How has college choice changed? An update for the 2000s using the conditional logit model

Benjamin T. Skinner
Vanderbilt University

PMB # 414
230 Appleton Place
Nashville, TN 37203-5721

I would like to thank Brent Evans and Laura Perna for their insightful comments on previous iterations of this paper presented at the Association for Education Finance and Policy and the Association for the Study of Higher Education annual conferences, respectively. I would also like to thank Will Doyle, Dale Ballou, Richard Blissett, and participants at the above conferences for their comments and suggestions. All errors and mistakes in interpretation remain my own.

Abstract

In this paper I investigate the college enrollment decisions of nationally representative cohort of students who first attended in the mid-2000s. Fitting conditional logistic choice models, I find in line with prior research that while price, distance, and match are important in the choice between colleges, characteristics of the most-likely college are not consistently predictive of the likelihood of matriculation when controlling for student characteristics. I take advantage of rich data to extend prior work by modeling discrete steps in the enrollment decision process: application and enrollment conditional on application. Comparing parameters across models, I find choice characteristics generally have the strongest effects in the application stage. I also explore heterogeneous effects for two subpopulations: those with high SAT scores and those with low family income. My results support other research that suggests students may self-select out of potentially better college matches due to lack of information about actual costs or limited geographic opportunity.

1 Introduction

Though economic stability—let alone mobility—in our modern economy increasingly demands training and credentialing beyond a high school diploma, postsecondary attendance in the United States remains optional. Rather than make college compulsory, our nation's higher education policies broadly continue in the vein of those which were laid out in the Higher Education Act of 1965: to incentivize college-going among students believed most likely to benefit but who otherwise may not have the means. Toward this end, federal and state governments have enacted policies that attempt to reduce college costs by funding financial aid programs and building low-cost regional and community colleges.

Human capital models of college attendance, which argue that enrollment is inversely related to costs (Becker 2009; Turner 2004), lend credence to such programs. A number of empirical studies have found that students are sensitive to price, distance, and match when applying to colleges (Hoxby and Avery 2012; Lovenheim and Reynolds 2011; Niu and Tienda 2008; Drewes and Michael 2006; Avery and Hoxby 2004; Long 2004; Manski and Wise 1983) and therefore may be positively influenced by financial aid and increases in their schooling options. Investigating cross-generational trends, Long (2004) finds that while students who enrolled in the 1970s, 1980s, and 1990s preferred lower-cost and closer institutions when choosing between colleges, the strength of these effects decreased over time as college quality and student-college match grew to matter more. Though the costs associated with a 1972 graduate's most-likely college were predictive of his or her likelihood of enrolling in college, this was no longer the case for the class of 1992. Long cites this changing relationship to argue that cost-reduction policies such as increases in the availability of credit through financial aid may have been successful in relaxing restraints on student college choice.

My study serves as an update to this earlier work as it applies the same conditional logistic choice model (McFadden 1973) to college enrollment decisions among a new generation of students who graduated from high school in the early 2000s. Using student-level data from the Education Longitudinal Study of 2002 (ELS2002) (National Center for Education Statistics

2016a) along with postsecondary institution data gathered from the Integrated Postsecondary Education Data System (IPEDS) (National Center for Education Statistics 2016b) and the Delta Cost Project (The Delta Cost Project 2016), I model the between-college decision for nearly 6,800 students who enrolled within two years of earning their high school or general equivalency diplomas as a choice among 3,406 postsecondary institutions (approximately 23 million student-college choice pairs). Because the data include not only indicators for enrollment but also information regarding applications and acceptances, I am able to separate the enrollment decision into three separate outcomes: attendance, application, and attendance conditional on application. Alongside those of the full sample, I present results for two subsets of students: those with higher SAT scores ($>$ 1100 combined math and verbal) and those with low family income ($<$ \$25,000 per year).

Using the estimated parameters from the conditional logit choice models, I am able predict the probability of each student's enrollment—attenders and non-attenders both—at every college in the alternative option set. From these predictions, I select each student's most-likely college and fit a logistic regression model that includes characteristics of the college and student on the right-hand side in order to estimate the overall likelihood of college enrollment among the sample within two years of high school graduation. As a comparison to the three most-likely college models, I fit the same logit model using the characteristics of the nearest public two-year college. I present results for all four models alongside those generated for a subset of students who lie on the margins of attendance due to low SAT scores ($<$ 900 combined math and verbal) and/or low family income.

In this paper I follow Long's (2004) framework, modeling decisions, and covariate selection closely so that my results be as comparable as possible to those reported for earlier cohorts. Throughout, I note instances in which changes in the higher education landscape or available data require me to deviate from Long's specifications. I believe that replicating this prior work as I have done allows me to offer an important and clearly comparable update to the literature on the college enrollment decision.

Previewing my results, I find that costs have continued to decline in relevance for between college choice, with an increase in cost of $1000 suggesting approximately a 16% to 21% reduction in the odds of enrolling in a given institution. I find student-college match to be increasingly important. While students in the recent cohort show even stronger reluctance than those in the past to choose institutions with lower median SAT scores than their own (undermatch), they are comparatively more willing to choose schools at which the median SAT score is higher (overmatch). The newest cohort of students also show a favorable response to institutions with higher proportions of tenured and tenure-track faculty. They, like their 1992 predecessors, also remain more likely to choose a two-year over a four-year institution. Against trend, I find that while students attending four-year colleges have become more sensitive to distance (81% reduction per 100 miles) than their 1992 counterparts (73%), those attending two-year institutions have become less so (76% vs. 92%).

When separating the enrollment decision into its constitutive parts—where students choose to apply versus where students choose to attend conditional on having applied—I generally find the largest and most significant effects in the application stage. Because these results are estimated on different subsamples of students they should be interpreted with some caution. These results suggest, however, that college choice characteristics may be most salient in the pre-application stage of the enrollment decision. This finding has important implications for current aid policy since it offers further evidence that some students self-select out of colleges when they generally have less information about their actual expected costs than they would after offers of admission or targeted recruitment (Avery, Hoxby, et al. 2006).

Among subpopulations of high ability and low income students, I observe results generally similar to those of the full sample with a few notable exceptions. High SAT students, compared to the full sample, are less sensitive to distance and more sensitive to match. Compared to the full sample, low income students are more sensitive to distance across all models, and, perhaps surprisingly, no more likely to choose two-year institutions than four-year institutions.

In the second set of equations, I find that characteristics of the most-likely college continue to have little or non-significant impact on the decision of whether to enroll when controlling for a student's characteristics. This result holds across most-likely college specifications as well as the nearest public-two year college for both the full sample and the subsample of marginal students. Consistent across all specifications and groups, however, is the positive correlation between county unemployment rate and the decision to attend: approximately 3-7% increase in odds per 1% increase in the unemployment rate.

Given these findings, it may appear that college costs such as price and distance do not matter, on average, in the decision *to* enroll (as opposed to *where* to enroll). Due to the non-random spatial sorting of the population based on demographic characteristics, however, non-significant parameters on college costs in models that control for race/ethnicity, parental education, family income, and local labor markets may represent somewhat artificial conditions and smooth over spatial heterogeneity in college enrollment decision-making processes. I discuss this possibility in more detail in the paper's conclusion.

The rest of the paper is structured as follows: after a brief review of the relevant literature on college choice in section 2, I provide a description of my estimation strategy in section 3 and data in section 4. Results are discussed in section 5 followed by the conclusion in section 6.

2 Literature Review

The literature on the college choice process is varied and large. Perna (2006) notes two major theoretical frameworks for understanding how students (1) make the decision to attend in college and, conditional on this decision, (2) decide where to enroll: one based on a sociological model of status attainment and the other based on an economic model of human capital investment. Sociological models based on social and cultural capital (Coleman 1988; Bourdieu 1977) have been richly applied in the study of college choice. A full description of this literature, however, lies outside of the scope of this paper, which instead frames the enrollment decision as in terms of a human capital investment (Toutkoushian and Paulsen 2016; Becker 2009; Paulsen and Smart

2001).

Much of the college choice literature based in human capital theory has been concerned with the cost of attendance (Deming and Dynarski 2009; Avery and Hoxby 2004; Dynarski 2002; Kane 1996; Kane 1995; Leslie and Brinkman 1987; Manski and Wise 1983; Fuller, Manski, and Wise 1982). Since this literature broadly frames the enrollment decision as one meant to maximize lifetime utility (generally operationalized as wages or earnings), college costs become a key factor (Toutkoushian and Paulsen 2016; Turner 2004; Paulsen and Smart 2001). Within this framework, potential college students are viewed as rational actors who weigh their options—which may include not going to college—and choose the school or employment prospect that represents the biggest wage return for their investment. Changes in the cost of college, therefore, can shift a student's optimal choice. For those on the margins of attendance who face constrained college choice sets, an increase in costs may reduce the estimated returns to a college degree enough to cause them to delay enrollment or forego college altogether. Empirical evidence has borne this theory out, with a number of studies finding that increased college tuition and/or costs are associated with decreases in the likelihood that a potential student will enroll in a given school or any school (Avery and Hoxby 2004; Long 2004; Kane 1995; Leslie and Brinkman 1987; Fuller, Manski, and Wise 1982).

Distance between student and school represents an indirect cost that is often used in econometric models that account for college enrollment. A number of papers seeking to produce causal estimates of various higher education outcomes have used college proximity as an exogenous instrument in place of college enrollment in order to mitigate or remove the selection bias that confounds naive estimates (Doyle and Skinner 2016; Card 1999; Rouse 1995). Bettinger and Long (2009) use each student's most proximal college as an instrument for selection into remedial education courses in a study on the outcomes of such courses. Similarly, Xu and Jaggars (2013) use the distance between students and their campuses to instrument for the likelihood of enrolling in online course section in a paper investigating student persistence and final grade in these courses. Shared by these studies is the belief that distance matters to students when selecting among colleges and course options.

Relatively fewer studies have explicitly investigated the links between college choice and distance. Long (2004) finds that distance mattered greatly for three nationally representative cohorts of students when choosing between colleges, with a 83% to 73% reduction in the odds of attending selecting a particular four-year university (66% – 95% for two-year colleges) per 100 miles. Over the smaller geographic area of Greater Baltimore, Jepsen and Montgomery (2009) find that among older students, the likelihood of attending a community college dropped 2.5% for every additional mile of travel distance. In a simulation in which students had to attend their second rather than first closest institution, an average change of approximately 5 miles, the authors predicted community college enrollment to drop by 19%. Most recently, Hillman and Weichman (2016) have studied the role that “college deserts,” locations with few if any proximal college options, have on college choice, especially for minority student populations. While it may be that distances between a student and nearby college options become less important as options for online education increase (Bowen 2013), recent studies on the efficacy of distance education courses on student outcomes are not promising for a wholesale replacement of face-to-face coursework (Xu and Jaggars 2013; Xu and Jaggars 2011). For the time being, distance matters for student college access and persistence.

Costs, however, are not the only determinants of college choice. Increases in the returns to the education at a particular college may shift the cost curve in the opposite direction and make it a comparatively better option. Students may be willing to spend more money or time to attend a high quality college if the expected return is higher than that of a less costly institution. Selectivity, rejection rate, retention rate, tuition, faculty salary, and student-to-faculty ratio have all been used as proxies for institutional quality in studies that have investigated its association with student outcomes (Long 2010; Black and Smith 2006; Black and Smith 2004; Eide, Brewer, and Ehrenberg 1998). Though the non-random sorting of students into colleges makes it difficult to produce causal estimates, three of the four above cited studies use quasi-experimental designs—propensity score matching or instrumental variables—to show that higher quality institutions may increase the likelihood of enrollment in graduate school as well as post-graduation wages.

Another measure of quality concerns a student's would-be institutional peers. Investigating changes in enrollment between the NLSY79 and NLSY97 cohorts, Lovenheim and Reynolds (2011) find that conditional on ability, changes in income became less important determinants of enrollment. Examining the Texas Top 10% program, Niu and Tienda (2008) find that changes in high school ranking were significantly associated with first college preference across unconstrained and constrained choice sets. Most recently, Hoxby and Avery (2012) find that high-income, high-achieving students were likely to consider academic match when applying to a range of safety, match, and reach schools. This set them apart from their equally high-achieving but low-income peers who did not follow this application pattern and were therefore more likely to "undermatch." Together with other measures, research suggests that many students consider institutional quality when making their enrollment decisions.

My paper represents a contribution to the literature for two reasons. First, I investigate previously described aspects of the college enrollment decision—cost, quality, and match—by analyzing that choice process for a more contemporary generation of high school graduates, many of whom graduated from high school in 2004 and enrolled in 2005 or 2006. In this way I offer an update to the human capital literature on the college enrollment decision. Second, I focus on replicating an earlier study, Long (2004), so that a clear comparison among all cohorts—1972, 1982, 1992, and 2004—may be made and longitudinal trends more accurately identified.

3 Method and estimation strategy

3.1 Modeling a student's choice of where to enroll

The college enrollment decision may be broadly separated into two parts. First, students must decide *whether* they will enroll.[1] Conditional on planning to attend, they must then choose *where* to enroll. It may also be the case for many students that the first step is subsumed in the second, that is, the *whether* only having salience if there appear to be no good options. Under this scenario, the enrollment decision is best understood as a choice between a discrete set of postsecondary options.

The conditional logit model, also known as McFadden's discrete choice model (McFadden 1973), has been used to predict an individual's choice when the option set is discrete and known. Unlike a multinomial logit model, which uses variation across individual attributes to model such choices, the conditional logit model relies on variation across choice attributes (Greene 2012). The decision about which mode of travel to take—walking, bus, subway, taxi—represents a canonical application of this model.

The conditional choice model applies well to the college enrollment decision as students choosing to enroll in college have a discrete and known set of choices. The probability that student i will enroll in college j is

$$Prob(Enroll_{ij}) = \frac{e^{X_{ij}\beta}}{\sum_S e^{X_{ij}\beta}}, \tag{1}$$

where X_{ij} is a vector of choice-specific covariates. If γ is a characteristic of the student and δ is a characteristic of the college, then $\gamma \times \delta$ represents an interactive characteristic of the student-college choice pair. Distance between the student and college would be one such example. Because the conditional logit model fully specifies every choice alternative for an

[1]The college enrollment process contains many steps, which is the reason I fit models on both enrollment and application behavior. For clarity and space, I speak only of enrollment in this subsection and the next. Application to college as an intermediate step in the enrollment process is implied.

individual, only those characteristics of the choice set that vary within the individual strata may be fitted. This means that only college-level attributes (δ) like student-faculty ratio and student-college interactions ($\gamma \times \delta$) like distance may be parameterized; invariant student-level covariates (γ) such as gender and race are differenced out of the equation.

For estimated parameters to be consistent, the conditional logit model must meet the independence of irrelevant alternatives (IIA) assumption. Strictly, the IIA requires that the odds of choosing *A* over *B* when other options are available is the same as when *A* and *B* are the only two options (McFadden 1973). Functionally, the IIA assumption means that all relevant alternative choices are known and included in the estimated model. While the alternative choice set may be a subset of all possible choices, the preferred choice should not change conditional on the inclusion of new options. Just as the transportation example assumes that the odds of a traveler choosing the bus over walking will remain the same regardless of whether riding a bicycle is an option, the college choice model assumes that a preference for Big State University over Small Private College will remain the same even when Mid-size Regional Tech is an option.

The IIA assumption is strong. While a number of methods for testing the IIA have been proposed, they remain problematic (Cheng and Long 2007). Should the inclusion of a new college in the alternative choice set affect the comparative odds of other binary choice pairs, then my results may be inconsistent. Against this threat, I am supported by two facets of my design. First, I include all Title IV postsecondary institutions with a physical location in 2004 in the choice set for each student. All students are modeled as having the same 3,406 options no matter how unlikely particular student-college pairs may be. I do not include institutions with only an online presence as they do not have physical location with which to measure distance to the student (nor does such a measure make sense). Few students completed their degree programs solely through distance education courses during the 2000s (Snyder, Brey, and Dillow 2016). Such that students viewed virtual institutions as viable higher education options or only took online courses, thereby weakening the link between distance and application/enrollment, their omission may bias my results.

Second, the college application process generally supports the IIA assumption. Because students apply to colleges independently of one another, the availability of college C should not affect the relative odds of applying to college A over college B. Similarly, acceptance to college C should not affect the odds of enrolling in college A over college B, assuming acceptance to all. Again, by design of the college enrollment process, it is unlikely that many undergraduate applicants attempt to use an admissions offer from a third school to leverage a better admissions package in a contest between two other schools. On the other side of the table, schools do not (and should not due to legal restrictions) collaborate when deciding offers of admission or aid. Such that students are able to signal their intentions to schools that then act on them when deciding on offers, the IIA is rejected and my results may be biased.

A closely related threat to IIA during the application process lies in alternatives that may be considered "close substitutes," McFadden 1973, p. 113. For my study, I am aided by the fact that colleges are distinct for potential students. Though two colleges may have similar attributes and appear virtually identical in the context of all postsecondary institutions, they will in fact represent distinct options due to their unique faculty composition and cultures, even in the unlikely case that all other attributes were identical.

3.2 Modeling a student's choice of whether to enroll

Though the conditional logit model of the choice *between* colleges requires variation within student strata and therefore cannot account for invariant student-specific characteristics such as gender, race/ethnicity, and socioeconomic status, these individual attributes are important in the enrollment decision. For the simple binary outcome of *whether* a student enrolls, I estimate a logistic regression that not only returns these student-level covariates but also attributes of the most-likely college, C_{most}.

In this model in which all sample students are included, both those who did and those who did not attend a postsecondary institution within two years of earning a high school or general equivalency diploma, C_{most} is the college where

$$Pr(Enroll_{ik}) \geq Pr(Enroll_{ij}) \forall j \neq k, \tag{2}$$

that is, where the probability of enrollment is the highest as compared to other options. These probabilities are estimated using the fitted parameters from conditional logit models for three outcomes: enrollment, application, and enrollment conditional on application. For attenders, C_{most} may not be the college actually attended. For students not in the first set of equations, non-attenders, this is the most-likely college as predicted based on their own student-college choice characteristics, which can be constructed despite the fact that they did not attend.

Following Long (2004), I also estimate the logistic regression equation using the characteristics of the nearest public two-year college, C_{near}. Because many studies use a student's nearest public two-year as an instrument for college choice (*e.g.,* Card 1999; Rouse 1995), a comparison between models using this college and the most-likely college is warranted. For all students C_{near} is simply the nearest public two-year college.

4 Data

Student-level data come from all waves of the ELS survey (National Center for Education Statistics 2016a). Administered by the National Center for Education Statistics, ELS has followed a nationally representative cohort of students who were high school sophomores in the 2001-2002 school year. With the administration of the first, second, and third follow-ups in 2004, 2006, and 2012, respectively, the original sample of students ($N \approx 15,000$) was tracked from high school graduation through early adulthood. Information about students' college enrollment was taken from the follow-up surveys. Student demographic information such as gender, race/ethnicity, parental education level, and family income were gathered from the base-year survey public-use files.

Other student-level covariates, including SAT or SAT-equivalent composite score and county of residence in the base year of the survey were taken from the ELS restricted-use data file. Student test scores were converted to percentiles so as to standardize the measure of student ability and facilitate interpretation of the interaction between a student's score and the median institutional score. Students without SAT or SAT-equivalent scores were dropped from the sample. Descriptive statistics of student characteristics, for the full analytic sample as well as the high SAT ($>$ 1100) and low income ($>$ \$25,000 per year), are shown in Table 1.

ELS restricted-use data also contains the unique institutional ID of all colleges to which students enrolled as well as those to which they applied and were accepted. These unique IDs were used to link the college of choice to its characteristics as found in IPEDS. Each student's base-year census block group of residence and institution's coordinates were used to compute the distance between each student-college pair. In all cases, the geographic center of the census block group as provided by the United States Census Bureau was used as the student's location. Census block groups generally contain 600 to 3,000 individuals and are contained within census tracts, meaning they represent one of the smallest geographic areas provided by the Census.[2] Though the measure of distance between each student-college pair is necessarily coarser than it would be were exact student addresses utilized, it represents an improvement on measures which rely entirely on zip codes, which can be problematic location proxies in spatial analyses due to their inconsistent shapes and centroid locations (Grubesic 2008). IPEDS did not report geographic coordinates of institutions in 2004, the senior year of ELS cohort and the one utilized in these analyses. Institutional coordinates were back-filled in an iterative process that first matched schools with the values given in later IPEDS survey years provided that the zip codes remained the same. If not matched or zip codes differed, the institution's mailing address was geocoded. For the very few instances where both steps failed to produce geographic coordinates, the centroid of the institution's zip code was used.

Most other college characteristics were taken directly from the 2004 administration of the IPEDS survey (National Center for Education Statistics 2016b). These include cost of attendance,

[2] https://www.census.gov/geo/reference/gtc/gtc_bg.html

both for in-state and out-of-state students, median SAT score of the student body, student faculty ratio, the percentage of tenured and tenure-track faculty, and full-time equivalent student enrollment. Table 2 offers descriptive statistics of these covariates for the sample of institutions used in the analyses.

Cost of attendance for each school was computed for both in-state and out-of-state students. In-state cost is the average in-state tuition and fees less average federal and state grants. Out-of-state cost is the average out-of-state tuition and fees less average federal grants. For each student-school choice, the student's state of residence was used to assign expected cost.

Median student body SAT scores were approximated by first combining each institution's math and verbal scores, reported at the 25th and 75th percentiles, and then taking the average. For institutions that reported ACT rather than SAT percentiles, the composite ACT percentiles were first converted to SAT equivalent scores and then averaged as before. Many institutions do not report score percentiles, either because they choose not to do so or because they follow open admissions policies. For these institutions, I first attempted to associate their respective *Barron's* competitiveness index measures with the mean of the SAT score range associated with that value. For institutions without a *Barron's* measure, which includes all public two-year colleges as well as those rated as non-competitive, I followed Long (2004) and assigned an SAT score value of 700. As with student SAT scores, median institutional scores were transformed into percentiles.

Representing a change from Long's specification, I replace the number of faculty holding a PhD with the percentage of tenured and tenure-track faculty as a measure of institutional quality. I do so for two reasons. First, the measure of faculty with a PhD is not readily available in current datasets. Second, studies of student application behavior (Drewes and Michael 2006) and survey responses (Umbach 2007) suggest that the more salient institutional quality measure for students may instead be the proportion faculty who are either tenured or on the tenure-track.

Instructional expenditures per full-time equivalent enrollment were computed using the measure of instructional expenditures reported in the Delta Cost Project Database (The Delta Cost Project 2016), which provides useful institutional finance measures not easily computed using raw IPEDS variables, and the IPEDS-reported measure of FTE enrollment.[3] Averages for these values across school types in the analytic sample are also shown in Table 2.

Measures of county-level unemployment rates that are used in the logistic regression on the decision to enroll come from the Bureau of Labor Statistics. As with the institutional data, these measures are for 2004, the senior year for the ELS cohort. Though perhaps of less importance among a cohort which saw approximately 80% of its members enroll in college, conditions in the local labor can change a student's baseline earnings potential, thereby adjusting the expected return to a college degree. By convention and theory, therefore, this covariate is included.

5 Results

5.1 Choice between colleges

5.1.1 Attendance

Results of the conditional logit model of student choice for members of the ELS cohort who attended college within two years of graduating from high school are shown in Table 3.[4] Because each postsecondary institution was considered an option for all sampled students, the design matrix involved over 23 million unique student-school choice combinations (approximately 6,780 students $\times$ 3,409 college options). Each choice pair was fit using the following predictors: cost, cost squared, distance in miles between student and school, distance squared, instructional expenditures per full-time equivalent student, both at level and squared, institutional student/faculty ratio, the percentage of tenured or tenure-track faculty, the difference between the student's SAT percentile and the school's student body SAT percentile, split by the higher of the

[3] Approximately 12% of institutions were missing information on instructional expenditures in 2004, meaning that instructional expenditures per FTE student could not be computed. Due to the computational demands of fitting the conditional logit model to large data set, a multiple imputation procedure was not feasible. Rather than drop these institutions from the option set (and bias the results) I employed Buck's method of conditional mean imputation (Little 1992; Buck 1960). The results of the conditional logit models obtained with these imputed values are qualitatively similar to those obtained in earlier specifications when the unadjusted fitted values were used.

[4] Conditional logit equations were estimated using the `asclogit` command in the Stata 14 statistical package (StataCorp 2015). Analysis datasets, predicted probabilities, and logistic regression equations were all generated or fit using the R statistical package (R Core Team 2016).

two values, an indicator variable for two-year institutions, and various interactions with the two-year indicator. Covariates also included but not shown are distance cubed, SAT variables squared, and full-time equivalent enrollment, both at level and squared. Odds ratios and z-scores are reported. To facilitate comparison between this and prior cohorts, results taken from Long (2004) are reprinted in the first six columns of Table 3. The last two columns of the table, which are in bold, show the new results for the class of 2004.

How has the college enrollment decision changed? Turning first to college costs I find that expected costs (tuition and fees less grants) have continued to decline in relevance for between-college choice among four-year institutions. Every $1,000 difference in cost between two colleges suggests an approximately 16% reduction in the relative odds of enrolling in the more expensive institution. Compared to the 53%, 42%, and 35% reductions found for the 1972, 1982, and 1992 cohorts, respectively, these results suggest that while the 2004 cohort still prefer lower costs at four-year institutions, they continue the trend of becoming less sensitive to cost over time. Considering two-year schools, however, the latest cohort have become a little more sensitive to cost than their 1992 counterparts: 21% vs. 15% reduction in the relative odds. This finding may be driven, at least in part, by the comparative increase in the number of private two-year options, especially for-profit two-year colleges with higher average costs, over public two-year options from the 1990s to 2000s.

Considering distance, I find that the 2004 cohort are split. While they are more sensitive to distance to four-year schools than their 1992 counterparts, they are less sensitive to two-year distance. For every 100 miles, the recent cohort shows around an 82% reduction in the odds of choosing a four-year institution and a 77% reduction for a two-year. These values are flipped compared to the 1992 cohort, who showed 73% and 92% reductions for the same levels, and more akin to the 81% and 66% shown by the 1982 cohort. Once more the change may be due in part to an increased number of for-profit universities. With their comparatively greater use of online coursework (even when a physical campus exists), these institutions may relax the role that distance plays in the choice between two-years. This explanation cannot explain, however, the similarity between the 2004 and 1982 cohort. The unexpected flip between two- and four-year institutional choice (one would expect students to be more sensitive to the distance to a two-year rather than four-year) may be an artifact of the sample. Regardless, distance on the whole remains a more salient consideration than cost for the 2004 cohort when choosing between schools.

Concerning instructional quality, I find no significant effect of instructional expenditures per full-time equivalent student on college choice for four-year universities, but an 18% increase per $1,000 for two-year schools. Compared to statistically significant relative odds increases of 10% and 50% at four- and two-year schools for the 1992 cohort, my findings suggest a weakening of the relationship. The class of 2004 shows a small, but significant positive preference for an increase in the student-to-faculty ratio, with an approximate 2% increase in relative odds per 10 to 1 increase in the number of students to faculty (controlling for full-time equivalent enrollments). Compared to past cohorts, I find a much stronger positive relationship for increases in the percentage of tenured and tenure-track faculty. Estimated jointly for both institutional levels, I find a 7% increase in odds of choosing a school per 10% increase in the proportion of these faculty. While relatively small, this estimate is an order of magnitude larger than past estimates on similarly scaled increases in percentage of faculty with a PhD. Because my operationalization differs from that of Long (2004), however, I cannot argue that students are increasingly swayed by faculty composition. This measure may simply be more salient to students, either directly or as a proxy for other desirable school characteristics, than those of the past.

For student-college match, I find that compared to all prior cohorts, 2004 cohort members are both less likely to prefer the college which has a median SAT score below theirs and more likely to choose the college with a comparatively higher median SAT score. For every 10% point increase in the student's SAT percentile over the student body median percentile, the odds a student will attend are reduced 36%. In the opposite direction, every 10% deficit in the student's SAT percentile relative to that of the school's student body median increases the odds of attendance by 54%. Differing from the 1982 cohort, who appeared to prefer "match" schools, all else being equal, the 2004 cohort have followed their 1992 peers in preferring "reach" schools and increasingly eschewing "safety" schools.

Finally, I find that the odds of choosing a two-year institution remained high and increasingly significant for the 2004 cohort. Students were 3.6 times as likely to choose a two-year college as a four-year institution. This finding is lower, but qualitatively similar to the 4.9 times increase estimated for the 1992 cohort. While positively signed, the 1972 and 1982 parameter estimates were not nearly as large and both were non-significant, suggesting no more relative preference for two-years over four-years. This finding is likely due in large part to the increased number of two-year institutions, especially in the for-profit sector, along with their accompanying growth in enrollments over the past few decades.

5.1.2 Application and attendance conditional on application

Due to the richness of the ELS data set, the choice between colleges for the 2004 cohort may be separated into its constitutive parts: application and enrollment conditional upon application. Table 4 presents results from these specifications in the middle and last pair of columns, respectively. For comparison, the first two columns repeat the results for unconditional attendance that were discussed in the prior section.

For the application conditional logit model, students were once again assumed to have the same full set of college choice options ($N = 4,209$), with the dependent variable this time set to one for all schools to which a student reported applying. As with the first model, only those students who reported at least one positive outcome were included in the estimation sample.[5] Unlike in the first model, students in the application model could have multiple positive outcomes, that is, apply to more than one school. In the last choice model, attendance conditional

[5] Application outcomes are reported in the second wave of the ELS survey. Final attendance outcomes are reported in the third wave. Because not all ELS students reported applications or began after the second wave collection (but within the two-year enrollment window), fewer students are included in the application model than in the unconditional attendance. Though applications could be logically imputed for those in the first model who attend but are missing application information, I have chosen only to include those students with complete application data in the second and subsequently third models.

on application, attendance was once again the outcome, but only those schools to which the student applied were included in the choice set. Because the conditional logit model requires variation in choice within each student strata, only those students who applied to more than one school were included in the third model sample.

Comparing the three model specifications, parameter estimates between the unconditional attendance and application models are remarkably similar in direction, size, and significance. When deciding upon which schools to apply, the 2004 cohort, as suggested by the first model, appears more amenable to schools that are less expensive, geographically closer, have higher instructional expenditures, more tenured and tenure track faculty, and have student bodies with as good as but preferably higher SAT scores. They also prefer, all else equal, two-year over four-year colleges. Turning to the third model, however, only distance to four-year colleges remains both statistically and practically significant. While distance retains its negative effect, it is much reduced. Conditional on having applied, students show only a 22% reduction in odds of attendance per 100 miles compared to 82% and 73% reductions shown in the first and second models, respectively.

These new results suggest that the determinants of the separated student enrollment decision, insofar as they concern the choice between college options, may better describe the application decision than the attendance decision. Two caveats apply. First, some students in the full sample apply to only one college. Such that the results from the second model are driven by collinearity between application and unconditional attendance outcomes, that is, application and attendance are effectively the same, parameter estimates should be similar. Second, the last model is fit to a subset of students who applied to more than one institution and therefore may represent a distinct and dissimilar subset of students. Insofar as their application and enrollment behavior is different from students who apply to only one institution, a comparison between the models may not be appropriate. As shown in Table 1, however, most students who attended college within two years applied to more than one school (mean ≈ 2.81). That the subsample included in the third model also represents a majority of the second model sample ($\approx 70\%$) further supports a comparison.

Taken together, the separated results suggest that the 2004 cohort's enrollment decisions may have been primarily a function of their self-directed application behavior. For many students, application and attendance were the same as they only applied to the school they subsequently attended. For others, the choice characteristics that drove their multiple application decisions—cost, distance, faculty composition, match—were less salient when choosing where to attend from within their self-selected options.[6] These results align with other research which finds that students may be making suboptimal decisions due to incomplete information or support (Avery, Hoxby, et al. 2006). Though my findings do not speak to the outcomes of each student's choice, they do provide evidence that much of the weight of the college enrollment decision may lie in the application stage when students do not generally have as much information regarding their actual college costs as they do once they have applied and received a personalized offer.

5.1.3 Subgroups: high SAT and low income

To explore possible differences in choice preference for subsets of the student population, I estimate the same three conditional logit models—unconditional attendance, application, and attendance conditional on application—for two different groups: high SAT students and low income students.

Results for students with a composite SAT score above 1100 are shown in Table 5. Across the models, high SAT students show generally similar preferences as the full sample. And like the full sample, their preferences appear most significant in the application stage, with estimated parameters most congruent between the first two models and losing significance in the third. There are, however, a few exceptions. College costs at two-year colleges show no significant change in the relative odds of selection in any model for this group. Distance also matters slightly less or loses statistical significance for high SAT students. In the first two models, the reduction in odds per 100 miles to four-year college is 65-75% percent. While still significantly negative,

[6] As a test on robustness, I also fit a conditional logit choice model on attendance conditional on acceptance. Most of the multiple applicants ($\sim$ 85%) were accepted to multiple schools and could be included in the model. Results were qualitatively similar to those for attendance conditional on application, and are available upon request.

the reduction is approximately 8-9% points less than that estimated for the full sample. Other distance parameters on two-year institutions and in the third model are only marginally significant.

The biggest difference between the high SAT subgroup and the full sample, however, lies in student college match. Compared to the full sample, high SAT students show a much greater preference for colleges at which the student body median SAT is higher ("reach") schools. For every 10% point difference, these students prefer such schools four to one in the unconditional attendance model. In the application model, the positive change in odds is 88% compared to 31% for the full sample. Taken together with the estimates I find on distance, high SAT students on average appear willing to travel a little farther if it means attending a school with a higher achieving student body.

Low income students, on the other hand, are not as sensitive to student college match. Results from their models are shown in Table 6. Categorized as students whose families reported earning less than $25,000 in the year prior to the ELS base year survey, low income students only show a significant negative preference for schools with lower median SAT scores in the application model—35% reduction per 10% point difference—which is qualitatively similar to that estimated for the full sample. Unlike the full sample and high SAT subgroup, low income students do not show significant preference for schools with higher SAT scores in any of the models. Across all models, low income students are more sensitive to distance to four-year colleges than the full sample. The trend reverses for two-year colleges, with distance having a less negative effect on the relative odds. While caution should be taken in interpreting these results due to the reduced sample size, it may be that scheduling flexibility often found at two-year colleges appeals to those low income students who have greater need to work, thereby increasing their willingness to apply to and attend them even if they are a little further away from home.

5.2 Choice whether to attend college

Having discussed the determinants of the between college choice for students, I turn now to logistic models on the decision of whether to enroll in college. For each student, the decision of whether to enroll was modeled as the function of both personal characteristics and characteristics of the most-likely college as predicted by the conditional logit model. Separate logit models were estimated for each most-likely college predicted by the three conditional logit models. A fourth model that used the characteristics of the nearest public two-year college was also computed. Across all specifications, the right-hand side covariates included: college cost for the student (level and squared), distance between the college and student (level, squared, and cubed), instructional expenditures per full-time equivalent student (level and squared), college full-time equivalent enrollment (level and squared), student SAT percentile, ordered category of student family income, ordered category of the student's parental education level, indicators for the student's gender and race/ethnicity, and home county-level unemployment rate in 2004. Results for these equations are presented in Table 7. The first four columns present estimates for the full analytic sample. The last four columns show estimates for a subset of students at the margins of attendance: those with family incomes below $25,000 or composite SAT scores below 900.

Across all full sample models, characteristics of the most-likely or nearest public two-year college have little predictive power concerning a student's likelihood of enrolling in college within two years of graduating from college. Distance, which shows a consistently negative association with enrollment in all four models, is only significant when using the most-likely college predicted by the unconditional enrollment choice model or the nearest public two-year college. Instructional expenditures are significant in models using the application and attendance conditional on application most-likely colleges, but oppositely signed. Other school characteristic parameters are not statistically significant. Among the subsample of students at the margins, only the parameter on instructional expenditures in the application-derived most-likely school meets conventional statistical significance.

By comparison, student characteristics are more predictive of the likelihood of enrollment.

As might be expected, students across all models in both groups are 36 to 46% more likely to enroll for every 10th percentile increase in their composite SAT scores. Similarly, enrollment likelihood is positively associated with family income and parental education. Aligning with recent trends that have seen increases in the relative proportions of women enrolling in college, female high school graduates are nearly 60% more likely than their male counterparts to enroll. Because of comparatively smaller sample sizes, the models generally lack the power required to clearly differentiate between racial/ethnic subgroups. That said, it appears that controlling for other characteristics, both black students in the full and marginal sample and Asian / Hawaiian / Pacific Islander students at the margins may be more likely to enroll, whereas students who identify as multiracial are less likely to enroll.

County-level unemployment rates are also predictive of enrollment at significant levels in six out of eight models; the remaining two parameters are of similar strength and direction and marginally significant. Even when controlling for student ability and socioeconomic status, a one percentage point increase in the unemployment rate is associated with a 3 to 4% increase in the odds of enrollment among the full sample. Among the subsample of students on the margins, the increase is 6 to 7%. These consistently significant results provide further evidence for a negative association between local labor market conditions and enrollment (Betts and McFarland 1995; Manski and Wise 1983). Even in a college-for-all era, members in the 2004 cohort appear to have considered immediate employment to be a viable alternative.

The similarity between my largely non-significant school parameters and those Long (2004) reports for the 1992 cohort suggests that the decision of whether to enroll has not significantly changed since that time. Conditional on student characteristics and local employment opportunities, characteristics of the most-likely and nearest two-year public college appear largely irrelevant for the 2004 cohort. Why is this the case? A few possibilities exist.

First, non-significance could be the result of misspecification. The most-likely colleges as predicted by the conditional logit choice models may not actually be that likely. Characteristics of that school, under this scenario, should not be significantly predictive of the decision to enroll.

As a check on this scenario, I performed a number of sensitivity analyses not reported in this paper. First, I considered how often the most-likely predictions matched the college actually chosen. In the unconditional attendance model, students who enrolled within two-years of high school graduation attended their most-likely college 18% of the time. Students applied to their most-likely college as predicted in the application model 33% of the time. For attendance conditional on application, however, the number is much lower: only 1% of students attended the school predicted by this model.

Next, for those students who did not attend or apply to the predicted most-likely school, I considered the difference in rank between that school and the ones they chose as well as the difference in predicted probability between the two. In the unconditional attendance model, the median rank of the school actually attended was 13, meaning the model predicted it as the thirteenth-most-likely school. The median difference in predicted probably between it and the most-likely school was $\approx 10\%$. For the application and attendance conditional on application models, these values were [6, $\approx 12\%$][7] and [966, $\approx 0.1\%$], respectively.

Finally, I considered how closely the most-likely schools matched the actual schools on a number of institutional characteristics. In the unconditional attendance model, the schools that students chose to attend were in the same state as the most-likely school 72% of the time. The actual and predicted school matched on level (two- vs. four-year), control (public, non-profit private, for-profit private), and sector (level and control jointly), 70%, 68%, and 44% of the time, respectively. In the application model, the match was as follows: state, 73%; level, 73%; control, 69%; and sector, 44%. For attendance conditional on application: state, 8%; level, 24%; control, 66%; and control, 21%.

Taken together these sensitivity analyses offer evidence that predictions from the conditional logit models fit the data well. Insofar as the selected schools have characteristics that match those of the predicted most-likely schools, then the mismatch between the two may be, in part, attributable to unobservable characteristics of the student-school choice. If predicted schools are

[7] These numbers consider the difference between the most-likely application school and the actual application school that was closest in the predicted rankings.

observationally similar to those actually chosen, then non-significant parameter estimates such as I find suggest that cost and distance, which have strong negative associations with the choice between colleges, may not similarly matter when choosing whether to enroll in college.

6 Conclusion

This paper replicates an earlier study on the determinants of the college enrollment decision with a new cohort of students. So that my findings for 2004 cohort may be placed in clear comparison with those for the 1972, 1982, and 1992 cohorts, I have followed the methods and model specifications of Long (2004), who investigates those cohorts, as closely as possible. Considering the between-college choice for the full sample, I find that the latest cohort of students remains sensitive to cost and distance. Though their cost sensitivity generally continues to abate, following the trend established by earlier cohorts, general sensitivity to distance among these students remains high. I further find that college match increasingly matters. While students are less likely to choose schools with student body SAT scores below their own, they prefer schools with higher average scores. Finally I find that students are moderately sensitive to the proportion of tenured and tenure-track faculty and continue to prefer, all else being equal, two-year institutions. Analyses for high-SAT and low-income subgroups follow similar patterns, though with small differences. Whereas high-SAT students are less sensitive to distance and more willing to "overmatch," low-income students are both more sensitive to four-year distance and less sensitive to two-year distance than the full sample.

As an extension on prior work, I separate the choice between colleges into the application and enrollment conditional on application stages. Where I find strong similarities between the first, unconditional attendance model and the application model, parameters tend to drop to non-significance in the conditional attendance model. Though changes in the sample require that comparisons be made with caution, I argue that cross-model comparisons provide evidence that the weight of the student enrollment decision may lie more in application stage than after offers of admission are made. Some students, rather than apply and face potential rejection, instead may be self-selecting out of colleges that would admit them at a lower-than-sticker cost and at which they would conceivably succeed.

Such self-selection (or self-non-selection, as the case may be) has important ramifications for both institutional and governmental policies meant to get students to apply and enroll. Institutions, especially those with high prestige and selective admissions, may be missing talented applicants (Avery, Hoxby, et al. 2006). State and federal aid policies may be overly complex and/or too ill-timed to give students accurate information about their actual expected cost (Dynarski and Scott-Clayton 2013). In either scenario, the results of my conditional logit models of between college choice suggest that policies could be improved so that students make more informed application decisions.

When deciding whether to attend college, my results suggest that the 2004 cohort students are like their 1992 counterparts in that the decision to enroll is largely independent of characteristics of their most-likely or nearest public two-year college. While cost and distance may matter when choosing between colleges, they do not appear to moderate the likelihood of enrollment versus non-enrollment. Instead, local labor market conditions and student characters that were not included in the conditional logit models such as gender, race and ethnicity, parental education and income, and academic ability are much more predictive. In the decision of whether to enroll, members of the 2004 cohort appeared more affected by their own characteristics than those of any model-predicted most-likely school.

A few important caveats to this interpretation apply. First, I do not include students who did not report SAT or SAT-equivalent test scores. My results, therefore, speak to those students who signal that their college choice sets are neither zero nor limited to their local open access postsecondary institutions. Such that students who do not take the SAT or equivalent exams are differentially affected by college costs and distance should they decide to enroll, my results may be less generalizable to the full population of high school graduates. Second, my results speak only to enrollment, not persistence or attainment. Though I find that, on the margins, cost and distance do not affect the decision to enroll, it may be the case that these aspects of college-going become more important as students work towards a degree. That only 42% of the 2004 cohort had obtained an associate's degree or higher (33% bachelor's or higher) by the third follow-up, despite nearly 80% having at least attempted college, suggests that persistence and attainment may be more important outcomes to model *vis à vis* cost and distance than first enrollment in future studies.

Finally, it may be the case that the inclusion of student and family characteristics in the final logistic models produces parameters on school characteristics that do not reflect non-random spatial sorting of student populations across the country. In the face of persistent racial and economic segregation (Souza Briggs and Wilson 2006), what does it mean that, controlling for socioeconomic characteristics, the characteristics of the most-likely college, specifically its distance from the student, are not statistically relevant in the decision of whether to enroll in college?

Ideally, I would explore this question by observing the college enrollment decision of representative students in locations across the entire country (not just those reported in my sample). In this thought experiment, students could be different from each other in all ways—gender, racial/ethnic identification, parental income—except they all would have identical levels of academic preparation. If these students demonstrated differential rates of college enrollment, I could compute the average distance to the nearest college for those who did not enroll and compare it to the average distance of those who did. A significant difference between the two values would lend support to the hypothesis that distance from college might remain an important component of the choice of whether to enroll in college.

Though I do not have access to such a population of students, I can simulate one and, based on their propensities of enrollment, compare average student-to-nearest-college distances. To do this I first use U.S. Census data to create a population of synthetic students, one representing each census tract in the lower 48 states and generate predicted probabilities of enrollment across the country. For this simulation, each synthetic student represents a simple amalgamation of his or her census tract's population characteristics. Because Census data do not give SAT scores, I generate four "modal" students for each tract, each with a different SAT percentile—30th, 50th, 70th, or 90th—representing different levels of college readiness. As with the ELS cohort students, I first predict each synthetic student's most-likely college to attend and then, using the characteristics of this college as well as those of the student, the likelihood that he or she will enroll in college. I generate four sets of predictions in total, one for each SAT percentile.

Figure 1 shows the heterogeneity in these predictions across the distribution of SAT score percentiles. Each map, one for each level of SAT percentile, is coded blue for predicted probabilities above than 0.5—more likely than not to attend within two-years of earning a high school diploma—and red for those below. Darker shades show predictions that are statistically significant ($p < 0.05$). The maps in Figure 1 show variation in the likelihood of enrollment, despite the fact that all synthetic students within each simulation are equally college ready. My models suggest that even for synthetic students with SAT scores in the 90th percentile (bottom right map in Figure 1), that is, those most likely to benefit from attending college and earning a degree, the odds of enrollment change depending on location.

When considering the population characteristics of only those census tracts with predictions that are statistically significant, high or low, I find statistically significant differences between the bottom and top quartiles of predicted probability. Tracts in the bottom quartile (those with the lowest predicted probabilities of enrollment) tend to have relatively more persons of color, lower average educational attainment, and lower median income than those in the top quartile. Computing the distance to the nearest public, two-year, and public two-year institution for each census tract centroid, I find that tracts that produced non-enrollees in the simulations tend to be farther removed from the nearest institution. Figure 2 shows the differences in median distance to nearest college between tracts with non-enrollees and those with enrollees. Looking at the upper right-hand facet in which synthetic students were given the 50th percentile of SAT scores, the median distance to the nearest public two-year institution for non-enrollee tracts is around 12 miles. For enrollee tracts, the median distance is less than five miles. Results are qualitatively similar across the range of SAT percentiles.

Bringing all evidence to bear, a rejection of the importance of college opportunity on the likelihood of a student's attendance based on a model that includes student characteristics may be unwarranted if measures of college opportunity are highly collinear with those characteristics. The results produced by the simulation suggest this may be the case. With the increased attention given to students' "geography of opportunity" in recent years (Hillman and Weichman 2016; Tate 2008), continuing to disentangle the endogenous relationship between place and demographic characteristics of student populations remains a pressing task for research concerned with the spatial facets of the college enrollment decision. As data on newer cohorts becomes available, the determinants of their college enrollment decision should be similarly investigated so that policies started in the middle of the last century may be adjusted to meet the needs of students in this one.

References

Avery, Christopher and Caroline M. Hoxby (2004). "Do and Should Financial Aid Packages Affect Students' College Choices?" In: *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It.* Ed. by Caroline M. Hoxby. University of Chicago Press, pp. 239–299.

Avery, Christopher, Caroline M. Hoxby, et al. (2006). *Cost Should Be No Barrier: An Evaluation of the First Year of Harvard's Financial Aid Initiative*. Working Paper 12029. National Bureau of Economic Research. URL: `http://www.nber.org/papers/w12029`.

Becker, Gary S. (2009). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. 3rd. University of Chicago Press.

Bettinger, Eric P. and Bridget T. Long (2009). "Addressing the Needs of Underprepared Students in Higher Education: Does College Remediation Work?" English. In: *Journal of Human Resources* 44.3, pp. 736–771.

Betts, Julian R. and Laurel L. McFarland (1995). “Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments.” In: *The Journal of Human Resources* 30.4, pp. 741–765.

Black, Dan A and Jeffrey A Smith (2004). “How robust is the evidence on the effects of college quality? Evidence from matching.” In: *Journal of Econometrics* 121.1, pp. 99–124.

Black, Dan A and Jeffrey A Smith (2006). “Estimating the returns to college quality with multiple proxies for quality.” In: *Journal of labor Economics* 24.3, pp. 701–728.

Bourdieu, Pierre (1977). “Cultural Reproduction and Social Reproduction.” In: *Power and Ideology*. Ed. by J. Karabel & A.H. Halsey. Oxford University Press, p. 485.

Bowen, William G (2013). *Higher Education in the Digital Age*. Princeton, NJ: Princeton University Press.

Buck, S. F. (1960). “A Method of Estimation of Missing Values in Multivariate Data Suitable for use with an Electronic Computer.” In: *Journal of the Royal Statistical Society, Series B (Methodological)* 22.2, pp. 302–306. URL: `http://www.jstor.org/stable/2984099`.

Card, David (1999). “The causal effect of education on earnings.” In: *Handbook of Labor Economics* 3, pp. 1801–1863.

Cheng, Simon and J Scott Long (2007). “Testing for IIA in the multinomial logit model.” In: *Sociological Methods & Research* 35.4, pp. 583–600.

Coleman, James S. (1988). “Social Capital in the Creation of Human Capital.” In: *American Journal of Sociology*.

Deming, David and Susan Dynarski (2009). *Into college, out of poverty? Policies to increase the postsecondary attainment of the poor*. Working Paper 15387. Cambridge, MA: National Bureau of Economic Research. URL: `http://www.nber.org/papers/w15387`.

Doyle, William R. and Benjamin T. Skinner (2016). “Estimating the education-earnings equation using geographic variation.” In: *Economics of Education Review*.

Drewes, Torben and Christopher Michael (2006). "How Do Students Choose a University?: An Analysis of Applications to Universities in Ontario, Canada." In: *Research in Higher Education* 47.7, pp. 781–800.

Dynarski, Susan (2002). "The behavioral and distributional implications of aid for college." In: *American Economic Review* 92.2, pp. 279–285.

Dynarski, Susan and Judith Scott-Clayton (2013). *Financial Aid Policy: Lessons from Research*. Working Paper 18710. National Bureau of Economic Research.

Eide, Eric, Dominic J Brewer, and Ronald G Ehrenberg (1998). "Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance." In: *Economics of Education Review* 17.4, pp. 371–376.

Fuller, Winship C., Charles F. Manski, and David A. Wise (1982). "New Evidence on the Economic Determinants of Postsecondary Schooling Choices." In: *The Journal of Human Resources* 17.4, pp. 477–498. DOI: 10.2307/145612.

Greene, William H. (2012). *Econometric Analysis*. 7th. Boston, MA: Prentice Hall.

Grubesic, Tony H. (2008). "Zip codes and spatial analysis: Problems and prospects." In: *Socio-Economic Planning Sciences* 42.2, pp. 129–149. ISSN: 0038-0121.

Hillman, Nicholas and Taylor Weichman (2016). *Education Deserts: The Continued Significance of "Place" in the Twenty-First Century*. Viewpoints: Voices from the Field. Washington, D.C.: American Council on Education.

Hoxby, Caroline M. and Christopher Avery (2012). *The Missing "One-offs": the Hidden Supply of High-Achieving, Low Income Students*. Working Paper 18586. Cambridge, MA: National Bureau of Economic Research. URL: http://www.nber.org/papers/w18586.

Jepsen, Christopher and Mark Montgomery (2009). "Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college." In: *Journal of Urban Economics* 65.1, pp. 64–73.

Kane, Thomas J. (1995). *Rising Public College Tuition and College Entry: How Well Do Public Subsidies Promote Access to College?* Working Paper 5164. Cambridge, MA: National Bureau of Economic Research.

Kane, Thomas J. (1996). "College Cost, Borrowing Constraints and the Timing of College Entry." In: *Eastern Economic Journal* 22.2, pp. 181–194. URL: http://www.jstor.org/stable/40325703.

Leslie, Larry L. and Paul T. Brinkman (1987). "Student Price Response in Higher Education: The Student Demand Studies." In: *Journal of Higher Education* 58.2, pp. 181–204.

Little, Roderick J. A. (1992). "Regression With Missing X's: A Review." In: *Journal of the American Statistical Association* 87.420, pp. 1227–1237. URL: http://www.jstor.org/stable/2290664.

Long, Bridget T. (2004). "How have college decisions changed over time? An application of the conditional logistic choice model." In: *Journal of Econometrics* 121, pp. 271–296.

Long, Mark C. (2010). "Changes in the returns to education and college quality." In: *Economics of Education Review* 29.3, pp. 338–347.

Lovenheim, Michael F. and C. Lockwood Reynolds (2011). "Changes in Postsecondary Choices by Ability and Income: Evidence from the National Longitudinal Surveys of Youth." In: *Journal of Human Capital* 5.1, pp. 70–109. URL: http://www.jstor.org/stable/10.1086/660123.

Manski, Charles F. and David A. Wise (1983). *College Choice in America*. Cambridge, MA: Harvard University Press.

McFadden, Daniel (1973). "Conditional logit analysis of qualitative choice behavior." In: *Frontiers in Econometrics*. Ed. by Paul Zarembka. New York: Academic Press, pp. 105–142.

National Center for Education Statistics (2016a). *ELS: Education Longitudinal Study of 2002*. (Visited on 2016).

National Center for Education Statistics (2016b). *IPEDS: Integrated Postsecondary Education Data System*. (Visited on 2016).

Niu, Sunny Xinchun and Marta Tienda (2008). “Choosing colleges: Identifying and modeling choice sets.” In: *Social Science Research* 37.2, pp. 416–433.

Paulsen, Michael and John C. Smart (2001). *The Finance of Higher Education: Theory, Research, Policy, and Practice*. en. Algora Publishing. ISBN: 978-0-87586-135-7.

Perna, Laura W. (2006). “Studying college choice: A proposed conceptual model.” In: *Higher Education: Handbook of Theory and Research*. Vol. 21. Springer Netherlands. Chap. 3, pp. 99–157.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rouse, Cecilia Elena (1995). “Democratization or Diversion? The Effect of Community Colleges on Educational Attainment.” In: *Journal of Business & Economic Statistics* 13.2, pp. 217–224. URL: http://www.jstor.org/stable/1392376.

Snyder, T.D., C. de Brey, and S.A. Dillow (2016). *Digest of Education Statistics 2014*. NCES 2016-006. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Souza Briggs, X. de and W.J. Wilson (2006). *The Geography of Opportunity: Race and Housing Choice in Metropolitan America*. James A. Johnson Metro Series. Brookings Institution Press.

StataCorp (2015). *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP.

Tate, William F. (2008). “”Geography of Opportunity”: Poverty, Place, and Educational Outcomes.” In: *Educational Researcher* 37.7, pp. 397–411.

The Delta Cost Project (2016). *The Delta Cost Project Database*. (Visited on 2016).

Toutkoushian, Robert K. and Michael Paulsen (2016). *Economics of Higher Education: Background, Concepts, and Applications*. en. Springer. ISBN: 978-94-017-7506-9.

Turner, Sarah E. (2004). “Going to College and Finishing College: Explaining Different Educational Outcomes.” In: *College Choices: The Economics of Where to Go, When to Go,* *and How to Pay for It*. Ed. by Caroline M. Hoxby. Chicago: University of Chicago Press, pp. 13–62.

Umbach, Paul D (2007). “How effective are they? Exploring the impact of contingent faculty on undergraduate education.” In: *The Review of Higher Education* 30.2, pp. 91–123.

Xu, Di and Shanna Smith Jaggars (2011). “The Effectiveness of Distance Education Across Virginia’s Community Colleges: Evidence From Introductory College-Level Math and English Courses.” In: *Educational Evaluation and Policy Analysis* 33.3, pp. 360–377.

Xu, Di and Shanna Smith Jaggars (2013). “The impact of online learning on students’ course outcomes: Evidence from a large community and technical college system.” In: *Economics of Education Review* 37, pp. 46–57. DOI: 10.1016/j.econedurev.2013.08.001.

Table 1: Descriptive table of student characteristics.

	Full sample ($N = 9050$)			SAT > 1100 ($N = 3020$)			Income < \$25k ($N = 1270$)		
	All	Attended college	Did not attend	All	Attended college	Did not attend	All	Attended college	Did not attend
	-	74.85	25.15	-	85.61	14.39	-	63.95	36.05
Female	53.91	56.13	47.3	51.08	52.63	41.84	59.37	63.78	51.54
Black	11.34	10.31	14.4	2.71	2.74	2.53	24.66	22	29.39
Hispanic	9.56	8.71	12.12	5.16	5.1	5.52	16.44	15.45	18.2
Asian, Hawaii/P. Islander	9.91	10.26	8.87	12.83	12.48	14.94	17.47	20.15	12.72
Amer. Indian/Alaska Native	0.44	0.43	0.48	0.2	0.19	0.23	0.95	1.36	0.22
More than one race	4.3	3.95	5.31	3.67	3.28	5.98	4.98	4.08	6.58
Parental education	5	5.16	4.52	5.82	5.85	5.65	3.61	3.72	3.42
(categorical)	(2)	(1.96)	(2.02)	(1.77)	(1.75)	(1.87)	(1.98)	(1.97)	(1.99)
Family income	9.61	9.79	9.08	10.41	10.44	10.23	5.42	5.47	5.34
(categorical)	(2.21)	(2.14)	(2.35)	(1.86)	(1.81)	(2.11)	(1.57)	(1.56)	(1.59)
Composite SAT	1004.69	1034.44	916.14	1233.37	1236.41	1215.26	880.45	916.21	817.02
	(207.3)	(200.57)	(201.72)	(109.22)	(109.62)	(105.18)	(194.5)	(188.71)	(188.53)
Average # of applications	-	2.81	-	-	3.42	-	-	2.63	-
	-	(2.1)	-	-	(2.44)	-	-	(1.8)	-

Notes. Values are percentages unless otherwise noted (standard deviations in parentheses). Parental education categorical values are roughly ordered: 2 = graduated high school; 4 = graduated from a two-year school; 6 = graduated from a four-year school. All incomes less than \$25k correspond to income categorical values 1-7; 5 = (\$5k, \$10k], 9 = (\$35k, \$50k], and 10 = (\$50k, \$75k]. Per NCES regulations for restricted access data, observation numbers have been rounded to the nearest 10.

Table 2: Descriptive table of colleges and universities in choice set.

	Four-year			Two-year		
	Public	Private, non-profit	Private, for-profit	Public	Private, non-profit	Private, for-profit
In-state cost	829	10709	9369	294	5356	6802
	(1408)	(6176)	(5160)	(805)	(3861)	(4881)
Out-of-state cost	9105	13330	10594	2871	7282	8447
	(3791)	(6618)	(4829)	(2504)	(4364)	(4689)
Instructional expend. per FTE	6685	7605	3679	4681	6741	3806
	(7091)	(6802)	(2918)	(4139)	(7957)	(2878)
Student body median SAT	957	962	717	706	734	700
	(136)	(180)	(61)	(33)	(88)	(0)
Student faculty ratio	11.81	11.52	102.98	16.62	12.72	24.28
	(4.12)	(12.3)	(596.3)	(8.08)	(15.42)	(33.09)
% Tenured/tenure track faculty	54.76	36.07	1.04	21.99	6.25	0.67
	(19.38)	(32.75)	(8.96)	(23.61)	(17.81)	(6.43)
FTE student enrollment	9684	2255	1774	3582	321	474
	(9056)	(3249)	(7266)	(3435)	(271)	(487)
N	567	1171	271	996	98	303

Notes. Institutional characteristics come from IPEDS and the Delta Cost Project. Only Title IV postsecondary institutions that have a physical location in 2004 are included. In-state costs are in-state tuition and fees less average institutional, state, and federal grant awards; out-of-state costs are out-of-state tuition and fees less average institutional and federal grant awards. Median SAT scores are the average of the 25th and 75th percentile SAT composite scores reported by the institution. For institutions that did not report SAT percentiles, the median score was imputed using either the middle value in the range given by Barron's Competitive Index for its competitiveness category or, for all non-competitive and non-rated institutions (including two-year institutions), given a score of 700 per Long (2004).

Table 3: Results of conditional logistic choice model of student college decision among students who attended a postsecondary institution within two years of high school graduation.

	Long (2004)							
	1972		1982		1992		**2004**	
	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year	**Main effect**	**Two-year**
College costs								
Cost	0.4686**	1.4096**	0.5809**	0.7520**	0.6548**	0.8531**	**0.8387****	**0.7926****
(per $1000)	(32.32)	(7.47)	(26.68)	(3.84)	(39.21)	(4.5)	**(41.91)**	**(11.03)**
Distance	0.1665**	0.0534**	0.1954**	0.3382**	0.2668**	0.0805**	**0.1818****	**0.2292****
(per 100 miles)	(65.29)	(31.46)	(60.91)	(17.39)	(64.66)	(35.06)	**(48.99)**	**(10.49)**
Instructional Quality								
Instructional expend.	1.038	1.1248	1.0303	1.1595*	1.1035**	1.4999**	**1.0118**	**1.1824****
(per $1000)	(1.46)	(1.23)	(1.27)	(3.08)	(6.08)	(4.47)	**(3.05)**	**(8.93)**
Student faculty ratio	0.998		1.0003		1.0127		**1.0015****	
	(0.48)		(0.07)		(2.95)		**(9.59)**	
Tenure track faculty	1.0050**		1.0048**		1.0060**		**1.0710****	
(per 10% pts)	(7.18)		(5.46)		(6.2)		**(12.95)**	
Student college match								
Student SAT percentile	0.6525**		0.8662**		0.7129**		**0.6355****	
larger (per 10% pts)	(10.26)		(4.64)		(11.26)		**(12.54)**	
School SAT percentile	0.995		0.8324**		1.1809**		**1.5367****	
larger (per 10% pts)	(0.16)		(5.75)		(4.78)		**(5.60)**	
College level								
Two-year dummy	1.7242		1.1219		4.8538**		**3.5981****	
	(2.49)		(0.48)		(7.93)		**(11.69)**	
Students	5666	-	4881	-	5693	-	**6780**	**-**
Choice combinations	12118588	-	9651765	-	15011370	-	**23100000**	**-**

Notes. Data taken from the NCES Education Longitudinal Study of 2002, IPEDS, and the Delta Cost Project. Odds ratios are reported, with z-scores in parentheses. Standard errors (not reported) were clustered at the student level. Median SAT values of 700 were given to all two-year and non-selective institutions. Distances are calculated using the centroid of the student's census block group in 10th grade and the geocoordinates of each institution option. Costs are tuition and fees less average grants, taking into account whether student resides in the same state. Additional controls not reported include: cubic distance, SAT match variables squared, FTE, and FTE squared. Per NCES regulations for restricted access data, both observation cell sizes and unique cases have been rounded to the nearest 10 and 1000, respectively.

** indicates significance at the 5% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 4.25$)

* indicates significance at the 10% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 3.57$)

Table 4: Results of conditional logistic choice model of student college decision among students who attended a postsecondary institution within two years of high school graduation.

	Attend		Apply		Attend|Apply	
	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year
College costs						
Cost	0.8387**	0.7926**	0.8277**	0.8261**	0.9746	1.3001
(per $1000)	(41.91)	(11.03)	(62.11)	(8.76)	(2.67)	(2.74)
Distance	0.1818**	0.2292**	0.2728**	0.2286**	0.8239**	0.9182
(per 100 miles)	(48.99)	(10.49)	(69.18)	(11.80)	(6.42)	(0.99)
Instructional Quality						
Instructional expend.	1.0118	1.1824**	1.0251**	1.1018**	0.9941	1.0813
(per $1000)	(3.05)	(8.93)	(7.90)	(4.84)	(0.84)	(1.44)
Student faculty ratio	1.0015**		0.9985		0.9991**	
	(9.59)		(0.27)		(4.35)	
Tenure track faculty	1.0710**		1.0821**		0.9895	
(per 10% pts)	(12.95)		(16.23)		(1.01)	
Student college match						
Student SAT percentile	0.6355**		0.6197**		0.9402	
larger (per 10% pts)	(12.54)		(14.89)		(1.00)	
School SAT percentile	1.5367**		1.3058**		1.5593*	
larger (per 10% pts)	(5.60)		(6.91)		(3.81)	
College level						
Two-year dummy	3.5981**		2.8761**		1.5249	
	(11.69)		(9.84)		(2.13)	
Students	6780	-	6360	-	4460	-
Choice combinations	23100000	-	21700000	-	16000	-

Notes. Data taken from the NCES Education Longitudinal Study 2002, IPEDS, and the Delta Cost Project. Odds ratios are reported, with z-scores in parentheses. Standard errors (not reported) were clustered at the student level. Median SAT values of 700 were given to all two-year and non-selective institutions. Distances are calculated using the centroid of the student's census block group in 10th grade and the geocoordinates of each institution option. Costs are tuition and fees less average grants, taking into account whether student resides in the same state. Additional controls not reported include: cubic distance, SAT match variables squared, FTE, and FTE squared. Per NCES regulations for restricted access data, both observation cell sizes and unique cases have been rounded to the nearest 10 and 1000, respectively.

** indicates significance at the 5% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 4.25$)

* indicates significance at the 10% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 3.57$)

Table 5: Results of conditional logistic choice model of student college decision among high SAT (> 1100) students who attended a postsecondary institution within two years of high school graduation.

	Attend		Apply		Attend|Apply	
	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year
College costs						
Cost	0.8350**	1.0027	0.8243**	0.8935	0.9708	1.6808
(per $1000)	(28.22)	(0.02)	(39.30)	(1.92)	(2.29)	(1.48)
Distance	0.2699**	0.2659*	0.3459**	0.2694*	0.8596*	0.9822
(per 100 miles)	(35.63)	(3.86)	(45.93)	(3.84)	(3.76)	(0.06)
Instructional Quality						
Instructional expend.	1.0142	1.2430*	1.0287**	1.1548	0.9899	0.9582
(per $1000)	(2.15)	(4.12)	(6.28)	(2.38)	(1.03)	(0.33)
Student faculty ratio	0.9709*		0.9511**		1.0172	
	(4.18)		(8.32)		(1.29)	
Tenure track faculty	1.0589**		1.0759**		0.9792	
(per 10% pts)	(6.01)		(11.64)		(1.22)	
Student college match						
Student SAT percentile	0.6216**		0.6877**		0.8758	
larger (per 10% pts)	(8.80)		(8.59)		(1.46)	
School SAT percentile	4.0078**		1.8776**		2.0084	
larger (per 10% pts)	(5.46)		(4.94)		(2.24)	
College level						
Two-year dummy	5.1918**		4.7005**		2.6934	
	(5.71)		(5.07)		(1.93)	
Students	2590	-	2490	-	1930	-
Choice combinations	8815000	-	8495000	-	8000	-

Notes. Data taken from the NCES Education Longitudinal Study 2002, IPEDS, and the Delta Cost Project. Odds ratios are reported, with z-scores in parentheses. Standard errors (not reported) were clustered at the student level. Median SAT values of 700 were given to all two-year and non-selective institutions. Distances are calculated using the centroid of the student's census block group in 10th grade and the geocoordinates of each institution option. Costs are tuition and fees less average grants, taking into account whether student resides in the same state. Additional controls not reported include: cubic distance, SAT match variables squared, FTE, and FTE squared. Per NCES regulations for restricted access data, both observation cell sizes and unique cases have been rounded to the nearest 10 and 1000, respectively.

** indicates significance at the 5% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 4.25$)

* indicates significance at the 10% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 3.57$)

Table 6: Results of conditional logistic choice model of student college decision among low income (< $25k) students who attended a postsecondary institution within two years of high school graduation.

	Attend		Apply		Attend|Apply	
	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year
College costs						
Cost	0.8342**	0.8236*	0.8009**	0.8906	0.9578	1.1778
(per $1000)	(12.56)	(3.83)	(22.54)	(2.33)	(1.12)	(0.84)
Distance	0.0958**	0.3136**	0.2081**	0.3015**	0.6125**	0.8471
(per 100 miles)	(20.10)	(4.34)	(24.49)	(4.80)	(4.37)	(0.64)
Instructional Quality						
Instructional expend.	1.0044	1.2382**	1.0199	1.1922*	1.0224	1.3121
(per $1000)	(0.34)	(4.58)	(2.39)	(3.75)	(1.05)	(1.91)
Student faculty ratio	0.9961		0.9817		1.0252	
	(1.99)		(3.21)		(1.52)	
Tenure track faculty	1.0825**		1.0769**		1.0304	
(per 10% pts)	(4.98)		(6.47)		(0.89)	
Student college match						
Student SAT percentile	0.6521*		0.6505**		0.9011	
larger (per 10% pts)	(3.70)		(4.64)		(0.54)	
School SAT percentile	1.0754		1.0848		0.9316	
larger (per 10% pts)	(0.37)		(0.78)		(0.25)	
College level						
Two-year dummy	1.8972		1.3786		0.9577	
	(2.80)		(1.32)		(0.09)	
Students	810	-	760	-	530	-
Choice combinations	2756000	-	2592000	-	2000	-

Notes. Data taken from the NCES Education Longitudinal Study 2002, IPEDS, and the Delta Cost Project. Odds ratios are reported, with z-scores in parentheses. Standard errors (not reported) were clustered at the student level. Median SAT values of 700 were given to all two-year and non-selective institutions. Distances are calculated using the centroid of the student's census block group in 10th grade and the geocoordinates of each institution option. Costs are tuition and fees less average grants, taking into account whether student resides in the same state. Additional controls not reported include: cubic distance, SAT match variables squared, FTE, and FTE squared. Per NCES regulations for restricted access data, both observation cell sizes and unique cases have been rounded to the nearest 10 and 1000, respectively.
** indicates significance at the 5% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 4.25$)
* indicates significance at the 10% level, with adjustment for significance to take into account potential survey design effects ($z_{critical} \approx 3.57$)

Table 7: Comparison of college enrollment decision between models using most likely institution and models using nearest public two-year institution.

	All				Low Income/Low SAT			
	Most-likely (attend)	Most-likely (apply)	Most-likely (att|app)	Closest public 2-year	Most-likely (attend)	Most-likely (apply)	Most-likely (att|app)	Closest public 2-year
College costs								
Cost	1.0010	0.9892	0.8112	0.9920	1.0020	0.9988	0.9609	0.9917
(per $1000)	(0.05)	(0.75)	(0.90)	(0.16)	(0.04)	(0.03)	(0.11)	(0.12)
Cost^2	0.9996	1.0000	1.0456	1.0018	0.9996	0.9998	1.0239	1.0087
(per $1000^2)	(1.05)	(0.14)	(1.08)	(0.19)	(0.47)	(0.25)	(0.36)	(0.61)
Distance	0.3874**	0.9480	0.8576	0.2770**	0.3796*	0.9903	0.8651	0.1749*
(per 100 miles)	(2.76)	(0.21)	(1.61)	(2.01)	(1.74)	(0.03)	(1.00)	(1.87)
Distance^2	2.2653*	1.0036	1.0144*	2.2030	1.9171	0.9617	1.0134	10.7318
(per 100^2 miles)	(1.70)	(0.01)	(1.74)	(0.51)	(0.76)	(0.11)	(1.07)	(1.04)
Instructional Quality								
Instructional expend.	1.0048	1.0284**	0.8763**	1.0135	1.0227	1.0441**	0.9038	0.9946
(per $1000)	(0.32)	(2.34)	(2.54)	(0.89)	(0.81)	(2.19)	(1.28)	(0.18)
Instructional expend.^{2}	1.0002	0.9995	1.0067**	1.0000	1.0000	0.9992	1.0048	1.0006
(per $1000^2)	(0.46)	(1.52)	(2.03)	(0.11)	(0.02)	(1.48)	(0.84)	(0.67)
Student								
SAT %tile (10s)	1.3624**	1.3673**	1.3832**	1.3886**	1.3845**	1.4030**	1.4057**	1.4136**
	(13.55)	(14.51)	(17.49)	(18.58)	(8.16)	(8.33)	(9.50)	(10.14)
Family income	1.0524**	1.0544**	1.0551**	1.0497**	1.0624**	1.0670**	1.0645**	1.0608**
	(3.90)	(4.04)	(4.11)	(3.70)	(3.58)	(3.84)	(3.72)	(3.51)
Parental education	1.0574**	1.0610**	1.0601**	1.0572**	1.0560**	1.0618**	1.0594**	1.0552**
	(3.90)	(4.13)	(4.09)	(3.88)	(2.67)	(2.95)	(2.84)	(2.63)
Female	1.5785**	1.5857**	1.5775**	1.5826**	1.5741**	1.5892**	1.5693**	1.5737**
	(8.96)	(9.05)	(8.96)	(9.01)	(6.02)	(6.15)	(5.99)	(6.03)
Black	1.1407	1.1897**	1.1933**	1.1152	1.2069*	1.2830**	1.2728**	1.1756
	(1.57)	(2.07)	(2.12)	(1.30)	(1.81)	(2.40)	(2.35)	(1.56)
Hispanic	0.9499	0.9930	0.9936	0.9247	1.0521	1.1346	1.0994	0.9931
	(0.58)	(0.08)	(0.07)	(0.88)	(0.42)	(1.05)	(0.78)	(0.06)
Asian, Hawaii/P. Islander	1.0128	1.0542	1.1109	0.9864	1.4117**	1.5556**	1.5665**	1.3687**
	(0.14)	(0.56)	(1.09)	(0.14)	(2.40)	(3.08)	(3.02)	(2.15)
Amer. Indian/Alaska Native	1.1166	1.1773	1.1355	1.1208	1.5256	1.6291	1.5839	1.5475
	(0.30)	(0.45)	(0.35)	(0.31)	(0.94)	(1.08)	(1.02)	(0.97)
More than one race	0.7081**	0.7172**	0.7271**	0.7072**	0.8355	0.8541	0.8647	0.8355
	(2.88)	(2.78)	(2.66)	(2.89)	(0.97)	(0.85)	(0.78)	(0.97)
Labor Market								
County unem. rate (2004)	1.0340**	1.0333*	1.0432**	1.0307*	1.0625**	1.0614**	1.0717**	1.0585**
	(1.96)	(1.91)	(2.42)	(1.77)	(2.58)	(2.53)	(2.90)	(2.42)
N	9050	9050	9050	9050	3290	3290	3290	3290

Notes. Odds ratios are reported, with z-scores in parentheses. Students included in the low income/low SAT column reported family incomes lower than $25,000 per year or had an SAT or SAT-equivalent score ¡ 900. Additional controls include cubic distance, FTE college enrollment, and FTE squared. Per NCES regulations for restricted access data, observation numbers have been rounded to the nearest 10.
** indicates significance at the 5% level.
* indicates significance at the 10% level.

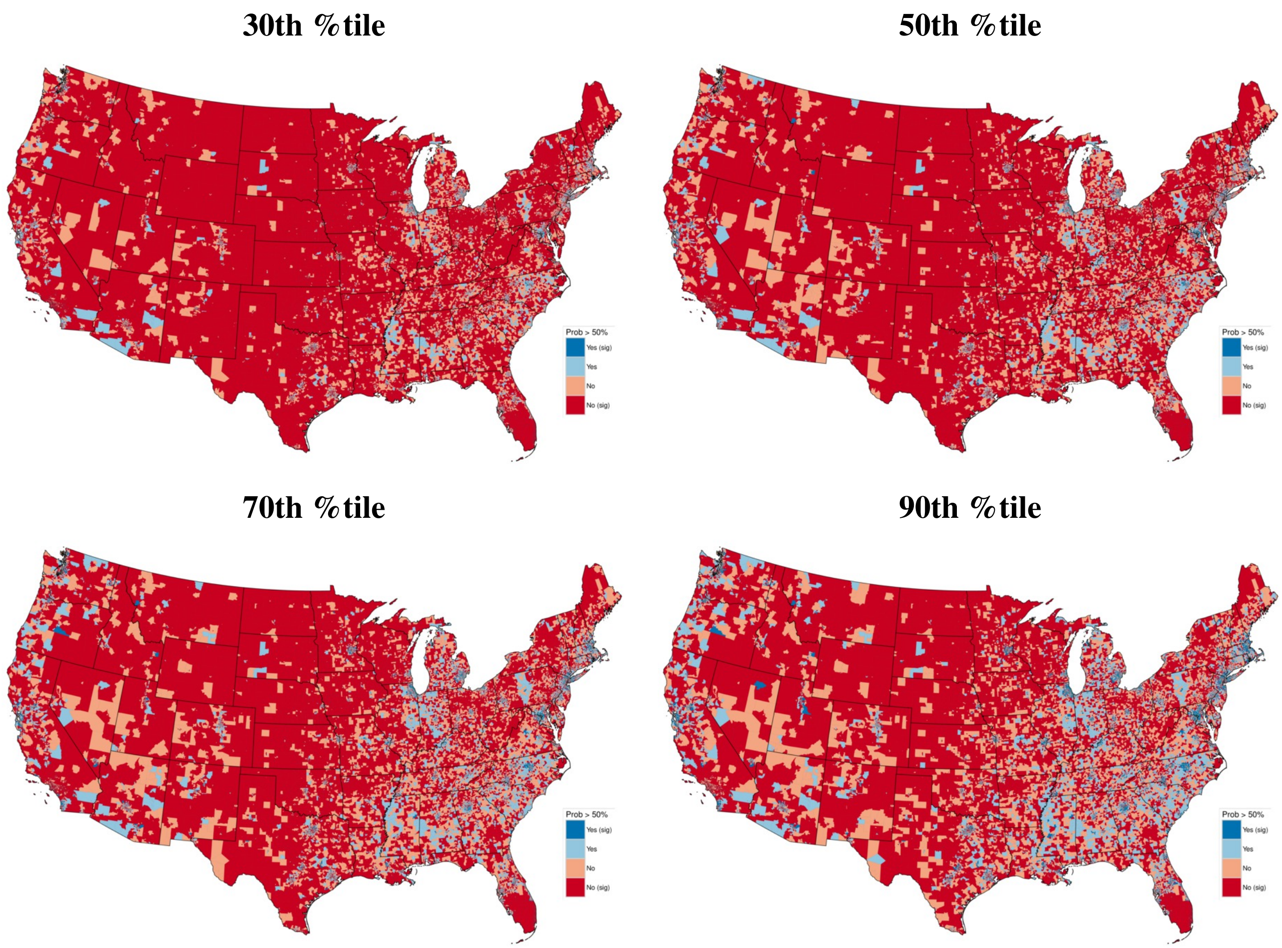

Figure 1: Visualization of predicted probability that an synthetic modal student at the census tract level will enroll in college within two-years of earning high school diploma/GED based on characteristics of his/her most-likely college choice. Shades of blue represent predicted probabilities greater than 0.5 (more likely than not); shades of red represent predicted probabilities less than 0.5 (less likely than not). In both cases, shades are darker when statistically significant. Because Census data used to create modal students does not contain SAT scores, each synthetic student was given multiple scores across the distribution of possible scores. Each map shows results for sythetic students given SAT scores at the stated point in the distribution: 30th, 50th, 70th, and 90th.

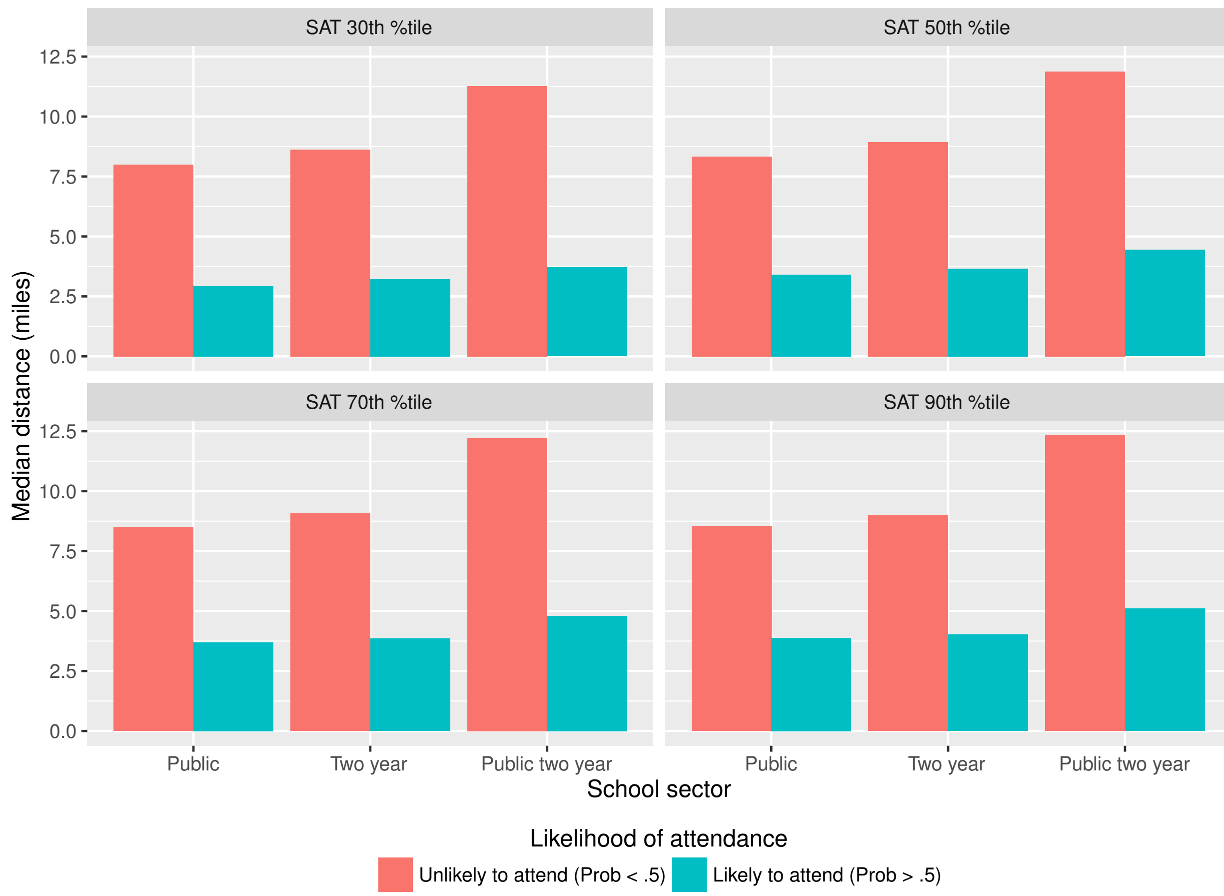

Figure 2: Median distances in miles to nearest public, two-year, and public two-year for census tracts that produced modal synthetic students likely and unlikely (Prob > 0.5) to enroll within two years of high school graduation across distribution of SAT percentiles. Only those tracts with statistically significant positive or negative predictions ($p < 0.05$) are included.