CrossMark

Choosing College in the 2000s: An Updated Analysis Using the Conditional Logistic Choice Model

Benjamin T. Skinner[1]

Abstract   In this paper I investigate the college enrollment decisions of a nationally representative cohort of students who first attended in the mid-2000s. I find that while cost, distance, and match continued to be important in the choice between colleges, characteristics of the most-likely college choice appear less important in the choice of whether to enroll at all when controlling for student characteristics and local labor market conditions. Subpopulation analyses on students with high SAT scores and students with low family income, two groups that remain the focus of many financial aid policies, indicate some differences in the way these particular students chose college. Extending prior work by modeling discrete steps in the enrollment decision process—application and enrollment conditional on application—I find choice characteristics were most significant in the application stage. These results support other research that shows students may self-select out of potentially better college matches due to lack of information about actual costs or limited geographic opportunity.

Keywords   College access · College choice · Conditional logistic choice model · Geographic opportunity

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, the country's higher education policies instead largely work 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

✉ Benjamin T. Skinner
   btskinner@virginia.edu

[1] University of Virginia, PO BOX 400879, Charlottesville, VA 22904, USA

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 relationships 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 their 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.

This study serves as an update to Long (2004) in that it applies the same conditional logistic choice model (McFadden 1973) to the college enrollment decisions of students who graduated from high school in the mid-2000s. Combining student-level data from the Education Longitudinal Study of 2002 (National Center for Education Statistics 2016) with postsecondary institution data, I first model the college enrollment decision of nearly 6800 students who enrolled within two years of earning their high school or general equivalency diplomas as a choice between 3406 postsecondary institutions (approximately 23 million student-college choice pairs). I next model the college enrollment decision as choice of whether to enroll at all—this time expanding the analysis sample to include students who did not attend college—using characteristics of both the individual student and their most-likely college choice as predicted by the first model. My results show that while cost, distance, and academic match remained important in the choice between colleges, characteristics of the most-likely college choice were less predictive of whether a student enrolled at all when individual characteristics and local labor market conditions are taken into account. I find some heterogeneity in the response to college characteristics when choosing between colleges among various student subgroups—high SAT, low family income, low SAT/low family income—but similarly little change in the choice of whether to enroll.

As an extension of prior work, I take advantage of rich ELS data on application behavior to split the enrollment decision into its component parts. In addition to modeling attendance collapsed into a single application-enrollment outcome, I also separately model a student's choice of (1) where to apply among all schools and (2) which among the schools in their application subset to select. Across all college-goers as well as high-SAT and low-income student subgroups, I find that the characteristics of the student-college choice are more significant in the application stage than the subsequent selection stage. This result supports other research which shows that students may self-select out of better college matches due to the comparatively lower level of information—particularly in terms of the real cost of attendance—by the choices made in the application period (Avery et al. 2006).

Though the majority of the students in the ELS sample are now over a decade past their high school graduation, their college enrollment decisions may be usefully compared to those of prior cohorts to better understand long-term trends, such as the increasing importance of academic match. Relevant to college choices being made by current high school graduates, evidence that many students in the 2004 cohort effectively made their college

choices in the application stage when actual costs were less likely to be known provides further support for interventions that increase information prior to application (Hoxby and Turner 2013) as well as policies that reduce the complexity of the financial aid process (Dynarski and Scott-Clayton 2013). Even in the face of changing demographics, growth in the online and for-profit sectors, and increased college costs (Eagan et al. 2016; Ma et al. 2016; Allen et al. 2016; Cottom 2017), the college enrollment decisions of students who first attended in the mid-2000s remain important to examine.

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 in terms of a human capital investment (Becker 2009; Toutkoushian and Paulsen 2016; Paulsen and Smart 2001).

Much of the college choice literature based in human capital theory has been concerned with the cost of attendance (Fuller et al. 1982; Manski and Wise 1983; Leslie and Brinkman 1987; Kane 1995, 1996; Dynarski 2002; Avery and Hoxby 2004; Deming and Dynarski 2009). 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 (Turner 2004; Toutkoushian and Paulsen 2016; 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 that they delay enrollment or forego college altogether. Empirical evidence has borne this theory out, with a number of studies finding that increased college costs are associated with decreases in the likelihood that a potential student will enroll in a given school or any school (Fuller et al. 1982; Leslie and Brinkman 1987; Kane 1995; Avery and Hoxby 2004; Long 2004).

Distance between student and school represents an indirect cost 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 to predict the likelihood of college enrollment (Baker and Doyle 2017; Card 1999; Rouse 1995; Doyle and Skinner 2016, 2017). Bettinger and Long (2009) use each student's closest college as an instrument for selection into remedial education courses. Similarly, Xu and Jaggars (2013) use the distance between student and campus to instrument their likelihood of enrolling in an online course section. Shared by these studies is the belief that distance matters to students when choosing between 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 an 83–73%

reduction in the odds of 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 they had to travel. 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 nearby college options, have on college choice, particularly 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), negative findings of recent quasi-experimental studies suggest that increased reliance on distance education courses may come at the cost of lower student retention and rates of success (Hart et al. 2016; Xu and Jaggars 2011, 2013; Bettinger et al. 2017; Huntington-Klein et al. 2017).

Costs, however, are not the only determinants of college choice. Increases in the returns to the education at a particular college can 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 (Black and Smith 2004, 2006; Eide et al. 1998; Long 2010). 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. Hoxby and Avery (2012) find that high-income, high-achieving students in their sample were likely to consider academic match when applying to schools with a range of average admissions test scores. This set them apart from their equally high-achieving but low-income peers who did not follow the same application pattern and were therefore more likely to "undermatch." Alongside other measures, research suggests that many students consider institutional quality when making their enrollment decisions.

This study represents a contribution to the literature for two reasons. First, I investigate aspects of the college enrollment decision—cost, quality, and match—for a more recent nationally-representative sample of high school graduates. Second, I focus on reproducing the methods and analyses of 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. While the higher education landscape has undoubtedly changed since most in the ELS cohort made their college enrollment decisions (Eagan et al. 2016; Ma et al. 2016), this study adds to our understanding of how past students have chosen between colleges and thus provides a firmer base for predicting how current and future students may do the same.

Method and Estimation Strategy

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. For many students, it may be the case that the first step is subsumed in the second, that is, the decision of whether to attend only having salience if the option set is poor. Assuming the desire to attend and the availability of choices, however, the enrollment decision is best understood as a choice between a set of postsecondary options.

The conditional logistic choice model, also known as McFadden's discrete choice model (McFadden 1973), can be 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 choice preferences, the conditional logistic choice 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 the model.

The conditional logit applies well to the college enrollment decision as students who wish 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 fully specifies every choice alternative for an 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 included in the model; 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 must meet the independence of irrelevant alternatives (IIA) assumption. Strictly, the IIA assumption requires that the odds of choosing *A* over *B* is the same whether or not other options are available (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 comparative preference for Big State University over Small Private College will remain the same even when Mid-size Regional Tech is an option. Should the inclusion of a new college in the alternative choice set affect the comparative odds of other binary choice pairs, then results may be inconsistent.

---

[1] For clarity and space in this section, I collapse the college enrollment decision into the single decision to attend. Application to college as an intermediate step in the enrollment process, modeled separately from attendance, is implied.

Unfortunately, the IIA assumption cannot be directly tested. While a number of statistical tests and ad-hoc methods have been proposed, they produce inconsistent results in applied work (Cheng and Long 2007). The large model matrix used in the study combined with the complexity of estimating the conditional logit also means that dropping options from the choice set and comparing results across full and sub-choice set, a common approach, is less practicable. Against threats of bias due to violations of the IIA assumption, however, the study design is supported by the fullness of the choice set, the nature of the college application and enrollment process, and the status of colleges as distinct choices in the eyes of potential applicants.

First, I include all Title IV postsecondary institutions with a physical location in the choice set for each student. I exclude only primarily distance education-based institutions as they did not have physical locations with which to measure distance to students. Such that sample students saw virtual institutions as viable higher education options or enrolled only in online courses, the omission of these schools could bias the results. Few students, however, completed their degree programs solely through distance education institutions in the 2000s (Snyder et al. 2016). Against the threat to the IIA assumption due to the omission of a college option, each student's choice set includes the same 3406 options—no matter how unlikely particular student-college pairs may have been—that together comprise the near-universe of relevant college options for students at the time.

Second, the college application and enrollment process supports the IIA assumption. Because students apply to colleges independently, 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. The IIA assumption may be violated if students were able to signal their application or enrollment intentions to schools that then collaborated when making admissions offers. By design of the college application and enrollment process, however, schools do not (and should not due to legal restrictions) collaborate when deciding offers of admission or aid, meaning that their decisions should be independent.

A final threat to the IIA assumption occurs when the choice set includes alternatives that may be considered "close substitutes," (McFadden 1973, p. 113). In the canonical example, public transit riders are unlikely to distinguish between red and blue buses, meaning that conditional choice models that include them as distinct travel alternatives will produce biased estimates of rider preferences. Threat due to "close substitutes" is a small concern for this study due to the fact that colleges represent distinct choices for students. Though two colleges may have similar attributes and appear virtually identical in the context of all postsecondary institutions, each represents a distinct option due to its location, branding, faculty composition, and culture. Quoting an earlier paper by McFadden, Cheng and Long (2007) write, "conditional logit models should only be used in cases where the outcome categories 'can plausibly be assumed to be distinct and weighed independently in the eyes of each decision maker,'" (p. 598). The importance of the college enrollment decision for students strongly suggests that this is the case and the conditional logistic choice model, therefore, is an appropriate means of examining the preferences underlying that choice.

Modeling a Student's Choice of Whether to Enroll

Though the conditional logit model of the choice *between* colleges 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 outcome

of *whether* a student enrolled, I use a logit link to regress an indicator for enrollment on characteristics of both the student and the most-likely college, C_{most}. In this model, C_{most} is the college, k, 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, j. Probabilities are computed using the fitted parameters from conditional logit models for three outcomes: attendance, application, and attendance 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 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 a logistic regression equation using the characteristics of each student's 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.

Data

Student-level data come from all waves of the ELS survey (National Center for Education Statistics 2016). Administered by the National Center for Education Statistics, ELS follows 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 census block group 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) subsamples, are shown in Table 1.

ELS restricted-use data contains the unique institutional ID of all colleges at 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 IPEDS record. Distances between each student-college pair were computed using the student's base-year census block group of residence and institution's coordinates. In all cases, the geographic center of the census block group, provided by the United States Census Bureau, was used as the student's location. One level below census tracts, census block groups represent one of the smallest geographic areas constructed by the Census, generally containing 600–3000 individuals.[2] Though the measure of distance between each student-college pair is necessarily coarser than it would

---

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

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)	–

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 = ($5 k, $10 k), 9 = ($35 k, $50 k), and 10 = ($50 k, $75 k). Per NCES regulations for restricted access data, observation numbers have been rounded to the nearest 10

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 does not report geographic coordinates of institutions for 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. These include cost of attendance, 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, including all public two-year colleges as well as those rated as non-competitive, I follow 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 number of faculty with a PhD is not readily available in current data sets. Second, studies of student application behavior (Drewes and Michael 2006) and survey responses (Umbach 2007) suggest that the proportion of tenure-line faculty may be a more salient measure of institutional quality for students.

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 generated 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.

---

[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 similar to those obtained when the unadjusted fitted values are used.

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	10,709	9369	294	5356	6802
	(1408)	(6176)	(5160)	(805)	(3861)	(4881)
Out-of-state cost	9105	13,330	10,594	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

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)

Measures of county-level unemployment rates that are used in the logistic regression on the decision of whether to enroll come from the Bureau of Labor Statistics. As with the institutional data, these measures describe 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 market can change a student's baseline earnings potential, thereby adjusting the expected return to a college degree and the likelihood a student will choose to enroll over immediately entering the workforce.

Results

Choice Between Colleges

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 6780 students × 3409 college options). Each choice pair observation was modeled using the following predictors: cost, distance in miles between student and school, instructional expenditures per full-time equivalent student, 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 two values, an indicator variable for two-year institutions, and various interactions with the two-year indicator.[5] To facilitate comparison with prior cohorts, odds ratios and *z*-scores taken from Long (2004) are reprinted in the first six columns of Table 3. The last two columns of Table 3 show the new results, also presented as odds ratios, for the class of 2004.

How did the college enrollment decisions of students who first attended in the mid-2000s compare to those of prior cohorts? Turning first to college costs, expected monetary costs (tuition and fees less grants) continued to decline in relevance for between-college choice among four-year institutions. Every $1000 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 preferred lower costs at four-year institutions, they continued the trend toward less cost sensitivity over time. Considering two-year schools, however, the 2004 cohort were a little more sensitive to cost than their 1992 counterparts: 21 versus 15% reduction in the relative odds of choosing the more costly institution for every $1000 differential. This increased sensitivity to cost in the two-year sector may reflect the growth of the for-profit sector during this period, which, despite being more expensive on average, gave students

---

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

[5] Full-time equivalent enrollment and various polynomial term were also included in the model but are not reported.

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)	

Table 3 continued

	Long (2004)						2004	
	1972		1982		1992			
	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year	Main effect	Two-year
Students	5666	–	4881	–	5693	–	**6780**	–
Choice combinations	12,118,588	–	9,651,765	–	15,011,370	–	**23,100,000**	–

Estimates in the first six columns for the 1972, 1982, and 1992 cohorts are reprinted from Long (2004) for comparison. Bolded estimates in the last two columns are results from new analyses that use data from the NCES Education Longitudinal Study of 2002, IPEDS, and the Delta Cost Project. Throughout, 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: squared and cubic distance, squared cost, squared instructional expenditures per FTE student, SAT match variables squared, FTE, FTE squared, and interactions between polynomials and the indicator for two-year institutions. Per NCES regulations for restricted access data, both observation cell sizes and unique cases have been rounded to the nearest 10 and 1000, respectively, in the last two columns of new estimates. Note that Long's specification used proportion of faculty with a PhD rather than the proportion on the tenure track

**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 ($z_{critical} \approx 3.57$)

more options for earning subbaccalaureate credentials and therefore an increased ability to choose based on relative costs (Deming et al. 2012; Baum et al. 2011; Baum and Ma 2012).

For every 100 mile difference, students were 82 and 77% less likely to choose the more distant four-year and two-year institutions, respectively. For four-year institutions, this relative odds reduction is greater than that showed by the 1992 cohort (73%), but closer to those of the 1982 (81%) and 1972 (83%) cohorts, suggesting that students have generally remained steady in their relative preference for closer four-year colleges and universities. For two-year institutions, the pattern is less clear with successive cohorts alternating between comparatively higher (1972: 95%; 1992: 92%) and lower (1972: 66%) odds reductions per 100 miles. Though this pattern may be an artifact of the data, it is also possible that increases in the number of online offerings since the 2000s, again at for-profit institutions offering subbaccalaureate credentials, may have reduced the role that distance played in the choice between two-year institutions (Allen and Seaman 2011, 2013). Regardless of precise values, however, the 2004 cohort continued to prefer closer schools at both institutional levels.

Concerning instructional quality, results show no significant association between instructional expenditures per FTE student and college choice at four-year universities for the 2004 cohort. Choosing between two-year institutions, however, the 2004 cohort showed a preference for higher expenditures, with 18% increase in relative odds per $1000. Compared to statistically significant relative odds increases of 10 and 50% at four- and two-year schools for the 1992 cohort, these findings suggest a weakening of the relationship. The 2004 cohort showed a small, but significant positive preference for an increase in the student-to-faculty ratio, with an approximate 2% increase in relative odds per 10–1 increase in the number of students to faculty. Even though these results control for FTE enrollments, they may reflect a trend among students to concentrate at larger institutions. In the fall of 1993, the 12% of institutions with more than 10,000 students enrolled 51% of all students (Snyder and Hoffman 1995). By the fall of 2005, an equal percentage enrolled 54% of all students, with some of the top enrollers having strong online components that would allow for greater numbers of students per faculty member (Snyder et al. 2007).

As another measure of institutional quality, the employment status of faculty appeared important to the 2004 cohort. Because my operationalization differs from that of Long (2004), I cannot argue that students became increasingly swayed by faculty composition. The proportion of tenure-line faculty may simply have been more salient to the 2004 cohort, either directly or as a proxy for other desirable school characteristics, than the proportion of faculty with a PhD was for earlier cohorts. Lack of comparison aside, students in the 2004 cohort showed a 7% increase in odds of choosing a school per 10% relative increase in the proportion of tenure-line faculty. While relatively small, this estimate is an order of magnitude larger than that shown by the 1972, 1982, and 1992 cohorts in their preferences for faculty with PhDs.

Regarding student-college match, 2004 cohort members were both less likely to prefer a college with a median SAT score below theirs and more likely to choose a 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 they attended were reduced 36%. In the opposite direction, every 10% increase in the school's median score over that of the student's increased their odds of attendance by 54%. Differing from the 1982 cohort, who appeared to prefer peer schools, all else being equal, the 2004 cohort followed their 1992 predecessors in preferring schools with higher achieving students and increasingly eschewing schools at which average SAT scores were lower.

Finally, I find that the odds of choosing a two-year institution remained high and significant for the 2004 cohort. While the 1972 and 1982 cohorts showed no strong relative preference for two-years over four-years, the 1992 cohort were 4.9 times as likely to choose a two-year institution. Like their immediate predecessors, the 2004 cohort were 3.6 times as likely to choose a two-year college as a four-year institution. As more high school graduates have chosen to go to college, the college-going population has changed to include more students of color, those from low income backgrounds, and first generation students who are more likely to choose two-year institutions (Baum et al. 2011). Results for the 2004 cohort appear reflective of this trend.

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 attendance 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 discussed above.

For the application conditional logit model, students were once again assumed to have the same full set of college choice options, 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.[6] Unlike in the first model, students in the application model could have multiple positive outcomes as they could apply to more than one school. In the third choice model, attendance conditional 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 appear to have been more amenable to schools that were less expensive, geographically closer, had higher instructional expenditures, more tenure track faculty, and had student bodies with as good as but preferably higher SAT scores. They also preferred, all else equal, two-year over four-year colleges. Turning to the third model, however, only the coefficient on distance to four-year colleges remains both statistically and practically significant. While distance retains its negative association, it is much reduced. Conditional on having applied, students showed 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 provide evidence that student choice preferences may have been more important in the decision of where to apply than in the subsequent decision of where to attend. In other words, student enrollment decisions could be better described as student

---

[6] 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.

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	23,100,000	–	21,700,000	–	16,000	–

Odds ratios are reported, with *z*-scores in parentheses. Standard errors (not reported) were clustered at the student level. See notes for Table 3 on variable construction and controls included in the model but not reported. For comparison, results reported in columns 1 and 2 are the same as those reported in columns 7 and 8 of Table 3. The outcome for the apply model (columns 3 and 4) is equal to one for all schools to which a student applied. The outcome for the attendance conditional on application model (last two columns) equals one for the school which the student attended, but with a choice set limited to schools to which the student applied. 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 ($z_{critical} \approx 3.57$)

application decisions. Two caveats to this interpretation apply. First, some students in the full sample applied 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, 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.

As shown in Table 1, however, the average number of applications was 2.81. Furthermore, the subsample used to fit the third model represents a majority of the second model sample ($\approx 70\%$). Taken together, results across models suggest that the 2004 cohort's

enrollment decisions were 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—appear to have been less salient when choosing where to attend from within their self-selected options.[7] These results align with other research which finds that students may be making suboptimal decisions due to incomplete information or support (Avery et al. 2006). Though my findings do not speak to the long-term outcomes of each student's choice, they do provide evidence that on the whole, the 2004 cohort effectively made their enrollment decisions during the application process, a time when students generally do not have the same level of information regarding actual college costs that they do after being admitted and receiving personalized offers.

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 logistic choice models—unconditional attendance, application, and attendance conditional on application—for two different groups: high SAT students and low income students. Because these two groups are the focus of many higher education policies, particularly merit and need-based financial aid policies, their college choice preferences remain important to investigate.

Results for students with a composite SAT score above 1100 are shown in Table 5. Across the models, high SAT students showed preferences generally similar to those of the full sample. And like the full sample, their preferences appear to have been 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 important 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 mattered slightly less for high SAT students, even losing statistical significance in some models. In the first two models, the reduction in odds per 100 miles to four-year college was 65–75% percent. While still significantly negative, 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 lies in student college match. Compared to the full sample, high SAT students showed a much greater preference for colleges at which the student body median SAT was higher. For every 10% point difference, these students preferred 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. All told, high SAT students appear to have been willing to travel a little farther if it meant attending a school with a higher achieving student body.

Low income students, on the other hand, were 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

---

[7] As a test on robustness, I also fit a conditional logit choice model on attendance conditional on acceptance. Most of the multiple applicants (∼ 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.

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	8,815,000	–	8,495,000	–	8000	–

See notes for Table 4

**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 ($z_{critical} \approx 3.57$)

survey, low income students showed only 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 did not show significant preference for schools with higher SAT scores in any of the models. Across all models, low income students were more sensitive to distance to four-year colleges than the full sample. The trend reverses for two-year colleges, with distance having had 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 appealed to low income students with greater need to work (Baum et al. 2011), thereby increasing their willingness to apply to and attend them even if they were a little farther away from home.

Table 6 Results of conditional logistic choice model of student college decision among low income (< $25 k) 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	2,756,000	–	2,592,000	–	2000	–

See notes for Table 4

**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 ($z_{critical} \approx 3.57$)

Choice Whether to Attend College

The decision of whether to enroll was modeled as a function of student characteristics and characteristics of either (1) a most-likely college as predicted by one of the conditional logit models or (2) the nearest public two-year institution. Across all specifications, the right-hand side covariates included characteristics of the student college choice as well as student-specific characteristics. Results for these equations are presented in Table 7. The first four columns present estimates for the full analytic sample, which now includes students who did not attend college within two years of graduating high school. 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. This subset of students is considered since they are likely more sensitive to the characteristics of their choice set when deciding whether to attend college and therefore more responsive to policies that adjust their options.

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)

Table 7 (continued)

	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
Female	1.5785** (8.96)	1.5857** (9.05)	1.5775** (8.96)	1.5826** (9.01)	1.5741** (6.02)	1.5892** (6.15)	1.5693** (5.99)	1.5737** (6.03)
Black	1.1407 (1.57)	1.1897** (2.07)	1.1933** (2.12)	1.1152 (1.30)	1.2069* (1.81)	1.2830** (2.40)	1.2728** (2.35)	1.1756 (1.56)
Hispanic	0.9499 (0.58)	0.9930 (0.08)	0.9936 (0.07)	0.9247 (0.88)	1.0521 (0.42)	1.1346 (1.05)	1.0994 (0.78)	0.9931 (0.06)
Asian, Hawaii/P. Islander	1.0128 (0.14)	1.0542 (0.56)	1.1109 (1.09)	0.9864 (0.14)	1.4117** (2.40)	1.5556** (3.08)	1.5665** (3.02)	1.3687** (2.15)
Amer. Indian/Alaska Native	1.1166 (0.30)	1.1773 (0.45)	1.1355 (0.35)	1.1208 (0.31)	1.5256	1.6291	1.5839	1.5475
More than one race	0.7081** (2.88)	0.7172** (2.78)	0.7271** (2.66)	0.7072** (2.89)	0.8355 (0.94)	0.8541 (0.85)	0.8647 (0.78)	0.8355 (0.97)
Labor market								
County unem. rate (2004)	1.0340** (1.96)	1.0333* (1.91)	1.0432** (2.42)	1.0307* (1.77)	1.0625** (2.58)	1.0614** (2.53)	1.0717** (2.90)	1.0585** (2.42)
N	9050	9050	9050	9050	3290	3290	3290	3290

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

Across all full sample models, characteristics of the most-likely or nearest public two-year college have little predictive power concerning the likelihood that a 2004 cohort student enrolled 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 attendance 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. Across all models, students were 36–46% more likely to enroll for every 10th percentile increase in their composite SAT scores. Similarly, enrollment likelihood was 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 were 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, black students in both samples and Asian / Hawaiian / Pacific Islander students at the margins may have been more likely to enroll, whereas students who identify as multiracial less likely.

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. Controlling for student ability and socioeconomic status, a one percentage point increase in the unemployment rate was associated with a 3–4% increase in the odds of enrollment among the full sample. Among the subsample of students on the margins, the increase was 6 to 7%. These consistently significant results provide further evidence for a negative association between local labor market conditions and enrollment (Manski and Wise 1983; Betts and McFarland 1995). Even in a college-for-all era, members of the 2004 cohort appear to have considered immediate employment a viable alternative.

The similarity between these largely non-significant school parameters and those Long (2004) reports for the 1992 cohort suggests that the decision of whether to enroll did not significantly change in that time period. Conditional on student characteristics and local employment opportunities, characteristics of the most-likely and nearest two-year public college seem to have been 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 logistic choice models may not actually have been that likely. If that were the case, characteristics of the school should not be significantly predictive of the decision to enroll. As a check on this scenario, I performed a number of sensitivity analyses. 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 approximately 10 percentage points (%pts). For the application and attendance conditional on application models, these values were [6, ≈ 12%pts][8] and [966, ≈ 0.1%pts], 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 sector, 21%.

Taken together these sensitivity analyses offer evidence that while not perfect, the predictions from the conditional logistic choice models fit the data generally 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 observationally similar to those actually chosen, then non-significant estimates suggest that cost and distance, which have strong negative associations with the choice between colleges, may not similarly have mattered for the 2004 cohort when choosing whether to enroll in college.

Discussion and Implications

In the choice between colleges, particularly in the application stage, I find that the 2004 cohort remained sensitive to cost and distance. Though their cost sensitivity follow the abatement trend established by earlier cohorts, general sensitivity to distance remained high. I further find that college match increasingly mattered. While students were less likely to choose schools with student body SAT scores below their own, they preferred schools at which the average scores were higher. I also find that students had a moderate preference for schools with greater proportions of tenure-line faculty and continued to prefer, all else being equal, two-year institutions. As a subgroup, high-SAT students were less sensitive to distance and even more willing to choose schools with comparatively higher average SAT scores; on the other hand, low-income students were both more sensitive to four-year distance and less sensitive to two-year distance than the full sample. In the choice of whether to enroll at all, however, results suggest that characteristics of a student's most-likely or nearest public two-year college were less important. While cost and distance may have mattered when choosing between colleges, I do not find consistent evidence that these considerations moderated the likelihood of enrollment when controlling for local labor market conditions and student characteristics. This finding holds across subgroups.

---

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

Separating the choice between colleges into the application and attendance conditional on application stages, I find that parameters lose significance in the attendance conditional on application model. Cross-model comparisons provide evidence that for many students in the 2004 cohort, the college enrollment decision was effectively made in the choice of where to apply rather than after offers of admission were received. Changes in the sample across models require that comparisons be made with caution. Yet if aspects of the college choice were more salient to students in the application stage, it is possible that some self-selected out of applying to colleges that would have, for example, offered them aid to offset the sticker price and at which they would have been successful—perhaps even more so than at the institution they chose.

A few limitations to this study should be noted. First, students who did not report SAT or SAT-equivalent test scores were not included. 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 did not take the SAT or equivalent exams were differentially affected by college costs and distance should they have decided to enroll, results may be less generalizable to the full population of high school graduates from this time. Second, these results speak only to enrollment, not persistence or attainment. Though I find that, on the margins, cost and distance did not affect the decision to enroll, these aspects of college-going may become more important as students work towards a degree. That only 42% of the 2004 cohort had obtained an associates 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, including student and family characteristics when modeling the choice of whether to enroll may produce parameters on school choice characteristics that do not reflect non-random spatial sorting of student populations across the country. In the face of persistent racial and economic segregation (de Souza Briggs and Wilson 2006), what does it mean that, controlling for socioeconomic characteristics, the characteristics of the most-likely or nearest public two-year college—specifically the distance from the student—are not statistically relevant in the decision of whether to enroll? 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 differ from each other in all observable 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.

In an online appendix, I explore this scenario by creating a synthetic population students that are meant to be representative of each census tract across the country. After assigning all students an equal SAT percentile (*e.g.*, everyone is at the 50th percentile), I perform the same two-step analytic procedure used above: (1) predict each student's most-likely college and (2) estimate the likelihood of enrollment based on characteristics of the school and the student. I repeat this process across a range of SAT percentiles to simulate different levels of college readiness. Briefly, simulations show that even when all students have the same level of academic preparation, the likelihood of attendance is not the same across the country. In some locations, students are more likely than not to attend even with SAT scores in the 30th percentile; in other locations, students are unlikely to attend even with SAT scores

in the 90th percentile. Students most unlikely to attend generally represent tracts that have more persons of color, lower educational attainment among the adult population, and lower median income than tracts in which the student almost always enrolls. Returning to the motivating thought experiment, the median distance between tracts unlikely to have their students attend college and the nearest public institution ranges from 8 to 12 miles. For tracts with students likely to attend, the distance is only 3–5 miles.[9]

A rejection, therefore, of the importance of college proximity 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.

Though students have many options when choosing where to go to college, results from this study suggest they face information gaps that limit their choice sets in practice. Because college-choice preferences are strongest in the application stage, when information about aspects of the decision (such as actual cost) is low, many students are likely to make suboptimal decisions. Simulations further demonstrate how the distribution of college opportunity is not equitable across the country and may further limit effective choice sets for underrepresented populations. Together, these findings have important ramifications for higher education policy.

Institutions, especially those with high prestige and selective admissions, may be missing talented applicants who do not apply. One reason may be that they overestimate the actual cost they will bear (Avery et al. 2006). While net price calculators and the College Scorecard website now offer students more information with which to make better informed application decisions, these tools rely on students knowing about them, seeking them out, and understanding how to use them. Though certainly more useful than not, these tools may work better in the context of targeted interventions that provide students with personalized information about their college options and which have shown promise in getting students to apply to a wider range of schools (Hoxby and Turner 2013). At the institutional level, admissions offices concerned with admitting a diverse class should work to expand their applicant pools rather than looking for talented students only among applicants. Offering better information about actual college costs to students before they apply, especially for students likely to quality for need- or merit-based aid, is one step.

At the federal level, aid policies are overly complex and too ill-timed to give students accurate information about their actual expected cost (Dynarski and Scott-Clayton 2013). The IRS Data Retrieval Tool can help reduce the time it takes to fill out the Free Application for Federal Student Aid (FAFSA). Using this tool in conjunction with an earlier submission date, students may be more likely to (1) complete the FAFSA and (2) do so early enough to provide useful information when applying for schools.[10] More direct supports are also warranted. Experimental results have shown that help filing the FAFSA increased the likelihood of FAFSA submission, enrollment, and aid receipt (Bettinger et al. 2012). Such that counselors and community organizations can provide hands-on help completing

---

[9] More detail about the procedure and accompanying figures can be found in the online appendix.

[10] The unannounced shutdown of the Data Retrieval Tool from March to October 2017 due to concerns over data security make the future reliability and potential usefulness of this tool less certain.

financial aid forms, the more students there will be who have accurate information about their real college costs and can make informed decisions.

To support students who have fewer choices due to living in "college deserts" (Hillman and Weichman 2016), state higher education systems and individual institutions with the capacity could choose to increase online offerings. Growth in online enrollment over the past decade demonstrates that distance education is a viable option for many students (Snyder et al. 2016). Lower average completion and pass rates for students in online courses, however, may dampen the benefits gained from choice sets that are digitally enlarged (Hart et al. 2016; Xu and Jaggars 2011, 2013; Bettinger et al. 2017; Huntington-Klein et al. 2017). For this reason, an expansion of online learning should be undertaken with consideration both of student access to technological resources and the quality of online course options as compared to face-to-face settings.

Conclusion

This paper consciously repeats the analytic approach of an earlier study (Long 2004) on the determinants of the college enrollment decision using a new cohort of students. As data for newer cohorts becomes available, researchers should continue to investigate the student preferences in the college enrollment decision. Since the early 2000s, the college student population has become increasingly diverse (Eagan et al. 2016) while real costs have continued to rise (Ma et al. 2016). Online course offerings and the number of for-profit institutions have also increased (Allen et al. 2016; Cottom 2017). These changes to the higher education landscape mean that more recent cohorts could have different college choice preferences that should be examined. Such that future analyses can reproduce the design of this and Long's study, longitudinal trends will be more readily identified and large-scale policies usefully evaluated.

Appendix

In this study, characteristics of the most-likely college (or nearest public two-year institution) were not consistently predictive of whether a student would enroll in college within two-years of high school graduation. In particular, the parameter on distance, which was negatively correlated with the choice between colleges and negatively signed across logistic models, was only statistically significant at the 5% level in two of eight models. Student-specific characteristics such as SAT percentile, family income, parental education, and gender as well as local unemployment rates were more consistently predictive of enrollment.

In the discussion section of the paper, I note that while these results suggest that distance was less important for students when choosing whether to attend college, it may also be the case that correlation between model covariates masks underlying realities about college access. If students are not sorted randomly in relation to postsecondary options, but
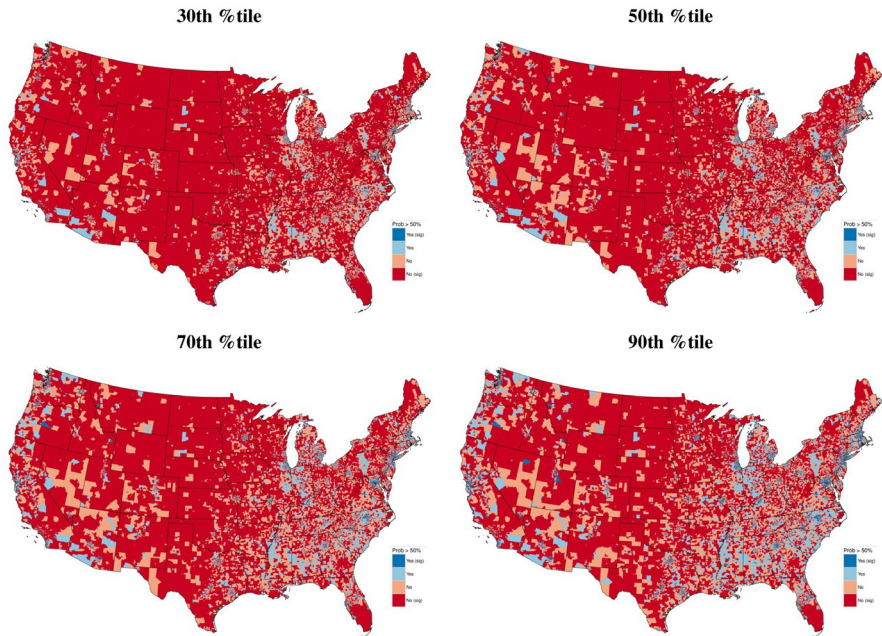
30th %tile

50th %tile

70th %tile

90th %tile

Fig. 1 Visualization of predicted probability that a 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 to attend); shades of red represent predicted probabilities less than 0.5 (less than likely to attend). 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 synthetic students given SAT scores at the stated point in the distribution: 30th, 50th, 70th, and 90th

instead are clustered based on observable characteristics such as family income, parental education level, and race/ethnicity, then it may be that controlling for these characteristics in a regression framework offers misleading insights into the relationship between distance and enrollment. To investigate whether distances to the closest public postsecondary institution are related to the likelihood of college enrollment, I proposed a thought experiment and method for carrying it out using synthetic data. I describe the procedure and results below.

I first use U.S. Census data to construct a population of synthetic students, each one representing a single census tract in the lower 48 states and created using its population characteristics. Specifically, synthetic students represent the "modal" person from the tract in that their characteristics are assigned using modal values of tract demographics. For example, in a tract in which women outnumber men, the student becomes female; in tracts where the plurality of residents have some college education, the student is assigned a parental education level of some college, and so on. Because Census data do not give SAT scores, I generate four 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
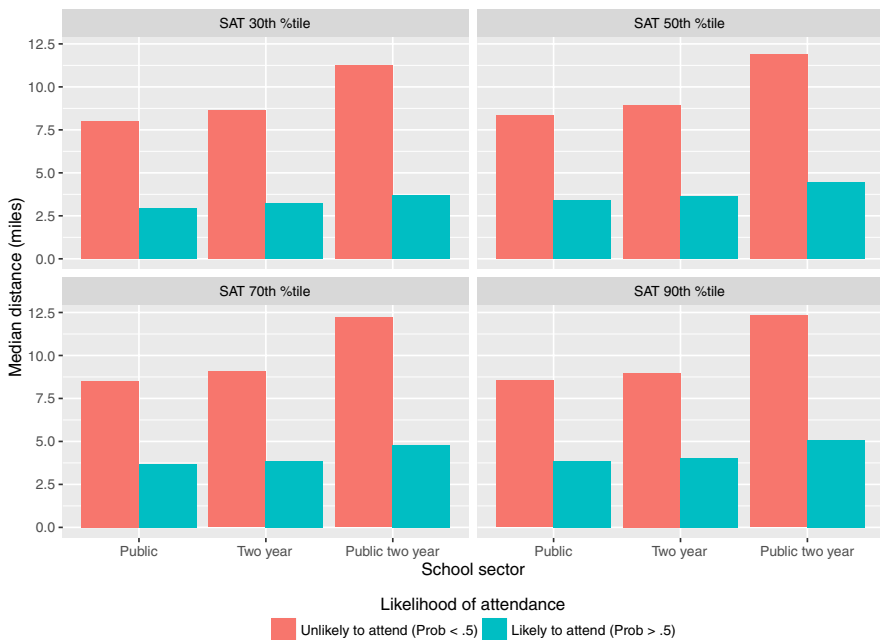
Fig. 2 Median distances in miles to nearest public, two-year, and public two-year for census tracts that produced modal synthetic students likely (Prob > 0.5) and unlikely (Prob < 0.5) to enroll within two years of high school graduation across distribution of SAT percentiles. Only those tracts with statistically significant positive or negative predictions ($p < 0.05$) are included

that they 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, shades census tracts according to the probability that its "modal" simulated student will enroll in college within two-years of high school graduation. Each tract 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 Fig. 1 show variation in the likelihood of enrollment, despite the fact that all synthetic students within each simulation are equally college ready in terms of their SAT scores. Simulation models suggest that even for synthetic students with SAT scores in the 90th percentile (bottom right map in Fig. 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 only those census tracts with probabilities of enrollment that are statistically significant, I find statistically significant differences in population characteristics between the bottom and top quartiles of predicted probability—those in which the synthetic student is least and most likely to enroll, respectively. Tracts in the bottom quartile of probable enrollment tend to have 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 similar across the range of SAT percentiles.

References

Allen, I. E., & Seaman, J. (2011). Going the Distance. Technical Report Babson Survey Research Group. http://www.babson.edu/Academics/centers/blank-center/global-research/Documents/going-the-distance.pdf.

Allen, I. E., & Seaman, J. (2013). Changing Course: Ten Years of Tracking Online Education in the United States. Technical Report. Babson Survey Research Group. https://files.eric.ed.gov/fulltext/ED541571.pdf.

Allen, I. E., et al. (2016). Online Report Card: Tracking Online Education in the United States. Technical Report. Babson Survey Research Group.

Avery, C., & Hoxby, C. M. (2004). Do and should financial aid packages affect students' college choices? In C. M. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 239–299). Chicago: University of Chicago Press.

Avery, C., 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.

Baker, D. J., & Doyle, W. R. (2017). Impact of community college student debt levels on credit accumulation. *The Annals of the American Academy of Political and Social Science*, *671*(1), 132–153. https://doi.org/10.1177/0002716217703043.

Baum, S., Little, K., & Payea, K. (2011). *Trends in community college education: Enrollment, prices, student aid, and debt levels. Trends in higher education series* (p. 11b-3741). New York: College Board.

Baum, S., & Ma, J. (2012). *Trends in college pricing.*, Trends in higher education series New York: The College Board.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). Chicago: University of Chicago Press.

Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education: Does college remediation work? *Journal of Human Resources*, *44*(3), 736–771.

Bettinger, E. P., et al. (2012). The role of application assistance and information in college decisions: Results from the H&R block FAFSA experiment. *The Quarterly Journal of Economics*, *127*(3), 1205–1242.

Bettinger, E. P., et al. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review, 107*(9), 2855–2875. https://doi.org/10.1257/aer.20151193. http://www.aeaweb.org/articles?id=10.1257/aer.20151193.

Betts, J. R., & McFarland, L. L. (1995). Safe port in a storm: The impact of labor market conditions on community college enrollments. *The Journal of Human Resources*, *30*(4), 741–765.

Black, D. A., & Smith, J. A. (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics*, *121*(1), 99–124.

Black, D. A., & Smith, J. A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics*, *24*(3), 701–728.

Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel & A. H. Halsey (Eds.), *Power and ideology* (p. 485). Oxford: Oxford University Press.

Bowen, W. 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. *Journal of the Royal Statistical Society, Series B (Methodological)*, *22*, 302–306.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics*, *3*, 1801–1863.

Cheng, S., & Long, S. J. (2007). Testing for IIA in the multinomial logit model. *Sociological Methods & Research*, *35*(4), 583–600.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, S95–S120.

Cottom, T. M. (2017). *Lower ed: The troubling rise of for-profit colleges in the new economy*. New York: The New Press.

de Souza Briggs, X., & Wilson, W. J. (2006). *The geography of opportunity: Race and housing choice in metropolitan America*., James A. Johnson Metro Series Washington, D.C: Brookings Institution Press.

Deming, D., & Dynarski, S. (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.

Deming, D. J., Goldin, C., & Katz, L. F. (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *The Journal of Economic Perspectives*, *26*(1), 139–164.

Doyle, W. R., & Skinner, B. T. (2016). Estimating the education-earnings equation using geographic variation. *Economics of Education Review*, *53*, 254–267.

Doyle, W. R., & Skinner, B. T. (2017). Does postsecondary education result in civic benefits? *The Journal of Higher Education*. https://doi.org/10.1080/00221546.2017.1291258.

Drewes, T., & Michael, C. (2006). How do students choose a university? An analysis of applications to universities in Ontario, Canada. *Research in Higher Education*, *47*(7), 781–800.

Dynarski, S. (2002). The behavioral and distributional implications of aid for college. *American Economic Review*, *92*(2), 279–285.

Dynarski, S., & Scott-Clayton, J. (2013). Financial aid policy: Lessons from research. Working Paper 18710. National Bureau of Economic Research.

Eagan, K., et al. (2016). *The American Freshman: Fifty-year trends 1966–2015*. Technical Report. Los Angeles: Higher Education Research Institute, UCLA.

Eide, E., Brewer, D. J., & Ehrenberg, R. G. (1998). Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance. *Economics of Education Review*, *17*(4), 371–376.

Fuller, W. C., Manski, C. F., & Wise, D. A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *The Journal of Human Resources*, *17*(4), 477–498.

Greene, W. H. (2012). *Econometric analysis* (7th ed.). Boston, MA: Prentice Hall.

Grubesic, T. H. (2008). Zip codes and spatial analysis: Problems and prospects. *Socio-Economic Planning Sciences*, *42*(2), 129–149.

Hart, C. M., Friedmann, E., & Hill, M. (2016). Online course-taking and student outcomes in California community colleges. *Education Finance and Policy*, *13*(1), 42–71.

Hillman, N., & Weichman, T. (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, C., & Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. In: Stanford Institute for Economic Policy Research Discussion Paper 12–014.

Hoxby, C. M., & Avery, C. (2012). The missing "One-offs": The hidden supply of high-achieving, low income students. Working Paper 18586. Cambridge, MA: National Bureau of Economic Research.

Huntington-Klein, N., Cowan, J., & Goldhaber, D. (2017). Selection into online community college courses and their effects on persistence. *Research in Higher Education, 58*(3), 244–269. https://doi.org/10.1007/s11162-016-9425-z.https://doi.org/10.1007/s11162-016-9425-z.

Jepsen, C., & Montgomery, M. (2009). Miles to go before I learn: The effect of travel distance on the mature person's choice of a community college. *Journal of Urban Economics*, *65*(1), 64–73.

Kane, T. 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, T. J. (1996). College cost, borrowing constraints and the timing of college entry. *Eastern Economic Journal*, *22*(2), 181–194.

Leslie, L. L., & Brinkman, P. T. (1987). Student price response in higher education: The student demand studies. *Journal of Higher Education*, *58*(2), 181–204.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*(420), 1227–1237.

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

Long, M. C. (2010). Changes in the returns to education and college quality. *Economics of Education Review*, *29*(3), 338–347.

Lovenheim, M. F., & Reynolds, C. L. (2011). Changes in postsecondary choices by ability and income: Evidence from the National Longitudinal Surveys of Youth. *Journal of Human Capital*, *5*(1), 70–109.

Ma, J., et al. (2016). *Trends in college pricing*., Trends in higher education series New York: The College Board.

Manski, C. F., & Wise, D. A. (1983). *College choice in America*. Cambridge, MA: Harvard University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

National Center for Education Statistics. (2016). ELS: Education Longitudinal Study of 2002.

Niu, S. X., & Tienda, M. (2008). Choosing colleges: Identifying and modeling choice sets. *Social Science Research*, *37*(2), 416–433.

Paulsen, M., & Smart, J. C. (2001). *The finance of higher education: Theory, research, policy, and practice*. New York: Algora Publishing.

Perna, L. W. (2006). Studying college choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 21, pp. 99–157). Dordrecht: Springer.

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

Rouse, C. E. (1995). Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics*, *13*(2), 217–224.

Snyder, T. D., de Brey, C., & Dillow, S. A. (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.

Snyder, T. D., Dillow, S. A., & Hoffman, C. M. (2007). *Digest of Education Statistics 2006. NCES 2007-017*. Washington, D.C: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Snyder, T. D., & Hoffman, C. M. (1995). *Digest of Education Statistics 1995. NCES 95-029*. Washington, D.C: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

StataCorp. (2015). *Stata statistical software: Release 14*. College Station, TX: StataCorp LP.

Tate, W. F. (2008). Geography of opportunity: Poverty, place, and educational outcomes. *Educational Researcher*, *37*(7), 397–411.

The Delta Cost Project. (2016). The Delta Cost Project Database.

Toutkoushian, R. K., & Paulsen, M. (2016). *Economics of higher education: Background, concepts, and applications*. Dordrecht: Springer.

Turner, S. E. (2004). Going to college and finishing college: Explaining different educational outcomes. In M. C. Hoxby (Ed.), *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 13–62). Chicago: University of Chicago Press.

Umbach, P. D. (2007). How effective are they? Exploring the impact of contingent faculty on undergraduate education. *The Review of Higher Education*, *30*(2), 91–123.

Xu, D., & Jaggars, S. S. (2011). The effectiveness of distance education across virginia's community colleges: Evidence from introductory college-level math and english courses. *Educational Evaluation and Policy Analysis*, *33*(3), 360–377.

Xu, D., & Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, *37*, 46–57.