile. This disparity is even more extreme for higher/lower percentile income families. Reardon, Baker, and Klasik (2012) show that students from families earning more than \$75,000 (in 2001 dollars) were dramatically overrepresented in the most selective categories of colleges, while students from families earning less than \$25,000 were notably underrepresented at these same schools. Such disparities are not new, but the underrepresentation of low-income students at highly selective schools has increased over time (Alon 2009; Astin & Oseguera 2004; Belley & Lochner 2007; Karen 2002). This trend has paralleled an increase in income stratification within the US, as well as an increase in the achievement gap between high- and low-income students (Reardon 2011).

Resources and the College Application Process

In this paper we explore five possible explanations for the overrepresentation of higher-resource students at higher-quality colleges.

Differential high school academic achievement. There is a strong correlation between family income and academic achievement. Whether it is because of the greater resources wealthy families are able to put in to educating their children, including through residential choices, or because parents in high-income families generally have higher education levels themselves, children from high-income families tend to outscore their low-income peers across a wide battery of achievement measures (Reardon 2011). Given the weight college admissions offices place on such achievement measures, this correlation may go a long way to explaining the income advantage at selective colleges.

Application enhancement. Regardless of academic ability, higher-income students may engage in activities that enhance their likelihood of admission to more selective colleges. For example, participation in extracurricular activities (overseas trips, athletics, music or arts activities, volunteer activities), enrollment in SAT/ACT prep classes, or retaking of the SAT/ACT may all work to improve students' desirability to colleges. The time and money often required to participate in these activities may be prohibitive for low-resource students in a way they are not for higher-resource students.

Unequal information. The college destinations of low-and high-resource students may be different from each other because they apply to different sets of schools. Part of this apparent difference may be the result of differential access to information. There are three types of information that are important for students as they decide where to apply to and enroll in college: information about the potential costs and benefits of different colleges, information about their own desirability to colleges relative to other students, and information about the likelihood of admission to different colleges, given their desirability. Relative to higher-income students, lower-income students have less information about these three factors on which to base their application decisions (Avery & Kane 2004; Hoxby & Turner 2013; McDonough 1997). Further, it may be that low-resource students not only lack information about colleges, but the information they have is flawed or incorrect. For example, low-income students are generally poor at estimating both the cost and benefits of college attendance (Avery & Kane 2004; Grodsky & Jones 2007). As a result, low-resource students may not think some colleges would be as accessible or beneficial for them as similarly skilled high-resource students.

Perceived utility of college enrollment. Even with good information, and given equal

chances of admissions, high- and low-resource students may not hold equal perceptions of the value of applying to and attending a highly selective institution. Students may have preferences that lead to different utility valuations over a host of college characteristics. One particular subset of these preferences may involve differential sensitivity to college cost. This might occur, for example, if low-income students disproportionately perceive the economic and/or social costs of attending such an institution to be higher than the potential benefits. Hoxby and Turner (2013) find little evidence that lower income students value selectivity less than higher income students, however, at least among high-achieving students. This is based on their observation that low-income students who have been provided with detailed cost information make similar application decisions as high-income students. It is still an open question whether differential preferences affect college sorting at other points of the achievement distribution.

Number of applications. The number of college applications students submit is associated with the likelihood of four-year college enrollment in general (Smith 2012). Applying to more schools likely also increases the odds of admission to selective schools, at least for students on the margin of being admitted to any such school. If the time and cost associated with submitting multiple college applications prevents low-resource students from submitting as many applications as high-resource students, then this mechanism may also explain differential sorting into selective colleges by socioeconomic status.

Cost. While tuition and financial aid are undeniably important parts of students' college choices, we do not include these elements of cost explicitly in our model. This choice is, in part, because we wish to focus on the other, less well-studied, processes described above. There are, however, two ways in which college cost considerations enter our model. First, to the extent that

sticker price and college quality are generally correlated, low-resource students may see less utility in attending a higher quality school, which is captured in our resource-dependent utility of college enrollment. Second, because sticker price isn't the whole story with respect to college cost, students with more information about financial aid options may still prefer higher quality colleges, despite the higher sticker price (see Hoxby & Turner 2013). Thus, differential utility and differential information both account for some of the effects of college cost on students' college choices in our model.

Agent-based Models

Agent-based models are ideal for studying college sorting both because the actions of students are interdependent and the institutional environment is dynamic. Student actions are interdependent because, given the finite number of seats available in any one college, one student's application decision affects the likelihood of admission of all other students. These decisions also affect the application decisions of students in the future because the quality of colleges changes according to the students who enroll, and because colleges' selectivity changes over time in response to changes in their applicant pools. Because of the dynamic nature of the process, students must adjust their decisions in response to both the change in valuation of a college and the change in likelihood of admission. Allowing the agent-based model to run over a number of years allows for the observation of emergent patterns in this interdependent and dynamic process.

Very few studies have used agent-based models as a means to study issues in education, and even fewer have used this method to study college sorting. Maroulis et al. (2010) use real-

world data on schools and students in Chicago to explore the potential effects of introducing intra-district choice to the school system. Howell (2010) conducts a structural estimation based on nationally representative data to determine what would happen to college diversity if colleges were prevented from using affirmative action in admissions decisions. Hendrickson (2002) designed an agent-based model to demonstrate that such a model could indeed be used to approximate the college enrollment decisions made by real students. She accomplished this task by having students use very simple strategies to apply to colleges (e.g. apply to all schools or apply to schools randomly) and compared her results to real world observed college choices.

We extend Hendrickson's work in two main ways. First, we use a more sophisticated (but still highly-stylized) simulation of students' application decisions. Second, we run a series of "policy experiments" to investigate the impact of each of our proposed mechanisms linking socioeconomic status and college destinations. Agent-based simulation methods are ideal for conducting such experiments because experimentally testing these mechanisms in the real world is nearly impossible.

Data and Method

In this section, we describe our agent-based model, the empirical basis for its input parameters, and the analyses that we perform using its output. In order to ensure that we provide enough information for readers to understand and potentially replicate our simulations, our description includes all of the applicable elements suggested in the ODD (Overview, Design concepts, Details) protocol for describing agent-based simulations (Grimm et al. 2006; Grimm et al. 2010).

Motivation for Model

The goal of our model is to develop intuition about how student characteristics and behavior influence the sorting of students into colleges of varying quality. Figures 2 and 3 present an overview of the agents and processes in our simulation.

Agents

Our model includes two types of entities: students and colleges (see Figure 2). Students have two attributes that we call “resources” and “caliber.” The “resource” attribute is intended to represent a unidimensional composite of the various forms of socioeconomic capital available to a student and that may affect the college application process (e.g. income, parental education, and knowledge of the college application process). The “caliber” attribute is intended to represent a unidimensional composite of observable markers of academic achievement, potential for future academic success, and other characteristics valued by colleges (e.g. grades, standardized test scores, application essay quality, extracurricular activities, unique talents or skills, etc.). We refer to this as “caliber,” rather than “academic preparation” simply to indicate that colleges may value non-academic student characteristics as well. For ease of interpretation, however, we represent caliber on an SAT-like scale.

Colleges have a single attribute, “quality,” which is intended to capture the average desirability of a college to prospective students. We operationalize quality as the average of the caliber of students enrolled in the school, with recent years’ classes weighted more than earlier years. Although in the real world the average caliber of enrolled students may not correspond strictly to the quality of a college’s educational experience, in practice, average student caliber is

widely used as a rough proxy for quality. Prospective applicants have more information about the characteristics of enrolled students (average SAT scores, for example) than they do about the quality of instruction, for example.

Students and colleges in our model each have straightforward objectives: students wish to maximize the utility received from the college in which they enroll and colleges wish to maximize the average caliber of their enrolled students.

Both colleges and students have imperfect information and idiosyncratic preferences regarding one another. As a result, any two students may not rank colleges identically and any two colleges may not rank students identically. Operationally, this is implemented in the model by adding random noise to each student's perception of each college's quality, and by adding noise to each college's perception of each applicant's caliber. Moreover, students do not have perfect information about their own caliber. Again, this is operationalized in the model by adding random noise to each student's perception of her own caliber.

A key feature of the model is that the amount of noise added to students' perceptions of their own caliber and of college quality is allowed to be a (decreasing) function of their resources. In this way, higher-resource students have more accurate information about their own caliber and about colleges' quality, which enables them (as we will see) to better target their applications.

Model Operation

Our model moves through three stages: application, admission, and enrollment (see Figure 3). The completion of these three stages represents one year. During the application

stage, students observe (with some uncertainty) the quality of each of the pool of colleges in a given year and select a portfolio of colleges to which they apply. In the admission stage, colleges rank applicants by their caliber (again with some uncertainty), and admit the highest-ranked applicants, up to a total number of students that colleges believe will be sufficient to fill their available seats. In the enrollment stage, students compare the schools to which they have been admitted and enroll in the one with the one they perceive to be of highest quality. At the end of each simulated year, the selectivity, yield, and quality of each college are updated based on the admission and enrollment outcomes. The colleges, with their updated characteristics, are then considered by a new cohort of students in the next year of the model, when the three stages of the process are repeated.

Both students and colleges are able to observe and adapt to one another's previous actions. Students observe the admissions outcomes of prior cohorts of students, from which they infer how the probability of admission is related to the difference between a student's caliber and the quality of a given college. From this, they estimate their likelihood of admission to every college given their perceptions of their own caliber and of each college's quality. This predicted likelihood is used in conjunction with the perceived utility of attending particular colleges to determine students' application sets.

Colleges determine the number of students to admit by observing their own prior yield rates—the percent of their accepted students who ultimately enrolled in their college. Colleges will admit more students if they did not fill their seats in prior years and admit fewer students if they enrolled more students than they had seats.

A more detailed description of the agents and processes in our model can be found in Appendix A.

At the end of each model run, we have highly detailed information of student and college behavior in each year, but focus on the emergent behavior in the final year of the model. We use this behavior to construct three specific measures of student sorting into colleges. First, we examine the relationship between resources and the rate at which students enroll in any college. Second, we examine the relationship between resources and the rate at which students enroll in one of the top ten percent of colleges (as ranked by quality) in our model. Finally, we examine the relationship between student resources and the quality of the colleges students attend. Taken together, these outcomes allow us to answer three important questions about our simulated world: (1) Who is attending college? (2) Who is attending elite colleges? And (3) how closely aligned is college quality to student resources? We focus on the five pathways through which students' resources and caliber might affect the sorting of students into college, described above, to evaluate the extent to which the mechanisms described above, either individually or in combination, affect the sorting of students into colleges.

Model Parameters

We select parameters that determine student and college attributes, perceptions, and behaviors that approximate what we find empirically using real-world data or, where that is not possible, what we judge to be plausible values. Table 1 outlines the parameter values we use and their sources.¹

Experiments

This simulated world, with flexible parameters and multiple pathways through which student resources can affect college quality, gives us the ability to understand how students might be sorted by resources across colleges and gives us intuition about which kinds of interventions would be the most effective in reducing this stratification. In order to do this, we run our model under a set of experimental conditions. The parameter values associated with resource pathways in each experiment are outlined in Table 2.

We examined how changes in resource pathways affected three main outcomes: likelihood of enrolling in college, likelihood of attending a top-10% college, and the relationship between student resource and college quality. In order to minimize the influence of random error on our results, we ran our model 100 times using each set of parameters that we discuss below.²

We also performed a Latin Hypercube analysis in order to explicitly and simultaneously quantify the contribution of each of the five mechanisms of interest on our three different outcomes (Bruch & Atwood 2012; Segovia-Juarez et al. 2004). We divided a range of possible values for each of the five parameters into 10 even parts. We then sampled one of these values from each of the five parameters, without replacement, and ran the agent-based model using the resulting 10 combinations of parameter values. This sampling method ensures that, in expectation, the 5 parameters are not correlated with each other. Using the results of the 10 runs of the model, we ran regressions predicting measures of disparities in enrollment outcomes between high- and low-resource students. Specifically, in the final year of each model run, we compute the gap in likelihood of each of our three outcomes of interest (college enrollment,

enrollment in a top 10% college, and college quality) between (1) the 10th and 90th percentile of family resources, (2) the 50th and 90th percentile of family resources, and (3) the 10th and 50th percentile of family resources on all five parameters. We select these three specific outcomes based on examination of the outcome functions obtained under experimental conditions, discussed below.

Results

In the sections that follow, we present our results in two ways. First, we present the graphical results of eight different model scenarios: a model where student resources are not allowed to influence the college sorting process, a model where the parameters have been set to simulate real world conditions as observed in nationally representative data sets (as outlined above and in Table 1), and our six main policy experiments. These figures present our three main outcomes across the full distribution of student resources and allow us to note general patterns in how particular resource pathways affect college sorting. Second, we present the results of our Latin Hypercube analysis, which work to quantify the results of the graphical analysis for different sections of the resource distribution.

Model 1: Basic – No Resource Influence

As expected, the model that does not include any of the resource pathways produces an equal distribution of students from varying resources across colleges. Higher resource students are no more likely than lower resource students to enroll in any college or a top 10% college.

Further, there is no relationship between student resources and college quality. At every point of the resource distribution, the probability of each of these outcomes is equal.

Model 2: Real World Baseline – All Resource Pathways

In our next model—our baseline model—we allow resources to affect college quality via all five pathways. As we describe above and in Table 1, when possible we chose values for each of our pathways based on empirical data. Using these plausibly realistic values for parameters, we find patterns that are similar to what we see empirically, which serves to demonstrate the capacity of our model to mimic real world behavior. For example, in terms of patterns of applications and admissions, the relationships between college quality and the number of applications received, number of students admitted, and number of students enrolled (Figure 5) is similar to the real world, using data from the Integrated Postsecondary Education Data System (IPEDS, collected by the National Center for Education Statistics). The relationships between college quality and selectivity (admission rate) and yield (enrollment rate) (shown in Figure 6) are also quite similar to IPEDS data (graphs showing the same relationships using IPEDS data are in Appendix B).

These plausible parameter values dramatically change student enrollment outcomes from a world in which there is no resource influence. In this model, as compared with our basic model, students from high resource backgrounds are much more likely both to enroll in any college and to attend a top-10% college. Students from low resource backgrounds are correspondingly less likely. While students in the basic model all have about a 75 percent likelihood of enrollment in any college, turning on these five pathways increases the likelihood of

college enrollment for the students in the 90th percentile of family resources to over 90 percent while the likelihood for students from students whose families are in the 10th percentile of resources decreases to nearly 55 percent. This change in likelihood is even more dramatic for enrollment in one of the schools in the top 10 percent of our distribution. Whereas in the basic model all students have a roughly equal probability of enrolling in a highly selective school, with the five resource pathways turned on, the likelihood of enrollment for 90th percentile students is nearly 20 times what it is for 10th percentile students. There is also a strong relationship between student resources and college quality. Figure 7 shows each of these relationships. Again these simulations mimic the patterns evident in empirical data. For example, the relationship illustrated in Figure 7 is remarkably similar to the depiction of the same relationship using real-world shown in Figure 1.

The similarity between the application, admission, and enrollment patterns that result from this model and those observed in real-world data bolster our confidence that we have a reasonable starting point from which we begin testing alternative conditions.

Models 3-8: Policy Experiments

Figures 7-9 show the results of experiments 3-8. In general, the correlation between student resources and caliber has the strongest influence on the relationship between students' resources and their college destinations, while other resource pathways have more subtle, but still notable, effects.³

Eliminating the correlation between resources and caliber decreases the difference in probability of enrollment for very high and very low resource students from about 50 percent to

closer to 20 percent (Figure 7). Figure 8 shows that eliminating this correlation also has a large effect on differences in the probability of enrolling in a highly selective school. Without the correlation between student resources and student caliber, the students in the 90th percentile of resources are about four times as likely as those in the 10th percentile to enroll in highly selective school, compared with about 20 times as likely when all resource pathways are turned on. The effect on quality of enrollment is also large—without the resource-caliber correlation, students in the 90th percentile enroll in schools with an average quality 75 points higher than students in the 10th percentile, which is roughly half the difference that results when all resource pathways are engaged. While the correlation between resources and caliber is clearly the most powerful player, other pathways have non-negligible effects.

In the model where the application enhancement pathway is not active, there is a significant shift toward equality. If students are unable to enhance their perceived caliber, the relationship between student resources and probability of enrollment at any college decreases. The probability of a very high resource student enrolling decreases by about three percentage points (roughly from 93 percent to 90 percent) and the probability of a very low resource student increases by a similar margin (roughly from 55 percent to 59 percent). Probabilities for students toward the middle of the resource distribution do not change appreciably. The relationship between student resources and probability of enrolling in a top-10% school is also affected when we do not allow high resource students to enhance their caliber. Students in the bottom 60% of the resource distribution are about one percentage point more likely to attend a selective college, while students in the top 20% of the distribution are much less likely (up to six percentage points less likely).

In the model where resources do not affect the quality of information students have about their own caliber and college quality, the relationship between student resources and the probability of enrolling in any college remains remarkably unchanged. However, removing this pathway does affect a student's probability of enrolling in a top-10% college. Students from the middle of the resource distribution (between about 20 and 70 percent) have an increased probability of attending a highly selective school (up to two percentage points), while students at the very high end of the resource distribution have a decreased probability (about five percentage points less likely).

Eliminating the relationship between resources and the number of applications a student submits has a small but observable effect at the lower end of the resource distribution, increasing both the probability of college enrollment and the quality of college students in the bottom quartile attend. Intriguingly, the relationship between resources and the perceived utility of college quality does not appear to appreciably affect the outcomes of interest.

The last model in Figures 7-9 shows attendance behavior when only the relationship between resources and caliber is engaged (all other resource pathways are removed). Particularly striking in these figures is the fact that they look quite similar to the model in which all pathways *except* for this relationship are engaged. Thus, it appears that the other four pathways combined have an effect on college attendance similar to the effect of the resource-caliber pathway alone.

Latin Hypercube Analysis

In addition to visualizing our outcomes of interest under specific experimental conditions, we also conduct a more formal exploration of our parameters' influence using Latin Hypercube analysis. Although we lose some of the nuance of observing the functions depicted in Figures 7-9 (i.e. observing exactly where on the resource distribution particular mechanisms seem to have the most influence), we gain the ability to quantify and compare mechanism effects.

We use slightly different outcomes in the Latin Hypercube analyses. Here, we regress gaps in enrollment outcomes (i.e. differences between those at the 90th and 10th percentiles of the resource distribution, at the 90th and 50th percentiles, and 50th and 10th percentiles) on our five mechanisms of interest. Gaps are a convenient way to quantify inequality. In our model without resource pathways, the gaps are 0 for all three outcomes that we consider (flat relationship between resources and outcomes). As we allow student resources to affect the application and admission decisions, the relationships between resources and outcomes get steeper and significant gaps emerge. We chose these three gaps (90-10, 90-50, and 50-10) to analyze. The 90-10 gap tells us what the difference in substantive outcomes are between those at the very top and the very bottom of the resource distribution, while the 90-50 and 50-10 gaps let us say something about whether the gaps are being driven by the experiences of those at the top of the resource distribution (where we expect to see disparities in access to elite schools), the bottom of the resource distribution (where we expect to find disparities in access to any college), or both. Respectively, Tables 3 through 5 explore gaps in likelihood of college enrollment, likelihood of enrolling in a top-10 % college, and quality of college enrolled in.

As shown in Table 3, four of the mechanisms—the correlation between student resources and caliber, the relationship between resources and information, the relationship

between resources, the number of applications a student submits, and the ability for higher resource students to enhance their apparent caliber—have statistically significant relationships with the likelihood a student enrolls in college. For each of these four, an increase in the correlation is associated with an increase in the gap between the likelihood of students at the 90th and 10th percentile of the resource distribution enrolling in college. Most of the change in this gap comes from the influence of mechanisms on the low end of the resource distribution: for each of the mechanisms that significantly predict the 90-10 gap, none are significant in predicting the 90-50 gap, but three of them significantly predict changes in the 50-10 gap. For example, an addition of one application in the relationship between number of applications submitted and standardized resources increases the 90-10 college gap in probability of college enrollment by 7.4 percentage points, and increases the 50-10 gap by 5.5 percentage points. The number of applications mechanism does not significantly change the 90-50 gap. These results confirm the results in the experimental conditions described above where the number-of-applications mechanism appears particularly to affect the likelihood of college enrollment for students at the lower end of the resource distribution. Additionally, a 0.1 increase in the correlation between resources and caliber increases the 90-10 gap by 5.8 percentage points and the 50-10 gap by five percentage points.

Although the size of the relationships are only about half as large, Table 4 shows that three mechanisms significantly predict gaps in the probability of attending a top 10% college—the correlation between resources and caliber, the ability of high resource students to enhance their apparent caliber, and the relationship between resources and information quality. As in the experimental conditions above, in the case of enrollment in top-10 percent colleges most of the

changes in the gaps appears to come from the top of the resource distribution—the 90-50 gap—rather than the lower half of the distribution.

Finally, Table 5 shows how the 90-10, 50-10, and 90-50 gaps in enrolled-college quality change in response to changes in response to each of the five mechanisms. Four mechanisms are significantly related to the 90-10 gap—the correlation between student resources and caliber, the relationship between resources and information, the relationship between resources and the number of applications a student submits, and the ability for higher resource students to enhance their apparent caliber. Of those four, all are also related to the 90-50 gap, while only the relationship between resources and number of applications is significantly related to the 50-10 gap.

Discussion and Conclusion

In this paper we used agent-based modeling to simulate the college application and selection process. Our model is highly stylized, focusing on only the ways in which student resources and caliber might affect the way in which students behave during the college sorting process. Left out of this model are parameters such as college costs, financial aid, or colleges' strategic admissions decisions based on student resources, race or other factors. Despite the simplifying assumptions we made to create our model, we were able to successfully replicate real-world patterns of application and enrollment. We were then able to conduct experiments by manipulating parameters that determine the specific ways in which student resources might influence student behavior and enrollment outcomes. Based on “virtual counterfactuals” obtained from these experiments, we are able to develop some intuition about the relative

importance of mechanisms that drive observed resource stratification in the college sorting process. We then supplemented our experiments with a Latin Hypercube analysis that allows us to quantify the influence of the mechanisms within our simulated system.

The most striking finding from both the policy experiments and the Latin Hypercube analysis is the very large role that the relationship between student resources and student caliber plays in the socioeconomic sorting of students into schools. This result suggests that changes in the income-achievement gap may substantially affect the degree of college stratification. While clearly important, the relationship between socioeconomic status and achievement is not easy to influence with policy tools, so attempts to decrease enrollment gaps via this relationship may prove relatively intractable in the short term.

Other pathways through which resources affect college destinations, however, may point to more fruitful avenues for policy evaluation and innovation. While none of the effects of these non-achievement gap mechanisms is particularly large on its own, together they are; the accumulation of several small processes has substantial effects on stratification. The bottom right panel in figures 7-9 shows the effect of eliminating all resource pathways except for the resource achievement gap.

Three of the pathways that we explored appear particularly promising for future examination. When tested on its own, reducing the ability of high resource students to enhance their apparent caliber, decreasing disparities in informational quality between high and low resource students, and weakening the link between students' resources and the number of applications that they submit each erodes the relationship between socioeconomic status and college enrollment in our model. These results suggest that student- or institution-level policies

(such as application coaching and college information provision to students in low-income schools or encouraging affirmative action-like policies for dimensions other than race/ethnicity) could have notable impacts on how students sort into colleges.

Interestingly, we did not detect any influence in outcomes generated by only manipulating the difference in college utility evaluations between high and low-resource students. This result suggests that although resources may be related to how students perceive lifetime returns from college quality, this relationship does not appear to drive college stratification given the way we have structured our model. This finding, while perhaps counterintuitive, does comport with the findings of Hoxby and Turner (2013).

While our experiments do not substitute for policy evaluation, they do help to build intuition about the relative importance of difference processes and the importance of evaluating the effects of enacting multiple policies at the same time.

Our model forms a basic framework to which we can add additional complexity and address other policy-relevant questions, including those related to financial constraints. In this paper we focus on how students' information and behavior influence the college sorting process. From this base we can allow colleges to play a more active and strategic role in selecting students as well as add more nuance by including cost considerations for both students and schools.

Much previous work has shown that students are increasingly sorting into colleges in ways that reflect their socioeconomic origins. While the societal dangers of this trend are clear, the mechanisms behind it are harder to disentangle. This paper provides important intuition

about which of these mechanisms are most influential in the college sorting process and thus suggests potential policy interventions that could begin to reverse college sorting trends.

References

- ALON, S. (2009). The evolution of class inequality in higher education: Competition, exclusion, and adaptation. *American Sociological Review*, 74, 731-755.
- ALON, S., & Tienda, M. (2007). Diversity, opportunity, and the shifting meritocracy in higher education. *American Sociological Review*, 72, 487-511.
- ASTIN, A., & Oseguera, L. (2004). The declining “equity” of American higher education. *The Review of Higher Education*, 27, 3, 321-341.
- AUTOR, D. (2010). The polarization of job opportunities in the U.S. labor market: Implications for employment and earnings. *Community Investments*, 23, 2, 11-41
- AVERY, C., & Kane, T. (2004). Student perceptions of college opportunities: The Boston COACH Program. In C. Hoxby, (Ed.) *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 355-391). Chicago: University of Chicago Press.
- BAILEY, M. J., & Dynarski, S. M. (2011) Inequality in postsecondary education. In Duncan, G. C. & Murnane, R. (Eds.) *Whither Opportunity?* (117-132). New York: Russell Sage Foundation.
- BASTEDO, M. N., & Jaquette, O. (2011). Running in place: Low-income students and the dynamics of higher education stratification. *Educational Evaluation and Policy Analysis*, 33, 3, 318-339.
- BAUM, S., Ma, J., & Payea, K. (2010). *Education Pays 2010: The Benefits of Higher Education for Individuals and Society*. Princeton, NJ: The College Board.
- BECKER, B. J. (1990). Coaching for the Scholastic Aptitude Test: Further synthesis and appraisal. *Review of Educational Research*, 60, 373-417.
- BELLEY, P., & Lochner, L. (2007). The changing role of family income and ability in determining

- educational achievement. *Journal of Human Capital*, 1, 1, 37-89.
- BLACK, D., & Smith, J. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, 121, 99-124.
- BOWEN, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton: Princeton University Press.
- BRUCH, E., & Atwood, J. (2012). *Agent-based models in empirical social research*. Working paper.
- BUCHMANN, C., Condron, D. J., & Roscigno, V. J. (2010). Shadow education, American style: Test preparation, the SAT and college enrollment. *Social Forces*, 89, 2, 435-461.
- CABRERA, A. F., & La Nasa, S. M. (2000). Understanding the college-choice process, *New Directions for Institutional Research*, 107, 5-22.
- CARNEIRO, P., Heckman, J.J., & Vytlacil, E. J., (2011). Estimating marginal returns to education. *American Economic Review*, 101, 6, 2754-2781.
- DALE, S., & Krueger, A. B., (2011). Estimating the return to college selectivity over the career using administrative earnings data. National Bureau of Economic Research Working Paper No. 17159.
- ENGBERG, M. E., & Wolniak, G. C. (2010). Examining the effects of high school contexts on postsecondary enrollment. *Research in Higher Education*, 51, 132-153.
- ELLWOOD, D. T., & Kane, T. J. (2000). Who is getting a college education? Family background and the growing gaps in enrollment. In S. Danziger & J. Waldfogel (Eds.) *Securing the Future: Investing in Children from Birth to College* (pp. 283-324). New York: Russell Sage Foundation.

- GRIMM, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., & DeAngelis, D. L. (2006). A standard protocol for describing individual-based and agent-based models. *Ecological Modelling, 198*, 1-2, 115–126.
- GRIMM, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: A review and first update. *Ecological Modelling, 221*, 23, 2760-2768.
- GRODSKY, E., & Jackson, E. (2009). Social stratification in higher education. *Teachers College Record, 111*, 10, 2347-2384.
- GRODSKY, E., & Jones, M. T. (2007). Real and imagined barriers to college entry: Perceptions of cost. *Social Science Research, 36*, 2, 745-766.
- HENRICKSON, L. (2002). Old wine in a new wineskin: College choice, college access using agent-based modeling. *Social Science Computer Review, 20*, 400-419.
- HOEKSTRA, M. (2009) The effect of attending the flagship state university on earnings: a discontinuity-based approach. *The Review of Economics and Statistics, 91*, 4, 717-724.
- HOSSLER, D., Schmit, J., & Vesper, N. (1999). *Going to College: How Social, Economic, and Educational Factors Influence the Decisions Students Make*. Baltimore: The Johns Hopkins University Press.
- HOWELL, J. S. (2010). Assessing the impact of eliminating affirmative action in higher education. *Journal of Labor Economics, 28*, 1, 113-166.

- HOXBY, C. M. (2004). *College choices: The economics of where to go, when to go, and how to pay for it*. Chicago: University of Chicago Press
- HOXBY, C. M. (2009). The changing selectivity of American colleges. *Journal of Economic Perspectives*, 23, 4, 95.
- HOXBY, C. M., & Avery, C. (2012). *The missing "one-offs": The hidden supply of high-achieving, low income students*. National Bureau of Economic Research Working Paper No. 18586.
- HOXBY, C. M., & Turner, S. (2013). *Expanding college opportunities for low-income, high-achieving students*. Stanford Institute for Economic Policy Research Discussion Paper No. 12-014.
- KAREN, D. (2002). Changes in access to higher education in the United States: 1980-1992. *Sociology of Education*, 75, 3, 191-210.
- LONG, M. C. (2008). College quality and early adult outcomes. *Economics of Education Review*, 27, 588-602.
- MAROULIS, S., Guimerà, R., Petry, H., Stringer, M. J., Gomez, L. M., Amaral, L. A. N., & Wilensky, U. (2010). Complex systems view of educational policy research. *Science (Washington)*, 330, 6000, 38-39.
- MCDONOUGH, P. M. (1997). *Choosing Colleges: How Social Class and Schools Structure Opportunity*. Albany: State University of New York Press.
- PAULSEN, M. B., & Smart, J. C. (2001) (Eds.). *The Finance of Higher Education: Theory, Research, Policy, and Practice*. New York: Agathon Press.

PERNA, L. W. (2006). Studying college choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 21, pp. 99-157). New York: Springer.

PERNA, L. W., & Titus, M. A. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *The Journal of Higher Education*, 76, 5, 485-518.

POWERS, D. E., & Rock, D. A. (1999). Effects of coaching on SAT I: Reasoning test scores. *Journal of Educational Measurement*, 36, 2, 93-118.

REARDON, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In Duncan, G.C. & Murnane, R. (Eds.) *Whither Opportunity?* New York: Russell Sage Foundation.

REARDON, S. F., Baker, R., B., & Klasik, D. (2012). *Race, income, and enrollment patterns in highly selective colleges, 1982-2004*. Center for Education Policy Analysis.

RIVERA, L. (2011). Ivies, extracurriculars, and exclusion: Elite employers' use of educational credentials. *Research in Social Stratification and Mobility*, 29, 1, 71-90.

RODERICK, M., Nagaoka, J., Coca, V., & Moeller, E. (2008). *From high school to the future: Potholes on the road to college*. Chicago: Consortium on Chicago School Research at the University of Chicago.

RODERICK, M., Nagaoka, J., Coca, V., & Moeller, E. (2009). *From high school to the future: Making hard work pay off*. Chicago: Consortium on Chicago School Research at the University of Chicago.

SEGOVIA-JUAREZ, J. L., Ganguli, S., & Kirschner, D. (2004). Identifying control mechanisms of granuloma formation during *M. tuberculosis* infection using an agent-based model. *Journal of Theoretical Biology*, 231, 3, 357-376.

US DEPARTMENT OF EDUCATION (2006). National Center for Education Statistics. Education Longitudinal Study (ELS), 2002 and 2006: Base Year through Second Follow-up.

WECKER, M. (2012). Where the Fortune 500 CEOs went to school. *U.S. News and World Report*, May 14: online.

ZUMETA, W. (2011). *Does the U.S. need more college graduates to remain a world class economic power?* Presented at the Conference "Preparing Today's Students for Tomorrow's Jobs in Metropolitan America."

Endnotes

¹ Many of our parameter estimates come from the Education Longitudinal Study of 2002 (ELS:2002). ELS:2002 is a nationally representative data set collected by The National Center for Education Statistics. It follows 10th grades in 2002 through secondary and postsecondary education and includes high school transcript data, surveys of students and parents, and postsecondary application behavior. We set r to 0.3 in our baseline model, a conservative estimate based on ELS:2002, where the observed correlation between students' SAT scores and the socioeconomic status index is 0.43 (US Department of Education 2006). We set b to 0.1 based on research that shows that students who take SAT-coaching classes typically raise their SAT scores by approximately 25 points, which is about 12 percent of a standard deviation of SAT scores on the 1600-point SAT scale (Becker 1990; Buchmann, Condron, & Roscigno 2010; Powers & Rock 1999). We set c to 0.5 based on the relationship between the socioeconomic index and the number of schools a student applies to in the ELS:2002 data set (US Department of Education 2006).

² Every figure that depicts one of our three main outcomes aggregates all 100 runs, with a line showing a running mean of outcome values by student resource percentiles across runs bounded by a shaded area indicating standard error values.

³ A closer examination of the sets of colleges to which students apply under these different experimental conditions can help us to understand why we are seeing these patterns. The

figures in Appendix C show the maximum, minimum and mean quality of schools that high and low resource students at all points of the caliber distribution apply to.

Table 1

Model Parameters (baseline model)

Parameter	Value	Source
<i>Basic model set up</i>		
Number of students	8000	N/A
Number of colleges	40	N/A
College capacity	150 students/college	N/A
Student to seat ratio	4:3	ELS
College quality	$quality \sim N(1070, 130)$	ELS
Student caliber	$caliber \sim N(1000, 200)$	College Board
Student resources	$resources \sim N(0, 1)$	N/A
Correlation between resources and caliber	$r=0.3$	ELS
Quality reliability (how well students see college quality)	$0.7 + a*resources; a=0.1$	N/A
Own caliber reliability (how well students see their own caliber)	$0.7 + a*resources; a=0.1$	N/A
Caliber reliability (how well schools see student caliber)	0.8	N/A
Apparent caliber (perceived caliber, increased or decreased through “caliber enhancement”)	$perceived\ caliber + b*resources; b=0.1$	Becker 1990; Buchmann, Condron, and Roscigno 2010; Powers and Rock 1999
Number of Applications	$4 + INT(c*resources); c=0.5$	ELS
Student evaluation of college utility	$-250 + d + (1+ e)*perceived\ quality; d=-500, e=0.5\ if\ resources>0$	N/A

Note: Quality and caliber reliability bound by minimum values of 0.5 and maximum values of 0.9.

Table 2

Resource Pathway Parameters by Experiment

Experiment	Parameter Values					
	<i>r</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
All resource pathways off	0	0	0	0	0	0
Baseline model	0.3	0.1	0.1	0.5	-500	0.5
Experiment 1	0	0.1	0.1	0.5	-500	0.5
Experiment 2	0.3	0.1	0	0.5	-500	0.5
Experiment 3	0.3	0.1	0.1	0	-500	0.5
Experiment 4	0.3	0	0.1	0.5	-500	0.5
Experiment 5	0.3	0.1	0.1	0.5	0	0
Experiment 6	0.3	0	0	0	0	0

Table 3

Latin Hypercube Sensitivity Analysis of Parameters of Interest on Gaps in Probability of Enrollment

VARIABLES	90 th	90 th	50 th	Parameter Space	
	Percentile – 10 th Percentile	Percentile – 50 th Percentile	Percentile – 10 th Percentile	Min	Max
Correlation(Resources, Caliber)	0.581*** (0.039)	0.096 (0.066)	0.485*** (0.055)	.1	.5
Resources/Information Relationship	0.405** (0.079)	0.250 (0.136)	0.155 (0.113)	0	.2
Resources/#Apps Relationship	0.074*** (0.006)	0.019 (0.010)	0.055** (0.008)	0	2
Utility Slope Differential	0.003 (0.005)	-0.001 (0.009)	0.004 (0.008)	0	2
Resources/Application Enhancement	0.580*** (0.056)	0.082 (0.096)	0.498** (0.080)	0	.2
Constant	0.032+ (0.013)	0.045 (0.023)	-0.012 (0.019)		

Note: Standard errors in parentheses. *** p<0.001; ** p<0.01; * p<0.05; + p<0.1.

Table 4

Latin Hypercube Sensitivity Analysis of Parameters of Interest on Gaps in Probability of Enrollment in Top-10% College

VARIABLES	90 th Percentile –	90 th Percentile –	50 th Percentile –	Parameter Space	
	10 th Percentile	50 th Percentile	10 th Percentile	Min	Max
Correlation(Resources, Caliber)	0.238** (0.045)	0.269** (0.045)	-0.031 (0.018)	.1	.5
Resources/Information Relationship	0.244+ (0.092)	0.279* (0.093)	-0.035 (0.038)	0	.2
Resources/#Apps Relationship	0.013 (0.007)	0.007 (0.007)	0.005 (0.003)	0	2
Utility Slope Differential	0.005 (0.006)	0.004 (0.006)	0.001 (0.003)	0	2
Resources/Application Enhancement	0.234* (0.065)	0.265* (0.066)	-0.031 (0.027)	0	.2
Constant	0.039+ (0.015)	-0.007 (0.016)	0.046** (0.006)		

Note: Standard errors in parentheses. *** p<0.001; ** p<0.01; * p<0.05; + p<0.1.

Table 5

Latin Hypercube Sensitivity Analysis of Parameters of Interest on Gaps in Enrolled-College Quality

VARIABLES	90 th Percentile –	90 th Percentile –	50 th Percentile –	Parameter Spa	
	10 th Percentile	50 th Percentile	10 th Percentile	Min	Ma
Correlation(Resources, Caliber)	192.432*** (17.737)	164.271*** (7.006)	28.160 (13.920)	.1	.5
Resources/Information Relationship	104.922* (36.453)	99.210** (14.400)	5.713 (28.608)	0	.2
Resources/#Apps Relationship	16.158** (2.699)	6.749** (1.066)	9.409* (2.118)	0	2
Utility Slope Differential	-0.028 (2.441)	-0.994 (0.964)	0.966 (1.916)	0	2
Resources/Application Enhancement	140.959** (25.743)	132.554*** (10.169)	8.405 (20.203)	0	.2
Constant	22.704* (6.089)	0.234 (2.405)	22.469** (4.779)		

Note: Standard errors in parentheses. *** p<0.001; ** p<0.01; * p<0.05; + p<0.1.

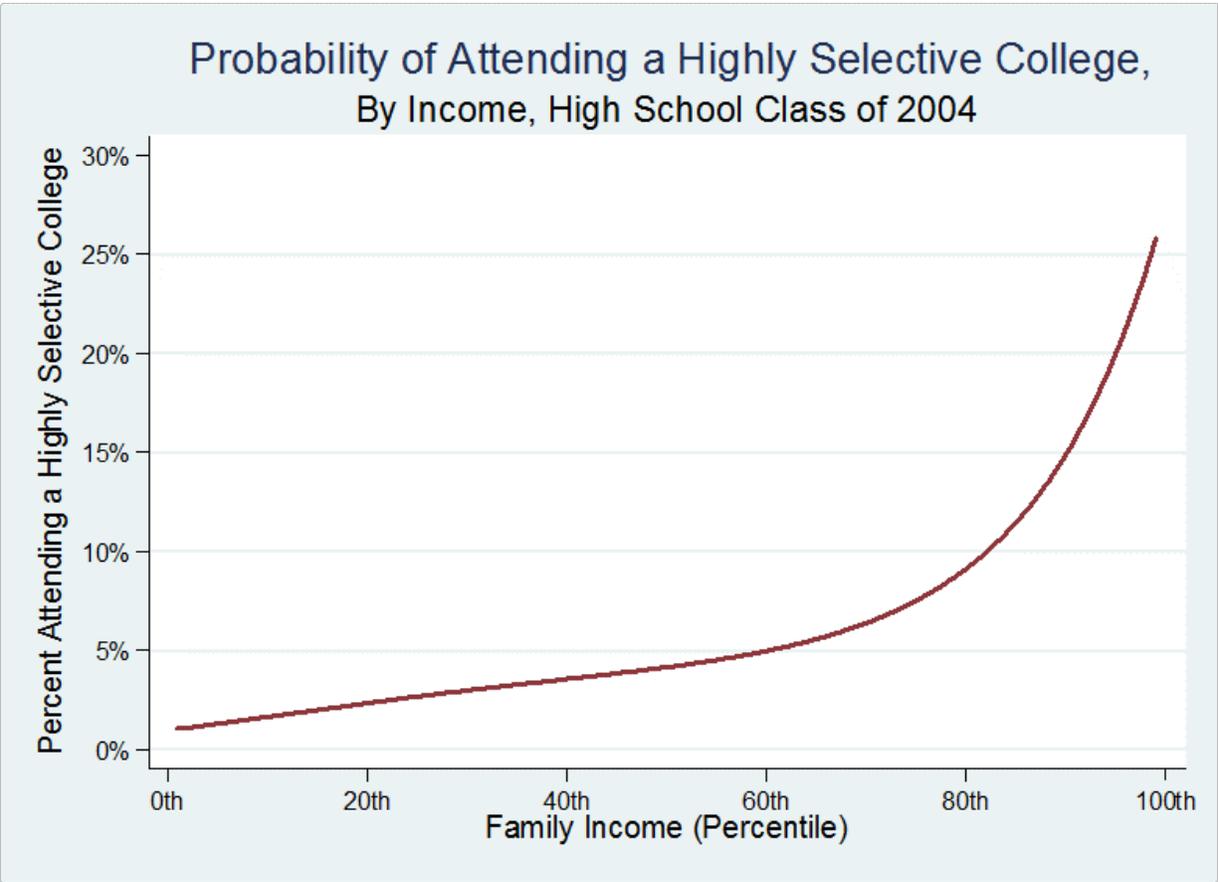


Figure 1. Probability of attending a highly selective college, by income, high school class of 2004.

Source: Authors' calculations from ELS:2002.

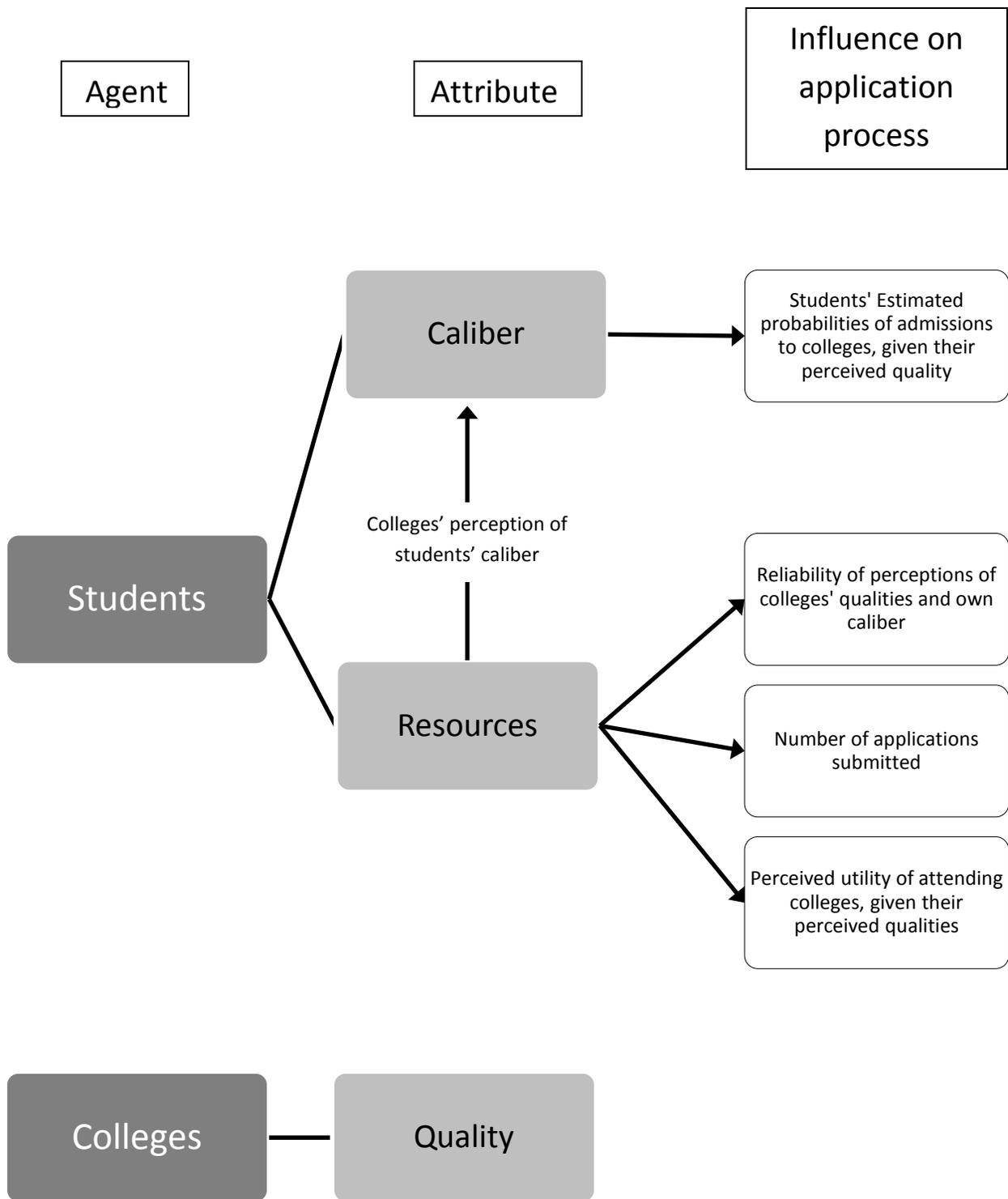
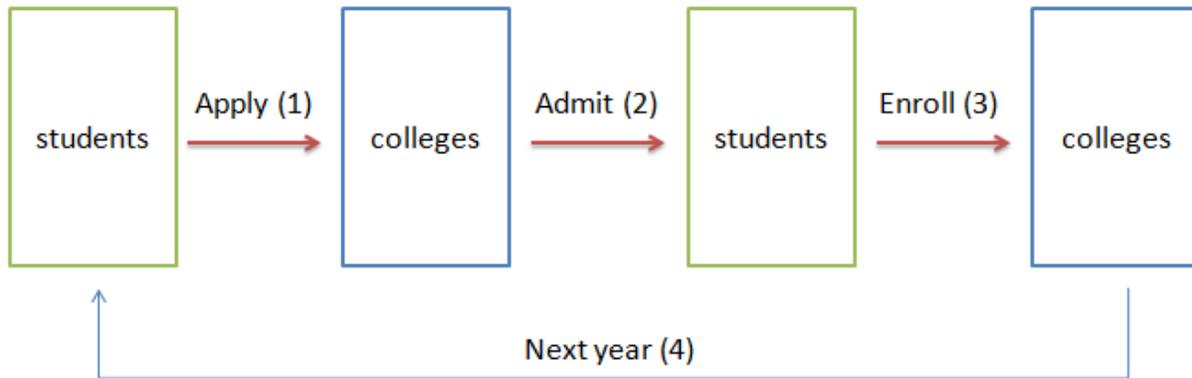


Figure 2. Agent types, attributes, and the attributes and model elements that they influence.



- 1) Limited by a finite number of applications, students apply to a set of colleges that maximizes their expected utility
- 2) Colleges order applicants by perceived caliber and admit a number of students based on prior enrollment yield
- 3) Students enroll in the college with the highest perceived utility to which they have been accepted
- 4) College selectivity, yield, and quality are updated based on outcomes of steps (1) – (3)

Figure 3. Overview of processes in the agent-based model

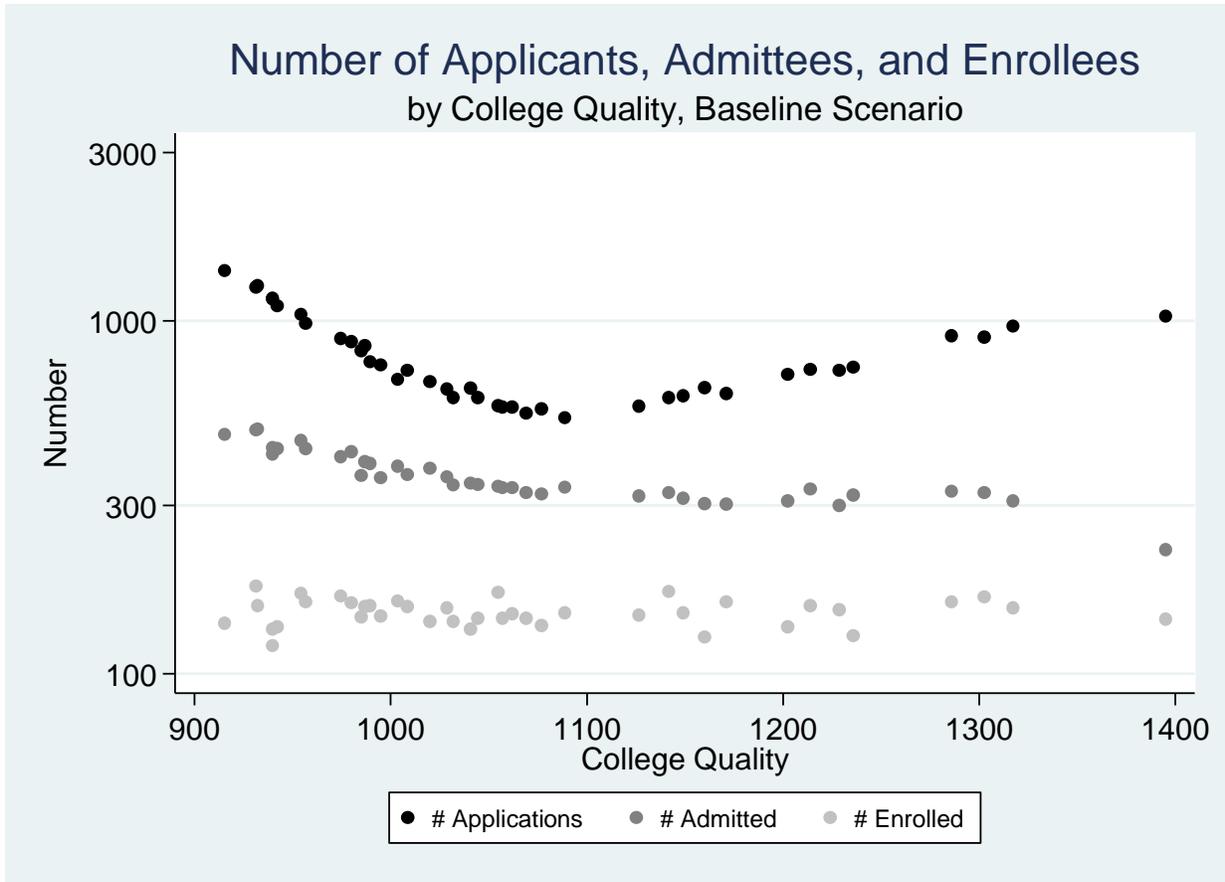


Figure 4. Number of applications, admittees, and enrollees, by college quality, baseline scenario.

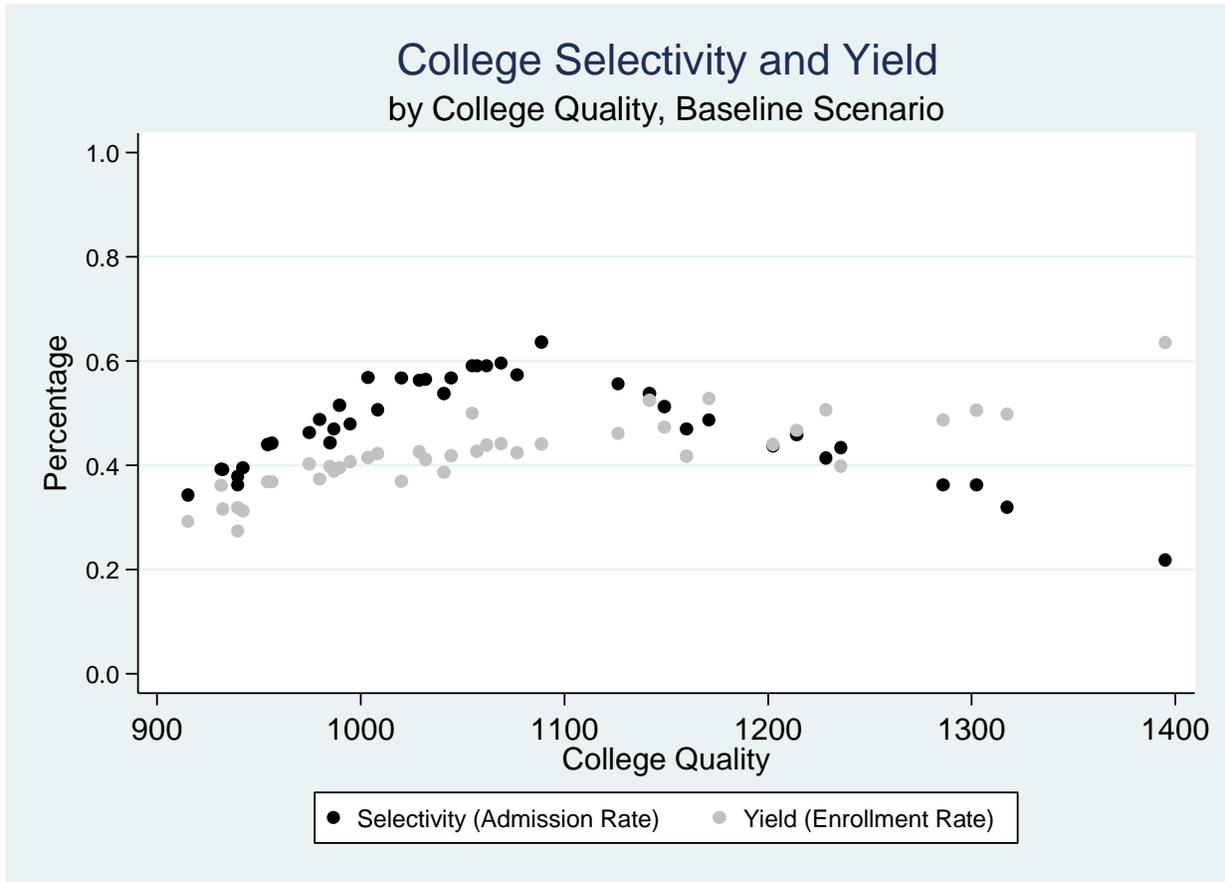


Figure 5. College selectivity and yield, by college quality, baseline scenario.

Student Outcomes by Student Resource Percentile

Baseline Model: All Resource Pathways Turned On

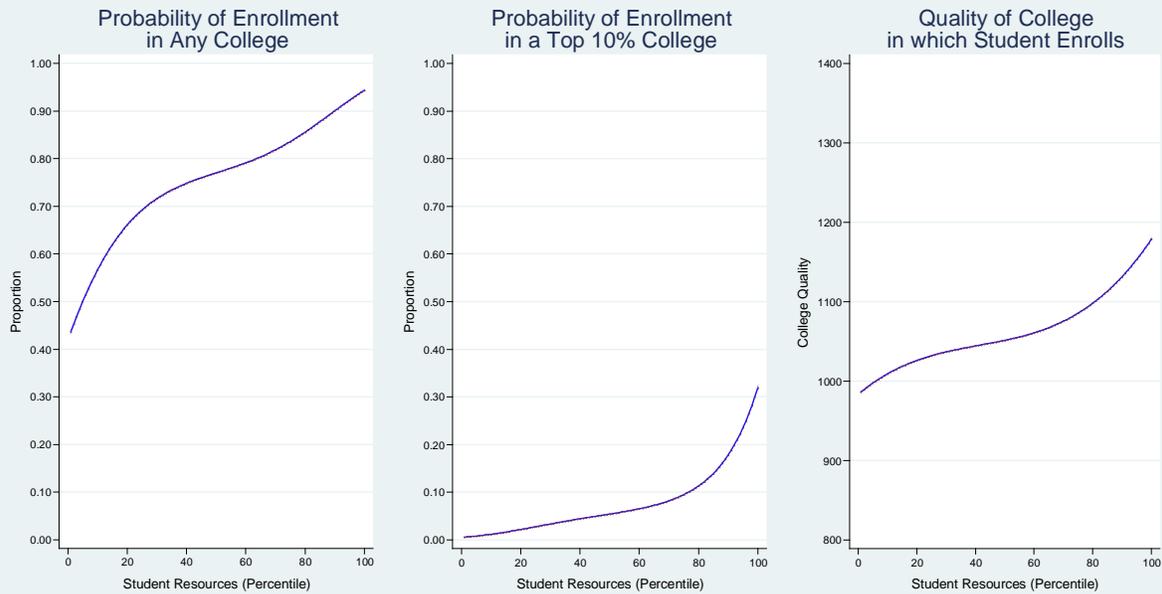


Figure 6. Probability of enrolling in any college, probability of enrolling in a top-10% college and quality of college enrolled in, by student resources percentile, year 30, baseline model.

Probability of Enrollment in Any College by Student Resource Percentile

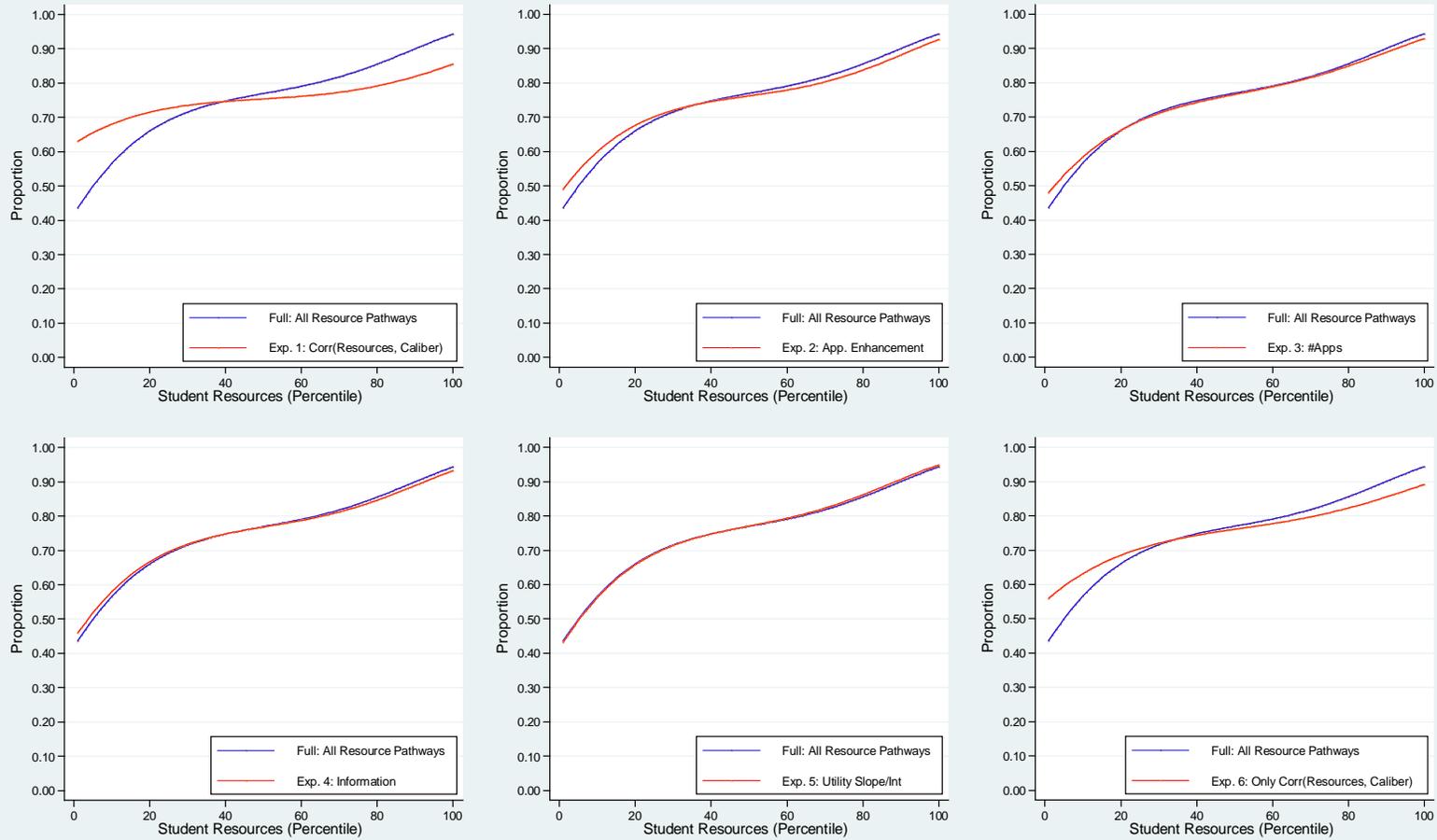


Figure 7. Probability of enrolling in any college, by student resource percentile and resource pathway, year 30.

Probability of Enrollment in a Top 10% College by Student Resource Percentile

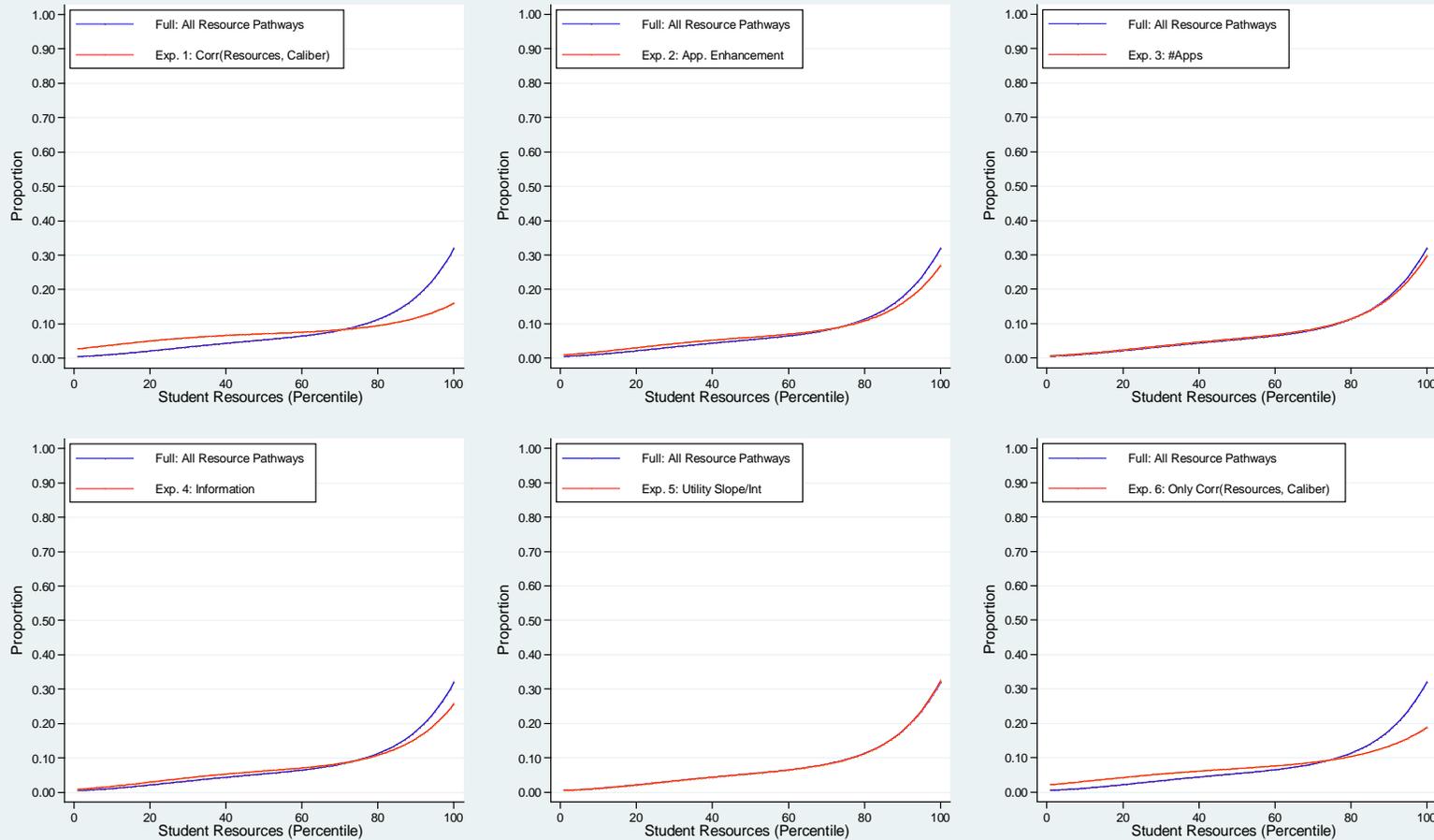


Figure 8. Probability of enrolling in a top-10% college, by student resource percentile and resource pathway, year 30.

Quality of College in which Student Enrolls by Student Resource Percentile

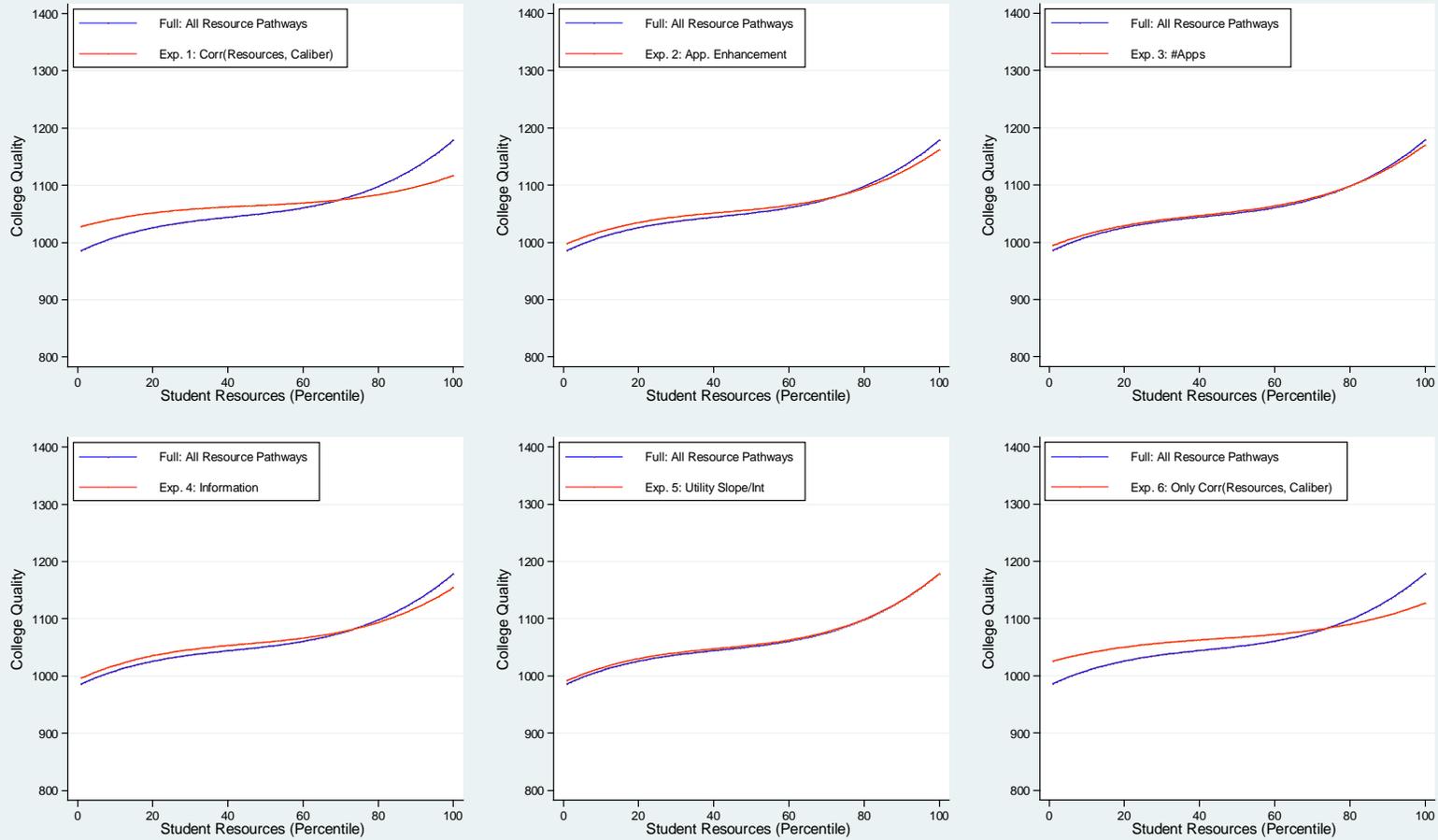


Figure 9. Quality of college enrolled in, by student resource percentile and resource pathway, year 30

Appendix A

Initialization

At the start of each model run, we generate J colleges with m available seats per year (for the sake of simplicity, m is constant across colleges). During each year of the model run, a new cohort of N students engages in the college application process. Initial college quality (Q), each student cohort's caliber (C), and each student cohort's resources (R) are normally distributed. We allow for a specified correlation between C and R .³ The values used for these parameters are specified in Table 1. We select these values to balance computational speed and distribution density as well as to match what we observe from real-world data (ELS 2002).

Submodels

Application. During this stage of our model, students generate an application portfolio, with each student selecting n_s colleges to which they will apply. Every student observes each college's quality (Q_c) with some amount of noise (u_{cs}), which represents both imperfect information and idiosyncratic preferences, and then uses perceived college quality (Q^*_{cs}) to evaluate the potential utility of attendance:

$$Q^*_{cs} = Q_c + u_{cs}; u_{cs} \sim N(0, \tau_s) \quad (\text{A.1})$$

$$U^*_{cs} = a_s + b_s(Q^*_{cs}) \quad (\text{A.2})$$

where a_s is the intercept of a linear utility function and b_s is the slope; both intercept and slope may differ between students. Students do not know their own caliber perfectly, but view it with both augmentation and noise:

$$C^*_s = C_s + c_s + e_s; e_s \sim N(0, \sigma_s) \quad (\text{A.3})$$

where c_s represents enhancements to caliber that are unrelated to caliber itself (e.g. test preparation, or application essay consultation) and e_s represents uncertainty. The values that are used for these parameters and their relationships with student resources are listed in Table 1. Based on their noisy observations of their own caliber and college quality, students estimate their probabilities of admission into each college:

$$P_{cs}=f(C_s^*-Q_{cs}^*) \tag{A.4}$$

where f is a function based on admission patterns over the prior 5 years. In each year f is estimated by fitting a logit model predicting the observed admissions decisions using the difference between (true) student caliber and college quality for each submitted application over the past 5 years. During the first 5 years of our simulation, the admission probability function has an α of 0 and a β of -0.015. These values were selected based on observing the admission probability function over a number of model runs; the starting values do not influence the model end-state, but do influence how quickly the function (and the model itself) stabilizes. A student's expected utility of applying to one college is the product of the estimated probability of admission and the estimated utility of attendance.

Students apply to sets of schools that maximize their overall expected utility. For example, if a student chooses to apply to three colleges, then she will select the set of three colleges that they believe has the greatest combined expected utility. In principle, this means that a student agent in the model computes the expected utility associated with applying to every possible combination of three colleges in the model, and then chooses the set that maximizes this expected utility. We develop a fast algorithm, described in Appendix D, that achieves this maximization without requiring the agent to compute and compare all possible

application portfolios. The assumption of rational behavior is an abstraction that facilitates focus on the elements of college sorting that we wish to explore. We recognize that real-world students use many different strategies to determine where they apply (e.g. Hoxby & Avery 2012).

Admission. Colleges observe the apparent caliber ($C_s + c_s$) of applicants with some amount of noise (like the noise with which students view college quality, this also reflects both imperfect information as well as idiosyncratic preferences):

$$C_{cs}^{**} = C_s + c_s + w_{cs}; w_{cs} \sim N(0, \phi_s) \quad (\text{A.5})$$

Colleges rank applicants according to C_{cs}^{**} and admit the top s_c applicants. In the first year of our model run, college's expected yield (the proportion of admitted students that a college expects to enroll) is given by:

$$Yield_c = 0.2 + .06 * \text{College Quality Percentile} \quad (\text{A.6})$$

with the lowest-quality college expecting slightly over 20% of admitted students to enroll and the highest quality college expecting 80% of admitted students to enroll. Colleges thus admit $m/Yield_c$ students in order to try to fill m seats. After the first year of a model run, colleges are able to use up to 3 years of enrollment history to determine their expected yield, with $Yield_c$ representing a running average of the most recent enrollment yield for each college.

Enrollment. Students enroll in the school with the highest estimated utility of attendance (U_{cs}^*) to which they were admitted.

Iteration. Colleges' quality values (Q_c) are updated based on the incoming class of enrolled students before the next year's cohort of students begins the application process:

$$Q'_c = 0.9 * Q_c + 0.1 * \text{College mean } (C_s) \quad (\text{A.5})$$

We run our model for 30 years (this appears to be a sufficient length of time for our model to reach a relatively stable state for the parameter specifications that we explore).

Appendix B

The following figures are based on IPEDS admissions statistics from the 2010-2011 admissions cycle. They were used to help confirm the calibration of our model. Figures B1 and B2 are intended to be compared to Figures 3 and 4, respectively.

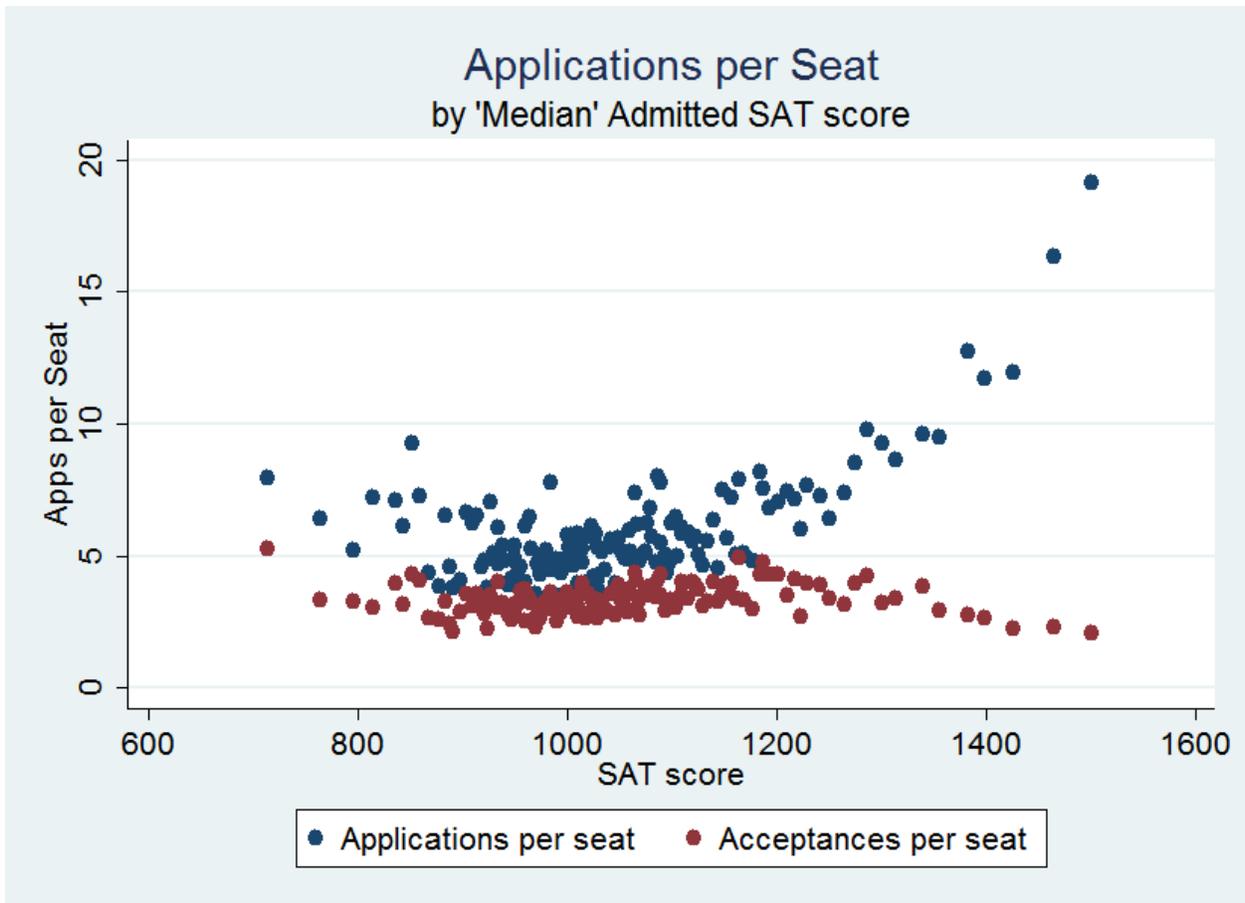


Figure B1. Applications and acceptances per enrolled student, by 'median' admitted SAT score.

From IPEDS data from 2010-2011. Median SAT score is approximated using half of the sum of the 25th and 75th percentile of SAT score.

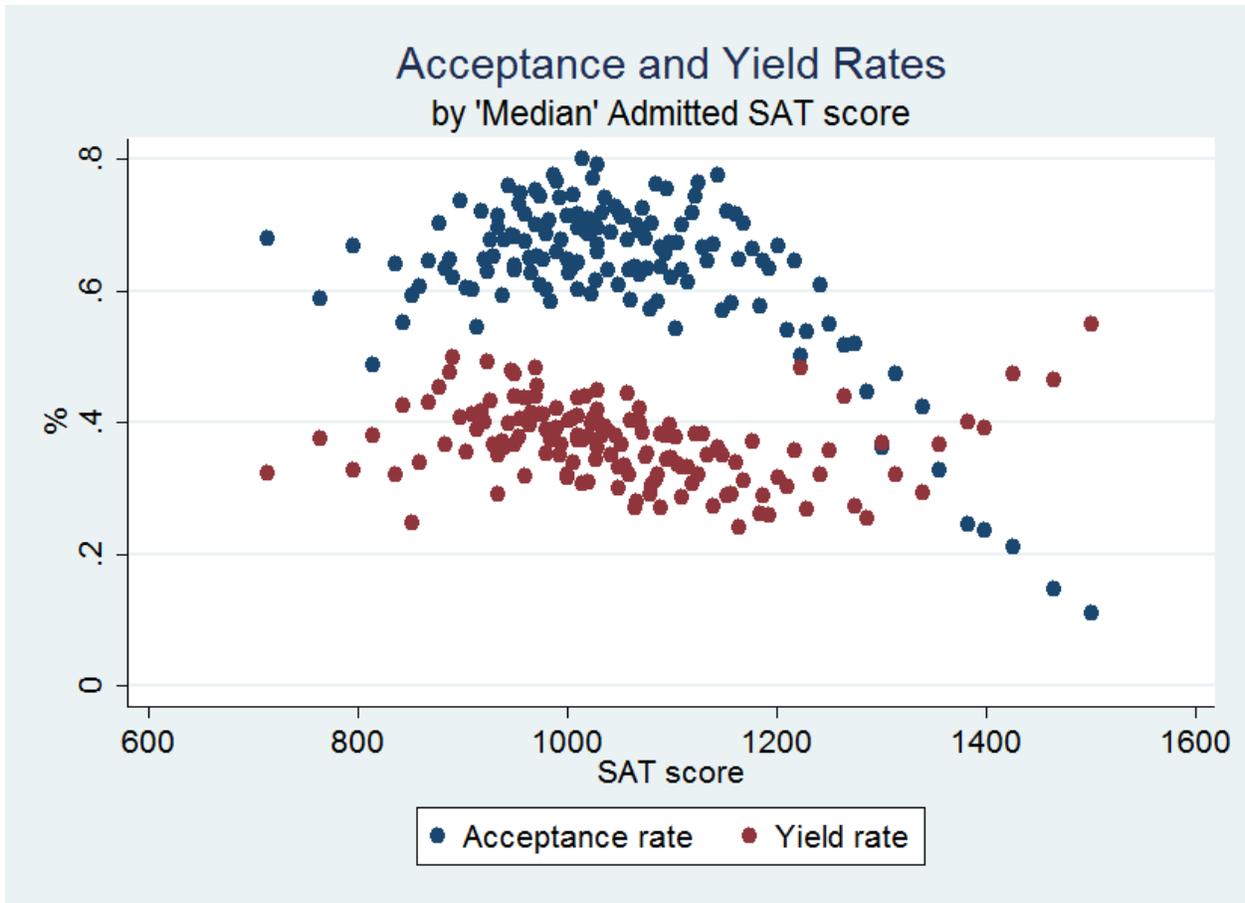


Figure B2. Acceptance and yield rates, by 'median' admitted SAT score. From IPEDS data from 2010-2011 Median SAT score is approximated using half of the sum of the 25th and 75th percentile of SAT score

Appendix C

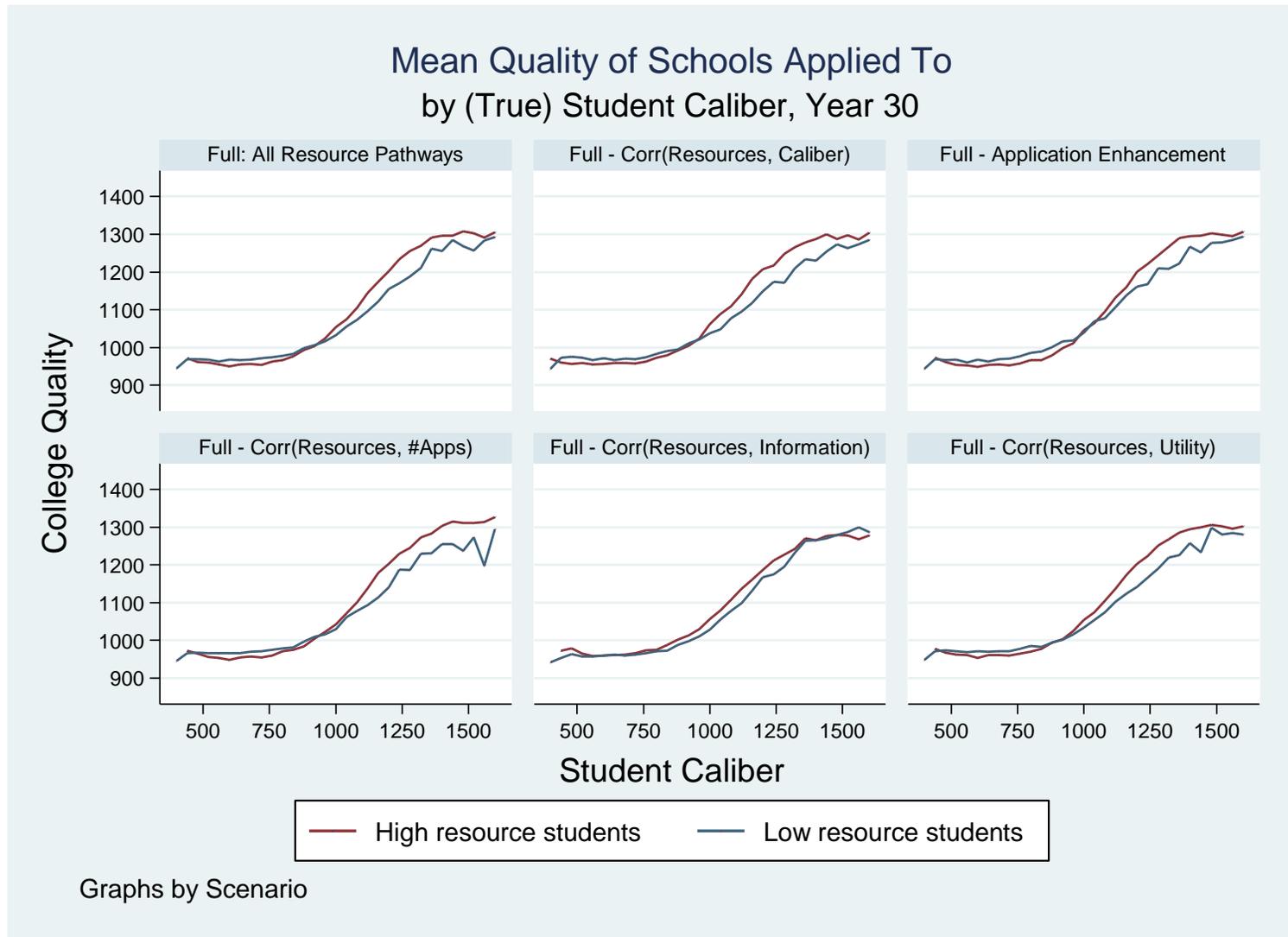
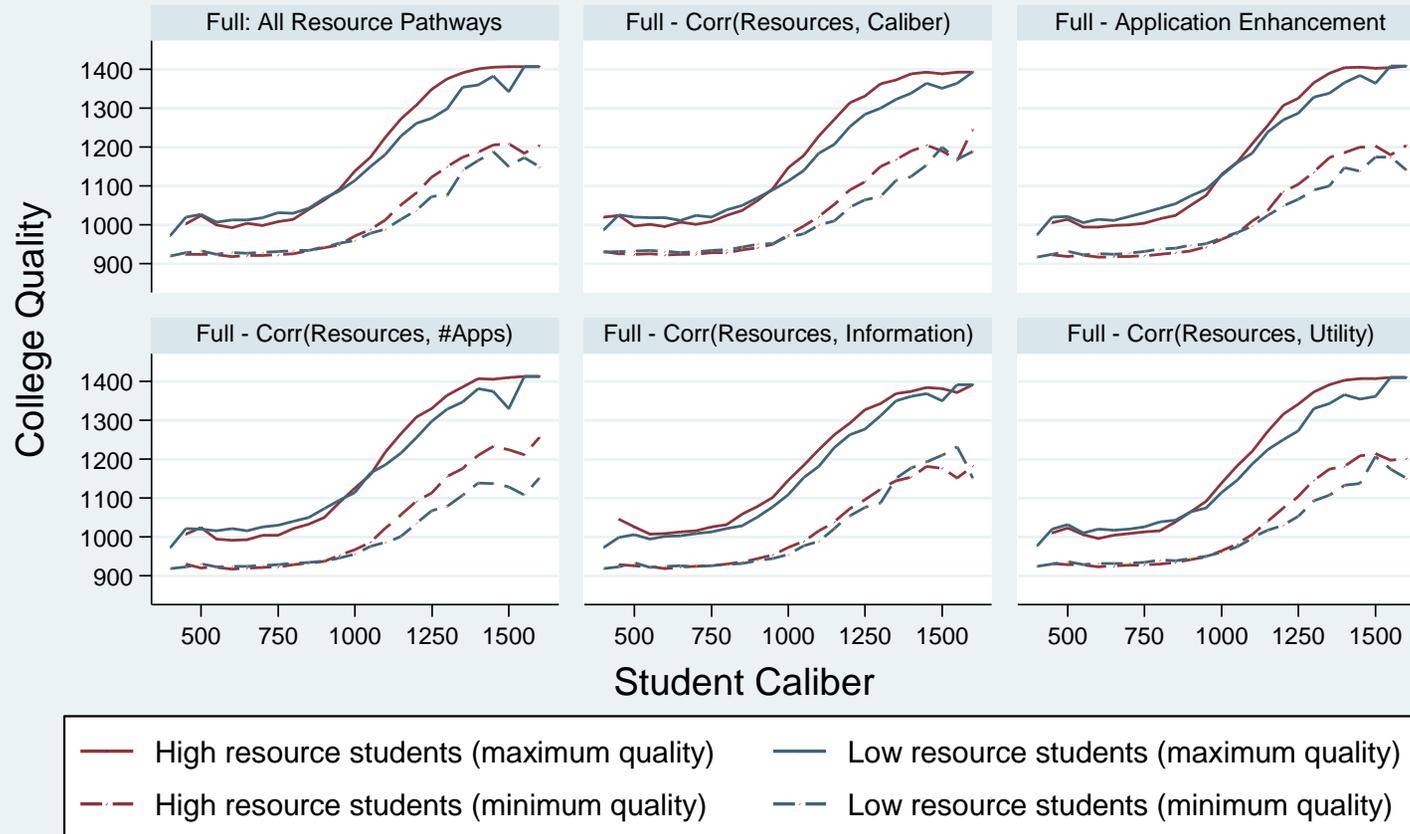


Figure C1. Mean quality of schools applied to, by (true) student caliber and scenario, year 30. Single run.

Average Maximum and Minimum Quality of Schools Applied To by (True) Student Caliber, Year 30



Graphs by Scenario

Figure C2. Average maximum and minimum quality of schools applied, by (true) student caliber and scenario, year 30. Single run.

Appendix D

Optimal College Portfolio Algorithm

Notation. Let $i = 1, N$ index students and let $j = 1, J$ index colleges. Let Q_j denote college quality (or utility). Suppose student i applies to some set $\mathbf{A}_i = \{a_1, a_2, \dots, a_n\}$ of n schools (where the set is ordered such that $Q_{a_1} < Q_{a_2} \dots < Q_{a_n}$). Denote the set of schools to which student i is admitted as \mathbf{B}_i , where $\mathbf{B}_i \subseteq \mathbf{A}_i$. Assume that a student will enroll in \mathbf{C}_i , the highest quality school to which she is admitted.

Now let $U_i(\mathbf{B}_i) = \max(Q_j, \dots, Q_k \mid j, \dots, k \in \mathbf{B}_i)$. In other words $U_i(\mathbf{B}_i)$ is the quality of the school that student i will enroll in if i is admitted to the set of schools \mathbf{B}_i . Define $U_i(\emptyset) = 0$ (if i is not admitted to any schools, then i experiences school quality of 0).

Now define $E_i(\mathbf{A}_i) = E[U_i(\mathbf{B}_i) \mid \mathbf{A}_i] = E[U_i(\mathbf{C}_i) \mid \mathbf{A}_i]$. That is, let $E_i(\mathbf{A}_i)$ indicate the expected value of the quality of the college student i will enroll in if she is applies to the set of colleges \mathbf{A}_i .

The student's challenge is to pick, given a number of applications, the set \mathbf{M}_i^n of n schools that will maximize $E_i(\mathbf{A}_i)$.

Assumptions. Let $A_{ij} = 1$ if student i is admitted to school j and $A_{ij} = 0$ otherwise. Let $P_{ij} = P_i(Q_j) = \Pr(A_{ij} = 1)$ be the probability that a given student will be admitted to a school of quality Q conditional on application to that school. We will assume that P_i is a monotonically decreasing function of Q for all i . Assume the utility of enrolling in no college is 0. Assume that the quality of available colleges range from a value of 0 to 1 (the scale is arbitrary), and that the function $P_i(Q) \cdot Q$ has a single local maximum on the interval (0,1). That is, there is at most a single point m_i in the interval [0,1] where $d(P_i Q)/dQ = 0$. Moreover, if there is such a point, then $d(P_i Q)/dQ > 0$ if 0

$\leq Q_j < m_i$ and $d(P_i Q)/dQ < 0$ if $m_i < Q_j \leq 1$. If there is no such point, then $P_i Q$ is monotonically increasing or decreasing over the interval $[0,1]$.

This last assumption about a single local maximum is met if $P_i(Q) = (1 + e^{a + bQ})^{-1}$ and $b > 0$ (i.e., if the probability of admission is described by a logit function of $-Q$).

Algorithm. Now, it can be shown that, under these conditions, and a given number of applications n , student i should apply the set of school \mathbf{M}_i^n defined the following way:

First, compute $E_i(j) = P_i(Q_j) \cdot Q_j$ for all $j \in 1, J$. Apply to the school with the maximum value of $E_i(j)$. Call this school a_1 . If student i were to apply to a single school, this would optimize the student's expected utility. In other words, $\mathbf{M}_i^1 = \{a_1\}$, and $E(\mathbf{M}_i^1) = P_i(Q_{a_1}) \cdot Q_{a_1}$.

Continue this process recursively to identify schools a_2, \dots, a_n . For $k = 2, \dots, n$, use the following algorithm. At the k^{th} step, check if there are schools with $Q_j > Q_{a_{k-1}}$. If not, pick the $n - (k - 1)$ highest quality schools that i has not yet applied to and include them in \mathbf{A}_i . This will complete the set \mathbf{M}_i^n that maximizes $E_i(\mathbf{A}_i)$.

If, however, there are schools with $Q_j > Q_{a_{k-1}}$, then for each such school j , compute $E_i(\mathbf{M}_i^{k-1}, j) = (1 - P_{ij})E_i(\mathbf{M}_i^{k-1}) + P_{ij}Q_j$. Pick the school with the maximum value of $E_i(a_i, j)$, and apply to this school as well. Call this school a_k . Now $\mathbf{M}_i^k = \{\mathbf{M}_i^{k-1}, a_k\} = \{a_1, a_2, \dots, a_k\}$.

Under these rules, \mathbf{M}_i^n will be the set of n schools that maximize the expected quality of the school in which student i will enroll.

We thus have that

$$\begin{aligned} E_i(\mathbf{M}_i^k) - E_i(\mathbf{M}_i^{k-1}) &= (1 - P_{iak}) E_i(\mathbf{M}_i^{k-1}) + P_{iak} Q_{ak} - E_i(\mathbf{M}_i^{k-1}) \\ &= P_{iak} (Q_{ak} - E_i(\mathbf{M}_i^{k-1})) \\ &> 0 \end{aligned} \tag{D.1}$$

because $Q_{ak} > \max(Q_{a1}, \dots, Q_{ak-1}) \geq E_i(\mathbf{M}_i^{k-1})$. In other words, adding a k^{th} school of higher quality than any of the $k - 1$ schools in \mathbf{M}_i^{k-1} will always increase the expected quality of one's college enrollment.