



Can't get there from here: The decision to apply to a selective college

Amanda L. Griffith^{a,*}, Donna S. Rothstein^b

^a Cornell University, Department of Economics, 385 Ives East, Ithaca, NY 14853, United States

^b U.S. Bureau of Labor Statistics, 2 Massachusetts Avenue, NE, Suite 4945, Washington, DC 20212, United States

ARTICLE INFO

Article history:

Received 31 October 2007

Accepted 13 January 2009

JEL classification:

I21

Keywords:

Educational economics

School choice

ABSTRACT

In an attempt to increase applications from low-income students, some selective 4-year colleges are developing programs to target and attract low-income students. However, relatively little research has looked at factors important in the college application process, and in particular, how these factors differ for low-income students. This paper uses data from the National Longitudinal Survey of Youth 1997 (NLSY97) to analyze factors influencing students' college application decisions, with a focus on the decision to apply to a selective 4-year college. We hypothesize that distance from a student's home to selective colleges may play a role in the application decision and differentially impact low-income students. Our results suggest that distance does matter, although the effects do not vary by family income level.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

This paper examines the factors that are important during the college application process, with a specific focus on the roles of family income and proximity of selective schools on the decision of whether to apply to a selective college. It employs a very rich national longitudinal data set with sample members applying to college in the early 2000s.

Attending a selective 4-year college can impact relative lifetime earnings (see, for example, Brewer, Eide, & Ehrenberg, 1999; Long, 2008), and the earnings premium may be larger for students from low-income backgrounds (Behrman, Constantine, Kletzer, McPherson, & Schapiro, 1996; Dale & Krueger, 2002). Yet low-income students are under-represented at elite colleges and universities (Heller, 2004; Hill, Winston, & Boyd, 2005). Bowen, Kurzweil, and Tobin (2005), in their book *Equity and Excellence in American Higher Education*, show that only 11 percent of students from families in the bottom quartile of the income distribution are enrolled at the 19 elite colleges and universities

in their sample. They argue that increasing the representation of students from low-income families has significant benefits, both in terms of increased social mobility for the low-income students themselves, and in terms of increased socioeconomic diversity within colleges, which has benefits for all students enrolled.

Recently, a number of selective institutions have implemented programs aimed at increasing the representation of students from low-income families. These programs, developed by both public and private elite institutions such as Harvard, Yale, Princeton, the University of Virginia, and the University of North Carolina, among others, are varied in their approaches. However, most have at their core a promise to cover most or all of the school's tuition for students with low family income. Many programs also include attempts to increase awareness of the institution and the opportunities available there for low-income students. The hope is that the various measures will lead to a larger applicant pool of low-income students at selective institutions, resulting in their higher representation in the matriculating classes. Preliminary results from Harvard suggest that although effects of the program are modest thus far, it appears to be succeeding (Avery et al., 2006).

To ensure that these programs can successfully target low-income student populations, one needs to examine

* Corresponding author. Tel.: +1 607 254 4777; fax: +1 607 255 4496.
E-mail address: alg53@cornell.edu (A.L. Griffith).

why so few low-income students apply to more selective 4-year institutions. According to a 2005 study by Hill et al. (2005) there is a sizeable pool of high ability, low-income students in the U.S. (as measured by test scores and reported family income). Although the cost of attending a selective college or university can be quite high, tuition costs may not be the only hurdle that low-income students face. Proximity to post-secondary institutions could be important in students' college application decisions, and may be a more significant factor for low-income students. About 46 percent of the more elite institutions in the U.S. are located in Northeastern states, yet many of the low-income students that could attend these colleges are located in geographically distant states.¹ For example, in the data set we use in this paper, only 12 percent of students with a grade point average of 3.5 or more who are in families in the bottom quartile of income live in the Northeast.

College proximity can have two distinct effects on college application decisions. First, distance can impose costs and make students less likely to apply to colleges far away from their homes. For example, students may want to attend college (and therefore will apply to colleges) close to home for convenience, lower travel costs, and for the option of living at home to avoid paying for room and board. One might expect that financial reasons for attending a college closer to home may be more pressing for students from low-income families.

Second, living close to a selective 4-year college can expose students to what these colleges have to offer and encourage students to try to attend a selective 4-year college. Do (2004) refers to this as a spillover effect, which may be particularly influential for lower income students. For example, living close to a college may raise awareness of opportunities available at post-secondary institutions and help create a college-going expectation for nearby youths. Living near a selective institution could have an additional spillover effect, increasing the probability that students would strive to attend a selective college or university. Both effects suggest that as distance to a selective college increases, the less likely a student is to apply to one.

In this paper, we use the National Longitudinal Survey of Youth 1997 (NLSY97) to assess the roles of college proximity and family income in the decision to apply to a selective 4-year college. We analyze the college application decision using a bivariate probit model with selection. The application process is shown in two stages: (1) the choice to apply to a 4-year college (selection), and (2) the choice to apply to any selective 4-year college versus only non-selective 4-year colleges. The selection model yields some interesting findings. First youths from families with low-income are much less likely to apply to 4-year colleges than those from families with high-income, even after controlling for test scores, high school grade point average, and many other family, school, and location characteristics. Second, family income does not have an effect on the type of 4-year colleges to which students applied. And third, proximity to selective colleges does matter—students are less likely to

apply to a selective 4-year college the further they live from one.

2. Prior literature

Several studies have focused on the college application decision, but few have focused on the decision to apply to a selective college or university.² Recent studies have found mixed results on the importance of family income in the decision to apply to a selective college. Two studies focusing on applications to a specific institution find contrasting results. Desjardins, Dundar, and Hendel (1999) examine the decision to apply to a large, high-quality public university in the Midwest. Their findings indicate that students from low- and middle-income families are more likely to apply to the institution than students from high-income families. Weiler (1994) looks at the decision to apply to a specific selective private institution in a suburban location. He finds that as parental income increases, students are significantly more likely to apply to the focus institution. Toutkoushian (2001) looks specifically at the application decisions of high school seniors in New Hampshire, and finds that low levels of parental income do not discourage students from applying to the more selective schools in the sample.

In a very recent paper, Koffman and Tienda (2008) examine how a 1996 law in Texas (HB588), which guarantees admission to public colleges and universities in Texas to high school seniors graduating in the top 10 percent of their class, affected the distribution of socioeconomic status of the applicant pool to two Texas flagship public universities. They find that the admission policy did little to change the application rate of students from poor high schools. These results suggest that even with guaranteed admission there are still hurdles to overcome in order to increase the application rates of low-income students. To remedy the situation, the authors advocate increased, targeted recruitment of top students from poor high schools. But to do this, one must understand what factors are important to low-income students in their application decisions.

Very little research has looked at how student proximity to a college or university impacts his or her college application decisions.³ Turley (2009) is an exception. She uses NELS:88 data to examine how college proximity influences the probability of applying to a 2- or 4-year college. Turley measures college proximity as the number of colleges within commuting distance of a student's home (12 miles for urban youths and 24 miles for rural/suburban youths). She finds a very small increase in the probability of applying to a 4-year college associated with a 1-unit increase in the number of 4-year colleges in close proximity. These results suggest that college proximity influences the college application decision, but she does not look at college selectivity, which is the focus of the current paper.

² See Hossler et al. (1989) for a review of many of the earlier articles examining the college application decision.

³ However, note that Card (1995) uses proximity to a 4-year college as an instrument for years of schooling. He finds that students living closer to 4-year institutions, on average, attained higher total years of schooling.

¹ We define elite or selective 4-year colleges as those ranked by Barron's *Profile of American Colleges* (2001) as most or highly competitive.

Two recent studies find that college proximity affects the college enrollment decision, and may have a larger effect for students from families with low-income. Do (2004), using data from *High School and Beyond*, examines college matriculation decisions for low- versus high-income students. Results suggest that low-income students are more likely to attend a high-quality college if they live near a good public university, with mixed results for the impact of living near other types of elite institutions. Using a sample of Canadian high school seniors, Frenette (2006) finds that students who live further away from top universities are less likely to attend one and that the effect is significantly stronger for students from low-income backgrounds.

3. Empirical approach

In order to investigate factors affecting college application choices, this study estimates a bivariate probit selection model (see, for example, Maddala, 1983). The first (selection) stage is the decision of whether to apply to a 4-year college versus a 2-year or no college. The second stage, which is the main focus of our paper, is then whether to apply to any 4-year selective college versus only non-selective 4-year colleges. The second stage is censored, in that the outcome is only observed for those who choose to apply to a 4-year college.

Students have different portfolios of 4-year college applications, and we reduce them to a 1 (at least one selective 4-year college), 0 (only non-selective 4-year colleges) dependent variable in the second stage. Ideally, we would like to have a dependent variable that reflects the richness of the application choices. However, as we will see in the next section, only 237 students apply to any selective college in our data set, with 61 percent of these applying to only one selective college and another 22 percent applying to only two selective colleges.

We use two variables to identify the first stage of the bivariate probit selection model: distance to a 2-year college and the State unemployment rate.⁴ The theory behind the use of these variables to identify the selection process is as follows: as the unemployment rate rises, parents and their children may steer away from expensive 4-year colleges in favor of 2-year colleges. Two-year college proximity is likely to affect the 2-year versus 4-year college decision, but unlikely to have an effect on a student's decision to apply to a selective versus a non-selective 4-year institution. In practice, we find that these two variables are statistically insignificant in the second stage.

Explanatory variables that may influence the college application decision are broken into five descriptive categories:

- (i) personal—gender, race, ethnicity, and test score;
- (ii) family—income, parent education, family structure, and family size;

- (iii) high school—type of institution, racial composition, and percent of low-income students;
- (iv) location—median income, urbanicity, region;
- (v) distance—distance to college type.

The rich data set used in this paper allows one to control for such an exhaustive list of characteristics. The hope is that the effects of these variables can shed light on the college application decision, particularly the roles that college proximity and family income play in the process.

4. Data and variables

This paper employs the NLSY97 to study the college application decisions of youths in the U.S. The NLSY97 consists of nearly 9000 youths who were born in the years 1980–1984. The youths were 12–17 when first interviewed in 1997, and have had annual in-person interviews ever since. In 2003 (round 7), the NLSY97 added a section on college choice for youths born in the years 1983 and 1984. Youths who attended at least 12th grade or received a GED report the colleges applied to in each application cycle, among other information.⁵ The survey repeated the section for the same two birth years in 2004. The NLSY97 geocode CD and confidential data available to researchers who come to the U.S. Bureau of Labor Statistics provide college UnitIDs, high school id codes, as well as residential zip code, county, and state for each survey year. The paper merges in high school characteristics from the Q.E.D. (Quality Education Data) and uses county of high school residence to link to information from the 2000 edition of the *County and City Data Book* (U.S. Department of Commerce, 2001).

The study obtains parent reports of household income from the round 1 NLSY97 parent questionnaire.⁶ Family structure, household size, and biological mother's education are also from round 1. The NLSY97 defines race and ethnicity as three mutually exclusive groups: non-black and non-Hispanic, black and non-Hispanic, and Hispanic. The survey oversamples the latter two groups. ASVAB (Armed Services Vocational Aptitude Battery) test scores are available for about 80 percent of the NLSY97 sample. From the summer of 1997 through the spring of 1998, NLSY97 youths took the computer-adaptive version of the ASVAB. Four of the subtests combine to form a composite measure of math and verbal aptitude. This aptitude measure is similar to the Department of Defense's Armed Forces Qualification Test (AFQT) score available in the NLSY79. NLSY97 survey personnel internally normed these tests and created the composite math and verbal aptitude percentile score (0 (lowest) to 99) provided in the NLSY97 data set.

The paper uses college UnitIDs to merge in information on selectivity and other characteristics of colleges from

⁴ State unemployment rate is from Table 572 of the 2001 edition of the *Statistical Abstract of the United States* (U.S. Bureau of the Census, 2002).

⁵ The sample is limited to those with a high school diploma or GED in the analysis that follows.

⁶ Household income is missing for 25 percent of the youths in the NLSY97 sample, with about half due to a missing parent interview. Descriptive statistics are shown for non-missing observations. In the analyses that follow, variables with missing observations are given a value of zero, and a dummy variable for the missing variable is included in the regression.

the 2000 IPEDS and the College Board's Annual Survey of Colleges.⁷ The rankings in *Barron's Profile of American Colleges* (2001) are used to define selective colleges (those with a most or highly competitive ranking). In 2000, 146 U.S. colleges are considered to be selective by the Barron's rankings. The average median SAT score is 1272 for the sample of selective institutions.

Finally, the study uses zip codes from a student's senior year of high school and the complete list of selective 4-year, non-selective 4-year, and 2-year colleges to create measures of college proximity.⁸ U.S. Gazetteer files from the U.S. Census Bureau provide the latitude and longitude of the centroid of each zip code in decimal degrees. One can then convert decimal degrees to radians, and then calculate the distance in miles between the zip code of student i and the zip code of school j .⁹ Once this is completed for all student and school zip code combinations, we use the minimum distance to each school type to define college proximity.

5. Descriptive statistics

Table 1 displays descriptive statistics for the NLSY97 sample, separated by type of school application and selectivity. The categories are: did not apply, 2-year college, any 4-year college, non-selective 4-year college, and selective 4-year college. The first three categories are mutually exclusive, and the last two are mutually exclusive subsets of the third column. Students who apply to multiple types of school are placed in the highest category (for example, applications to both a non-selective 4-year college and a selective 4-year college place the student in the selective category). On average, students who apply to at least one selective college, apply to 3.5 4-year colleges; the number is lower, 1.7, for those who apply to non-selective 4-year colleges. About 95 percent of students who apply to only non-selective 4-year colleges are accepted. The number is much lower for the selective category (73 percent), reflecting the increased difficulty of acceptance at more selective 4-year colleges.

Table 1 indicates an under-representation of students from low-income families in the selective college applicant pool. We divide household income into approximate quartiles based on Current Population Survey data of households with a 12–17-year old youth present.¹⁰ Only about 14 percent of students in families with income under \$25,000 apply to a selective 4-year college, compared to over 46 percent from families with income of at least \$70,000. Note that low-income students make up the

largest share of those who apply to no college or apply to a 2-year college only.

A number of authors (for example, Bowen et al., 2005) have noted the under-representation of low-income, high ability students in the pool of students who apply to selective schools. We find this in the NLSY97 as well. Suppose we limit our sample to those who scored in the 75th percentile or better on the math/verbal portion of the ASVAB. Of these high-scoring youths who are from families in the bottom two quartiles of income, 65 percent apply to a 4-year college, and of those, 25 percent apply to a selective 4-year college. In contrast, of high-scoring youths who are in the top quartile of family income, 85 percent apply to a 4-year college, and of those, 40 percent apply to a selective 4-year college. Thus we see income differentials in the first stage decision – whether to apply to a 4-year college at all – as well as in the second stage decision – whether to apply to a selective 4-year college.

Table 1 shows that, on average, youths who apply to a selective college have a significantly shorter distance to a 4-year selective college than youths who apply to non-selective 4-year colleges: 54 miles versus 98 miles. On average, students live significantly further from selective colleges than non-selective 4- and 2-year colleges. This differential reflects the small number of selective colleges and universities in the U.S. and their geographic distribution. For example, of the 146 colleges rated as selective, 46 percent are in the Northeast and about 57 percent are in the Northeast and California. To put this in perspective with respect to the NLSY97 sample, youths in the Northeast live less than 19 miles from a selective college, on average, but those who live in the South or Midwest average about 95 miles, and those in the West average about 149 miles.

The difference in average distance to a selective college also varies by income level. For example, in the NLSY97 sample, students from families in the lowest quartile of income live an average of 95 miles from a selective college, but students from families in the highest quartile of income live an average of 87 miles from a selective college. If instead, we look at a measure of whether a student has a selective college within 50 miles of his or her high school residence, we find that 51 percent of students from low-income households have a selective college in close proximity compared to 63 percent of students from high-income households.

6. Results

Marginal effects from independent probits and a bivariate probit with selection for the probability of applying to any 4-year college and the probability of applying to a selective versus non-selective 4-year college are shown in Table 2. The results from the two sets of equations are similar. A Wald test cannot reject the null hypothesis that the equations are independent ($p = .17$). Both variables used as exclusion restrictions are significant in the first stage. Given the richness of the NLSY97 data, we are able to account for much of the heterogeneity between youths through covariates.

Family income has a large effect in the first stage, but no effect in the second stage. This suggests that, all else

⁷ We exclude for-profit colleges in our analysis. At the time of the survey, for-profit schools were a very small part of the college application set. We delete 27 observations in which the respondent only applied to for-profit 4-year colleges. Note that only 9 students who apply to public and private 4-year colleges in our final sample also apply to a for-profit college.

⁸ We exclude for-profit colleges from these measures.

⁹ The formula is $4000 \times \arccos(\sin(\text{school}_j \text{ latitude}) \times \sin(\text{student}_i \text{ latitude}) + \cos(\text{school}_j \text{ latitude}) \times \cos(\text{student}_i \text{ latitude}) \times (\cos(\text{school}_j \text{ longitude} - \text{student}_i \text{ longitude})))$.

¹⁰ See U.S. Census Bureau (1997). Note that the NLSY97 round 1 parent interview asked parents to report income from calendar year 1996.

Table 1
Descriptive statistics, by type of college application.

	Did not apply	2-year college	Any 4-year college	Non-selective 4-year college	Selective 4-year college
Number of 4-year college applications	.000	.000	2.101 (1.541)	1.738 (1.182)	3.511 (1.919)
<i>Selective college</i>					
Apply to any selective	.000	.000	.205	.000	1.000
Accepted if applied	–	–	.733	–	.733
<i>Non-selective college</i>					
Apply to any	.000	.000	.942	1.000	.717
Accepted if applied	–	–	.961	.954	1.000
<i>Distance variables</i>					
Distance to selective 4-year college	105.033 (186.407)	86.934 (125.703)	88.833 (148.251)	97.747 (158.855)	54.192 (88.663)
Distance to non-selective 4-year college	10.600 (13.178)	10.861 (13.359)	9.658 (16.361)	10.190 (17.425)	7.591 (11.106)
Distance to 2-year college	12.463 (19.603)	9.633 (12.732)	11.544 (15.291)	12.381 (16.386)	8.290 (9.286)
<i>Personal characteristics</i>					
Female	.454	.502	.548	.556	.519
Black	.256	.244	.233	.255	.148
Hispanic	.249	.249	.124	.129	.105
Math/verbal percentile score	36.517 (25.234)	41.480 (24.185)	64.960 (25.358)	61.229 (24.995)	78.643 (21.768)
High school grade point avg.	2.648 (.733)	2.792 (.664)	3.295 (.601)	3.218 (.606)	3.594 (.477)
<i>Family characteristics</i>					
Income < \$25,000	.405	.327	.178	.189	.140
Income ≥ \$25,000 and < \$45,000	.252	.293	.233	.251	.167
Income ≥ \$45,000 and < \$70,000	.208	.233	.261	.269	.231
Income ≥ \$70,000	.135	.147	.327	.291	.462
Biological mother's years of education	11.844 (2.539)	12.285 (2.875)	13.922 (2.797)	13.743 (2.739)	14.615 (2.915)
Family size	4.694 (1.589)	4.584 (1.485)	4.437 (1.337)	4.457 (1.370)	4.359 (1.198)
Both biological parents	.469	.518	.651	.623	.759
Biological parent + step	.166	.141	.100	.109	.068
Biological mother only	.288	.275	.202	.218	.139
Biological father only	.032	.029	.022	.024	.017
<i>High school characteristics</i>					
Private	.017	.012	.028	.018	.065
Catholic	.009	.030	.074	.062	.120
Log (school size)	6.945 (.836)	7.012 (.774)	7.004 (.674)	7.000 (.670)	7.022 (.688)
Student/teacher ratio	17.026 (5.861)	17.520 (5.963)	16.583 (5.152)	16.518 (4.892)	16.846 (6.088)
% Black	20.856 (27.090)	19.627 (26.532)	21.342 (29.164)	22.079 (30.167)	18.011 (23.910)
% Hispanic	17.335 (25.016)	16.472 (23.763)	10.573 (19.453)	10.284 (19.231)	11.875 (20.432)
% Chapter I	25.632 (19.050)	25.308 (19.608)	21.827 (19.611)	22.824 (19.959)	17.324 (17.301)
<i>Location characteristics</i>					
Urban	.758	.773	.767	.746	.850
Log median income in county	10.486 (.227)	10.493 (.222)	10.531 (.258)	10.512 (.259)	10.605 (.236)
Midwest	.218	.216	.254	.266	.207
West	.261	.304	.177	.172	.198
South	.359	.348	.370	.371	.367
State unemployment rate	4.065 (.843)	4.111 (.772)	3.890 (.832)	3.896 (.842)	3.866 (.790)
N	965	546	1158	921	237

Note: Means, standard deviations in parentheses. Monetary values are in constant 1996 dollars. Distance is in miles. Means exclude any missing observations. Selective 4-year college is defined as having a *Barron's Profile of American Colleges* (2001) rating of most or highly competitive.

Table 2

Four-year college versus 2-year or no college application and selective 4-year college versus non-selective 4-year college application decisions, marginal effects from probits and bivariate probit with selection.

	Probability of applying to a			
	4-year college (probit)	Selective college (probit)	4-year college (bivariate probit)	Selective college
<i>Distance variables</i>				
Distance to selective 4-year college	-.010 (.010) ^a	-.047 (.012) ^{***,a}	-.001 (.001) ^b	-.026 (.008) ^{***,a}
Distance to non-selective 4-year college	-.001 (.001)	.001 (.001)	-.001 (.001)	.001 (.001)
Distance to 2-year college	.002 (.001) ^{**}	–	.002 (.001) ^{**}	–
<i>Personal characteristics</i>				
Female	.046 (.022) ^{**}	-.033 (.023)	.044 (.022) ^{**}	-.020 (.014)
Black	.163 (.039) ^{***}	-.026 (.038)	.164 (.039) ^{***}	-.021 (.022)
Hispanic	.030 (.039)	-.047 (.033)	.027 (.039)	-.030 (.018) [*]
Math/verbal percentile score	.067 (.001) ^{***,b}	.031 (.006) ^{***,b}	.068 (.005) ^{***,b}	.016 (.004) ^{***,b}
High school grade point average	.246 (.017) ^{***}	.150 (.023) ^{***}	.248 (.017) ^{***}	.078 (.012) ^{***}
<i>Family characteristics</i>				
Income < \$25,000	-.156 (.038) ^{***}	.049 (.050)	-.156 (.038) ^{***}	.038 (.034)
Income ≥ \$25,000 and < \$45,000	-.093 (.036) ^{***}	-.006 (.034)	-.092 (.036) ^{***}	.002 (.021)
Income ≥ \$45,000 and < \$70,000	-.069 (.035) ^{**}	-.041 (.028)	-.067 (.035) [*]	-.021 (.016)
Biological mother's years of education	.033 (.005) ^{***}	.003 (.005)	.034 (.005) ^{***}	.001 (.003)
Both biological parents	.140 (.032) ^{***}	.042 (.037)	.141 (.032) ^{***}	.021 (.023)
Biological mother only	.077 (.042) [*]	.014 (.053)	.081 (.041) ^{**}	.005 (.031)
Biological father only	.048 (.072)	-.062 (.063)	.055 (.072)	-.035 (.031)
<i>High school characteristics</i>				
Private	.150 (.088) [*]	.256 (.109) ^{**}	.159 (.087) [*]	.163 (.089) [*]
Catholic	.311 (.062) ^{***}	.093 (.056) [*]	.311 (.062) ^{***}	.043 (.034)
Log (school size)	.046 (.019) ^{**}	-.022 (.023)	.049 (.019) ^{**}	-.018 (.014)
Student/teacher ratio	.001 (.002)	.002 (.002)	.001 (.003)	.001 (.002)
% Black	.001 (.001)	.002 ^{**} (.001)	.001 (.001)	.010 (.005) ^{**} , ^b
% Hispanic	-.001 (.001)	.003 (.001) ^{***}	-.001 (.001)	.016 (.006) ^{***} , ^b
% Chapter I	.003 (.001) ^{***}	-.002 (.001) ^{**}	.003 (.001) ^{***}	-.0015 (.007) ^{**} , ^b
<i>Location characteristics</i>				
Urban	.003 (.030)	.064 (.024) ^{***}	.006 (.030)	.036 (.014) ^{**}
Log median income in county	.062 (.058)	.096 (.055) [*]	.051 (.058)	.063 (.033) [*]
Midwest	-.061 (.036) [*]	-.055 (.029) [*]	-.066 (.036) [*]	-.028 (.017)
West	-.146 (.040) ^{***}	.008 (.041)	-.144 (.040) ^{***}	.005 (.025)
South	-.064 (.034) [*]	.014 (.033)	-.064 (.034) [*]	.013 (.021)
State unemployment rate	-.040 (.016) ^{**}	–	-.036 (.015) ^{**}	–
Rho (ρ)	–	–	.998 (.009)	–
Log likelihood	-1305.921	-474.027	-1777.749	–
N	2669	1158	2669	–

Robust standard errors in parentheses. Conditional marginal effects are shown in the right-most column.

^a Marginal effect and standard error multiplied by 100.

^b Marginal effect and standard error multiplied by 10.

^{*} Indicate significance at 10% level.

^{**} Indicate significance at 5% level.

^{***} Indicate significance at 1% level.

equal (including test scores and high school grade point average), students with lower family incomes are less likely to be in the pool of applicants who apply to any type of 4-year college. However, given they get past this hurdle, low-income students are not any less likely to apply to a selective college. Relative to the highest income group, youths in families in the lowest income quartile are about 16 percentage points less likely to apply to any 4-year college, those in the next lowest income quartile are about 9 percentage points less likely to apply, and those in the second to highest income quartile are about 7 percentage points less likely to apply. Policies that aim to increase the applicant pool of low-income students at elite 4-year colleges may have to take into account that a number of these students are not applying to any 4-year college.

A number of background characteristics are significantly associated with the likelihood of applying to a selective college. For example, students with higher aptitude, as measured by test scores, are significantly more likely to apply to a selective 4-year college, relative to a non-selective 4-year college. High school grade point average has the same type of effect. High school environment can play a major role in influencing students' college choice, by preparing students academically and also possibly by providing information about the availability of opportunities at more selective institutions. Students who attend private high schools are significantly more likely to apply to a selective college by about 16 percentage points. Students from low-income families are much more likely to attend public high schools than their higher income peers.

Table 3

Alternative selective 4-year college proximity measures: selective 4-year college versus non-selective 4-year college application decision, marginal effects from probits.

	Descriptive statistics, by type of college application			Probit
	4-year college	Non-selective 4-year college	Selective 4-year college	Pr (apply to selective 4-year college)
Specification 1: Have selective 4-year college within a 50-mile radius	.592	.559	.717	.059 (.026)**
Specification 2: Have one selective 4-year college within a 50-mile radius	.205	.193	.249	.056 (.036)
Have two selective 4-year colleges within a 50-mile radius	.098	.091	.122	.096 (.050)*
Have three or more selective 4-year colleges within a 50-mile radius	.289	.275	.346	.057 (.039)
Specification 3: Number of selective 4-year college slots within a 50-mile radius/1000	3.147 (4.480)	2.920 (4.360)	4.027 (4.825)	.008 (.004)*
Specification 4: Have selective 4-year college within a 50-mile radius	.592	.559	.717	.090 (.030)***
Have selective 4-year college within a 100-mile radius, but not a 50-mile radius	.139	.142	.127	.085 (.047)*
Specification 5: Have selective 4-year college within a 100-mile radius	.731	.701	.844	.080 (.025)***
Specification 6: Number of selective 4-year college slots within a 100-mile radius/1000	6.852 (7.943)	6.435 (7.846)	8.474 (8.126)	.009 (.003)***
N	1158	921	237	1158

Note: Means, standard deviations in parentheses in first three columns. Marginal effects with robust standard errors in parentheses in last column. Specifications include controls for personal, family, high school, and location characteristics.

* Indicate significance at 10% level.

** Indicate significance at 5% level.

*** Indicate significance at 1% level.

Longer distances to a selective college decrease the probability of applying to one. The marginal effects from the bivariate probit suggest that a 75-mile increase in distance to a selective college (about half a standard deviation for the 4-year application group) decreases the likelihood of applying to one by about 2 percentage points. When we interacted distance with family income quartiles, we found that the distance effect does not vary with family income. To get a feel for the magnitude of the distance effect, note that a .3-point increase in high school grade point average (about half a standard deviation for the 4-year application group) increases the likelihood of applying to a 4-year selective college by about 2.3 percentage points. Distance to the nearest non-selective college does not have a significant effect in either stage. In addition, the distance to a 2-year college has a positive effect in the first-stage. As the distance to the nearest 2-year college increases, the likelihood of applying to a 4-year college increases.

To check the robustness of our key result regarding proximity to a selective college we create a number of alternative measures. The first set is for a 50-mile radius of the student's high school zip code: (1) the presence of a selective college within a 50-mile radius of the student's high school zip code, (2) a series of mutually exclusive dummy variables that describe whether the respondent has 1, 2, or 3 or more selective 4-year colleges within a 50-mile radius, and (3) the number of freshman slots at selective schools within the 50-mile radius divided by 1000. The second set includes two mutually exclusive dummy variables for the presence of a selective college within a 50-mile

radius and presence of a selective college between a 51- and 100-mile radius, as well as (1) and (3) above defined for a 100-mile radius rather than 50. In addition to verifying that our results are robust, these alternative measures allow us to examine whether the effect of college proximity is non-linear. On the one hand, students may only require one selective college within a certain radius to increase the likelihood that they will apply to one. On the other hand, an increase in the number of selective colleges in close proximity, and freshman slots at these colleges, may further raise the probability of applying to one.

Table 3 shows college proximity marginal effects for selective colleges from six different probits of the probability of applying to any selective 4-year college versus only non-selective 4-year colleges.¹¹ The probits control for all of the background variables included in the second stage estimates in Table 2. The results in Table 3 are very similar to those in Table 2: close proximity to a selective college raises the likelihood of applying to one. For example, having a selective college within a 50-mile radius increases the likelihood of applying to one by about 6 percentage points. Although it looks as though moving from having one to two selective colleges within a 50-mile radius increases the likelihood of applying to one, these two effects

¹¹ Given our prior finding that we could not reject the null hypothesis that the first and second stage equations are independent, we estimate probits only. However, the results are very similar when we estimate bivariate probits.

are not statistically different. The effect of the slots measure is modest—an increase in selective freshman slots within a 50-mile radius by 1500 (about one third of a standard deviation), increases the likelihood of applying to a selective college by 1.2 percentage points. Interestingly, the effects of having a selective college within a 50-mile radius (potentially commuting distance) versus only one within a 100-mile radius are not statistically different. In addition, the results from the last two specifications are similar to those that used the 50-mile radius. The earlier finding of a negative and significant effect of distance to a selective college on the probability of applying to one appears robust to alternative proximity measures.

This paper has suggested two hypotheses about the effects of distance: (1) distance imposes costs and makes students less likely to apply to colleges far away and (2) living close to a selective 4-year college exposes students to what this type of college has to offer and encourages students to try to attend a selective 4-year college. It is difficult to differentiate between the two stories because both suggest the same sign in the college application equations: as distance to a selective college increases, the less likely a student is to apply to one.

A pure distance cost story would suggest that students who live in close proximity to a selective college would apply to that one, rather than a selective college farther away. This does not appear to be the case. About 72 percent of students who applied to a selective college in the NLSY97 lived within 50 miles of a selective institution. However, of these students, only 38 percent applied to the closest (give or take 25 miles), and in fact on average applied to institutions much further away. We would expect the cost story to be more binding for lower income students, but the number is very similar (39 percent) for students from families in the bottom two quartiles of income. The probit results using alternative measures of college proximity, shown in Table 3, also suggest that a true cost story may not fully explain the importance of distance. For example, a pure distance cost story would suggest that having a selective college within 50 miles would increase the probability of applying to one more so than having one between 51 and 100 miles. However, in specification (4) we cannot reject the null hypothesis that the effects are the same. These findings suggest that distance costs alone do not appear to be driving the results for the most selective colleges.

Our results fit well with those of past studies. We estimate a bivariate probit with selection to attempt to tease apart the impacts of factors affecting the decision to apply to a 4-year college, and the decision to apply to any selective 4-year college versus only non-selective 4-year institutions. After controlling for a wealth of covariates and selection, we find that family income on its own is not a deterrent to applying to a selective college. These results are very similar to the findings of [Toutkoushian \(2001\)](#). Our results build on [Turley's 2009](#) finding that college proximity impacts the likelihood of applying to a 4-year college. We find that selective college proximity also impacts the likelihood of applying to a 4-year selective college. We do not, however, find that proximity effects differ by family income, as found by [Do \(2004\)](#) and [Frenette \(2006\)](#) for the selective college enrollment decision.

Of course, we would be remiss not to mention that the application stage is only the first part of the puzzle of how to increase the representation of lower income students at selective colleges. Students must be accepted at a selective college and then ultimately enroll. Do low-income students have similar acceptance rates to their high-income peers? [Bowen et al. \(2005\)](#) suggest that they do within their sample of 19 selective colleges and universities. We estimate some simple probits to see whether income levels are related to the likelihood of being accepted at a 4-year selective college, given the student applied to at least one. We control for all of the same variables in the selective college choice equations shown in Table 2, including test score and high school grade point average.

Our results indicate that students in the lowest income quartile are about 26 percentage points less likely to be accepted at a selective 4-year college relative to their peers in the highest income quartile. The addition of a control for the number of selective college applications the student submitted causes the marginal effect to increase. Clearly, we have a selection problem, in that we are conditioning our sample on students who apply to selective colleges. In addition, the number of low-income students who apply to selective schools in the NLSY97 sample is low. However, these results provide some suggestive evidence that even once low-income students get past the hurdle of applying to at least one selective college, something that very few low-income students do, they are less likely to be accepted. Perhaps this is due to non-need-blind admissions practices at some colleges. If a student is on the margin for acceptance, perhaps some colleges take into account the amount of funding each student would require to enroll. If this is occurring, lower income students may be at a disadvantage at the acceptance stage. Research using a larger data set could shed more light on this issue. But the results here point to another potential reason for the under-representation of low-income students at selective colleges—lower acceptance rates.

7. Conclusion

Low-income students are under-represented at selective 4-year colleges and universities. Remedying this problem could potentially increase social mobility for low-income students as well as boost socioeconomic diversity within colleges. A number of selective institutions have implemented programs aimed at increasing the representation of students from low-income families. Most of these programs involve the lowering or elimination of tuition costs for qualified low-income students. However, tuition costs of college may not be the only hurdle facing these students. About 46 percent of the more selective institutions in the United States are located in Northeastern states, and many of the low-income students that could attend these colleges are located in geographically distant states. Physical distance from a selective college may be an important issue to low-income students for both financial and non-monetary reasons, such as convenience, travel costs, and the option of living at home. A nearby college or university may also provide spillover effects by raising awareness of

opportunities available at colleges and creating a college-going expectation for nearby youths.

This paper uses data from the NLSY97 to analyze the relationship between various personal, family, school, and geographic background characteristics and the likelihood of applying to a selective 4-year college or university. Particular attention is paid to the influence of distance to selective colleges, and whether the effect differs for low-income students. Basic means show that students who apply to selective 4-year colleges live almost half as far from an elite institution as students who apply to non-selective 4-year colleges. In addition, a lower proportion of students from families in the lower two quartiles of income apply to a selective 4-year college compared to a non-selective 4-year college.

We estimate a bivariate probit with selection. The first stage (selection) is the probability of applying to any 4-year college, and the second stage is the probability of applying to any 4-year selective college versus only non-selective 4-year colleges. The results suggest that lower income students are much less likely to apply to any 4-year college. However, income does not appear to impact the likelihood of applying to a selective college. Distance to a selective 4-year college has a significant impact on the probability that a student will apply to a selective school. As the distance to the closest selective college increases, students are less likely to apply to this type of college, all else equal. Low-income students do not seem to be any more sensitive to distance than their high-income peers.

The findings from this paper can provide some suggestions for selective colleges that are trying to increase their representation of students from low-income families. First, a number of high-test score students from low-income families are not applying to any 4-year colleges. As advocated by Koffman and Tienda (2008), increased recruitment of top students from poor high schools may be productive. Second, the geographic mismatch of low-income students and selective institutions appears to be a factor in the college application process. Helping low-income students with travel costs and logistics may increase a school's attractiveness. In addition, educating students in lower income and education areas that are far from any selective colleges about the opportunities available at selective colleges may be fruitful. Finally, it appears that, all else equal, low-income applicants to selective colleges are less likely to be accepted. Future research that examines the acceptance decisions in more detail, as well as factors that affect enrollment behavior, can provide further insight into how to increase the representation of students from low-income families at selective 4-year colleges and universities.

Acknowledgements

Griffith's research was supported by a National Science Foundation Graduate Research Fellowship and the Cornell Higher Education Research Institute. The views expressed

are those of the authors alone and do not reflect the policies of the BLS or the views of other BLS staff members. We thank Alison Aughinbaugh, Ronald Ehrenberg, George Jakobson, Charles Pierret, seminar participants at the AEA and SOLE 2007 annual meetings, and an anonymous referee for helpful comments.

References

- Avery, C., Hoxby, C., Jackson, C., Burek, K., Poppe, G., & Raman, M. (2006). Cost should be no barrier: An evaluation of the first year of Harvard's Financial Aid Initiative. *N.B.E.R. Working Paper No. 12029*.
- Behrman, J., Constantine, J., Kletzer, L., McPherson, M., & Schapiro, M. (1996). The impact of college quality on wages: Are there differences among demographic groups? *Williams Project on the Economics of Higher Education, Discussion Paper No. 38*.
- Bowen, W. G., Kurzweil, M. A., & Tobin, E. M. (2005). *Equity and excellence in American higher education*. Charlottesville: University of Virginia Press.
- Brewer, D. J., Eide, E., & Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college quality on earnings. *Journal of Human Resources*, 34(1), 104–123.
- Card, D. (1995). The causal effect of education on earnings. In A. Orley, & D. Card (Eds.), *Handbook of labor economics*. Amsterdam: North Holland, pp. 1801–1863.
- College Board. (2000–2001). *Annual Survey of the Colleges of the College Board and Data Base*.
- College Division of Barron's Educational Series. (2001). *Barron's Profile of American Colleges*. 24th ed. Hauppauge, NY: Barron's Educational Series.
- Dale, S. B., & Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics*, 117(4), 1491–1527.
- Desjardins, S. L., Dundar, H., & Hendel, D. D. (1999). Modeling the college application decision process in a land-grant university. *Economics of Education Review*, 18(1), 117–132.
- Do, C. (2004). The effects of local colleges on the quality of college attended. *Economics of Education Review*, 23(3), 249–257.
- Frenette, M. (2006). Too far to go on? Distance to school and university participation. *Education Economics*, 14(1), 31–58.
- Heller, D. (2004). Pell grant recipients in selected colleges and universities. In D. Richard, & Kallenberg (Eds.), *America's untapped resources: Low-income students in higher education* (pp. 157–166). New York: Century Foundation Press.
- Hill, C., Winston, G., & Boyd, S. (2005). Affordability: Family incomes and net prices at highly selective private colleges and universities. *Journal of Human Resources*, 40(4), 769–790.
- Hossler, D., Braxton, J., & Coppersmith, G. (1989). Understanding student college choice. *Higher education: Theory and research*. New York: Agathon Press, pp. 231–288.
- Koffman, D., & Tienda, M. (2008). *Missing in application: The Texas Top 10% Law and socioeconomic diversity*. Unpublished working paper. Princeton University.
- Long, M. C. (2008). College quality and early adult outcomes. *Economics of Education Review*, 27(5), 588–602.
- Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- Toutkoushian, R. K. (2001). Do parental income and educational attainment affect the initial choices of New Hampshire's college-bound students? *Economics of Education Review*, 20(3), 245–262.
- Turley, R. (2009). College proximity: Mapping access to opportunity. *Sociology of Education*, 82(2), 126–146.
- U.S. Bureau of the Census. (1997). *Current Population Reports, P60-197, Money Income in the United States: 1996 (With Separate Data on Valuation of Noncash Benefits)*. U.S. Government Printing Office: Washington, D.C.
- U.S. Bureau of the Census. (2002). *Statistical Abstract of the United States, 2001*. U.S. Government Printing Office: Washington, D.C.
- U.S. Department of Commerce. (2001). *County and City Data Book, 2000*. U.S. Government Printing Office: Washington, D.C.
- Weiler, W. (1994). Transition from consideration of a college to the decision to apply. *Research in Higher Education*, 35(6), 631–646.