Available online at www.sciencedirect.com
SCIENCE DIRECT®

JOURNAL OF Econometrics

Journal of Econometrics 121 (2004) 271–296

www.elsevier.com/locate/econbase

How have college decisions changed over time? An application of the conditional logistic choice model

Bridget Terry Long*

465 Gutman Library, Harvard Graduate School of Education, Appian Way, Cambridge, MA 02138, USA

Abstract

This paper examines how individuals from 1972, 1982, and 1992 chose whether and where to attend college by estimating the importance of postsecondary costs and quality. Using the conditional logistic choice model to exploit extensive match-specific information between individuals and colleges and include the thousands of alternatives available to prospective students, the paper is able to approximate an individual's college outlook as never before. The results suggest that although tuition price was an important determinant of attendance for the class of 1972, college costs do not explain differences in enrollment for the class of 1992. However, price is still found to be an important factor when individuals choose between colleges, particularly among low-income students. Additionally, college quality has become a more important factor in college decisions.

JEL classification: I2; C2; J11

Keywords: College choice; Enrollment; Conditional logistic; College decision-making

1. Introduction

During the last 30 years, the college enrollment decision has become increasingly complex as higher education has transformed in many ways. First, American higher education grew from a collection of small, local markets to one that is integrated regionally and nationally (Hoxby, 1997). As predicted by models of industrial organization, this prompted colleges to differentiate themselves and increased the variation

* Tel: +1-617-496-4355; fax: +1-617-496-3095.
E-mail address: longbr@gse.harvard.edu (B. Long).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2003.10.004

in college options available to students. During the last thirty years, Hoxby finds that between-college variation in tuition, student quality, and subsidies grew substantially. In addition, during this time the list tuition price of colleges nearly doubled in real terms. Meanwhile, government and institutional financial aid programs expanded in both scope and amount with the introduction of the Pell Grant and subsidized loans. At the same time, the benefits of a college degree increased. Real incomes for those with a Bachelor's degree grew 14.6 percent from 1975 to 1998, while those with only a high-school degree experienced a 2.1 percent decline in real income.[1] With all of these major developments, have individuals altered the way in which they consider their college options? Theory predicts college demand should depend upon the benefits and costs of higher education, and therefore, one might expect postsecondary decisions to have changed in response to these trends. As such, this paper examines how the college decisions of high-school graduates from 1972, 1982, and 1992 changed during this period of great transformation.

In general, individuals have responded to the growing return to higher education by enrolling in greater numbers. In 1970, only 52 percent of high-school graduates attended college, but by 1998, this proportion had grown to 66 percent.[2] But with the major changes in the market for higher education, did the importance of various college factors also change? Have individuals responded to increases in financial aid and growing monetary benefits so that price does not affect decisions as much as previously? How has the role of distance changed as advances in transportation and communication as well as the increased availability of college information have made higher education options beyond a person's immediate community more accessible? As the market has become more segregated in terms of resources, has college quality become more important? Moreover, do students value quality more as recent studies suggest that particular college choices and the resources offered by schools have become increasingly important to eventual returns.[3] While many studies have estimated the impact of price on college enrollment, few provide any sense of how other college characteristics such as quality or distance factor into enrollment decisions. Furthermore, no study has examined whether the role of these factors has changed over time. This study addresses this hole in the literature by analyzing the changing role of college cost, quality, and distance in postsecondary decisions.

To understand how decisions about college are made, this paper approximates the outlook of a potential student using the conditional logistic model. Also known as McFadden's choice model, this methodology allows for the over 2000 possible college alternatives as well as takes advantage of important match-specific information between the individual and college.[4] These person-specific factors include the price charged (in-state versus out-of-state net Federal grants), distance to the college, and student similarity to prospective college peers. In this way, this paper serves as an improvement

[1] Source: U.S. Census Bureau, March CPS.

[2] Source: American College Testing Program, unpublished tabulations, derived from statistics collected by the U.S. Bureau of the Census.

[3] See Behrman et al. (1996), Brewer et al. (1999) and Hoxby and Long (1999).

[4] The application of the conditional logit to college decisions was first discussed by Manski and Wise (1983).

to previous work that has had to rely upon aggregated tuition measures or greatly simplified the matching process of students to colleges.

Two main questions are addressed. First, how do individuals decide *between* colleges conditional on attendance? Estimated off of the considerable variation within the market for higher education, the model parameters suggest how college price, quality, and distance factor into decisions between schools for three different time periods. Second, how do individuals decide *whether* to attend college at all? The paper uses logistic regression analysis with controls for family background and local labor market conditions to estimate how the likelihood of enrollment has changed over time in relation to college costs and resources. I approximate the enrollment decision by using the characteristics of the college predicted to be the most-likely choice by the conditional logistic choice model. As an alternative, the price and distance of the closest 2-year, public college are also used. These specifications serve as a better approximation of the relevant college option for a marginal student than the mean state characteristics used in earlier studies.

A variety of data sets are employed for this analysis. The National Longitudinal Study of the Class of 1972 (NLS72), High School and Beyond (HSB), and National Education Longitudinal Study (NELS) 1988 provide background and decision information on three cohorts of high-school graduates. These data sets were designed by the National Center for Education Statistics to compare successive cohorts and to permit the study of trends relevant to education. Still, care is taken to standardize the analysis of how decisions between colleges have changed over time since the college-going pool has changed substantially since 1972.

The results suggest that while college costs alone do not explain who does and does not attend college, low-income students in 1992 were as negatively affected by price as students in 1972 when choosing between colleges. Therefore, more attention needs to be given to where different groups of students attend college and the variation in resources by type of college. Additionally, quality is found to have become more influential in college choice. In terms of the enrollment decision, college costs are estimated to have been important determinants of college attendance for the class of 1972, but they do not explain differences in enrollment for the class of 1992. Therefore, the importance of non-monetary factors needs to be examined.

2. Literature review

2.1. The importance of price

Many studies have tested for the sign and magnitude of the impact of price on college enrollment. Leslie and Brinkman (1989) review studies from the 1970s and 1980s and conclude that a \$1000 (2001 dollars) change in college costs is associated with a four percentage-point difference in college enrollment rates. More recent studies have found similar results exploiting state cross-sectional differences to estimate the effect of price. For example, Kane (1995) uses the HSB, National Longitudinal Survey of Youth (NLSY) 1979, and the October Current Population Survey (CPS) to link individual enrollment decisions to the mean tuition costs of a state. He finds that states with higher public tuition levels had lower college entry rates and estimates an elasticity of demand of −0.20 for public 2-year tuition, a price effect similar to those found by Leslie and Brinkman. Cameron and Heckman (1999) find a slightly larger impact of a six percentage-point enrollment effect per $1000 using the 1979 NLSY and similar cross-sectional variation in costs.

However, the literature on the effect of price on enrollment suffers from several problems. First, using state-level variation in tuition costs when estimating the effect of price could possibly bias results. As Rouse (1994) and Kane (1995) caution in their papers, interpreting state variation as a natural experiment for tuition changes has the problem that omitted state factors may be correlated with enrollment and tuition. It is difficult to distinguish the impact of tuition from any other characteristic of the state that has remained constant over time. For example, unobserved state preferences for higher education could bias the results by making the error term negatively correlated with the mean tuition level of the state.

Another weakness of the enrollment literature is that it fails to capture the substantial heterogeneity in price within the market for higher education. Most studies use state-level tuition variables such as the mean public tuition level or the price of the nearest community college, also strongly correlated with state tuition levels. However, these specifications mask much of the heterogeneity in the market for higher education. While tuition levels vary significantly by state, most of the variation exists within a state. For example, in California during the 1991–1992 school year (relevant for this study), UC-Berkeley charged a tuition of $3248 while CSU-Fullerton, another public, 4-year university, charged $1480. On the other end of the spectrum of costs, Stanford University charged $16,635. Similar examples can also be found for most other states. Cost, however, is only one way in which states and colleges differ. State variation in the number and selectivity of schools available to students is also substantial. In terms of school resources for student-related activities, a proxy for college quality, colleges like UC-Berkeley and Stanford University spent over $21,000 per pupil on instruction, academic support, and student services in 1992, while CSU-Fullerton only spent about $6100 per pupil. While the distribution of colleges in a state could affect enrollment decisions, this would be masked using state-level variables, and therefore, college heterogeneity presents a major source of identification problems when estimating the effect of price.

A third drawback to the literature is the lack of perspective on how college decisions have changed over time. While the literature does provide estimates of the effect of price at different times over the last 30 years, since each study involves different controls, econometric models, and data samples that have not been adjusted in order to be comparable to each other, it is difficult to get a true sense of how the effect of price has changed over time. Cameron and Heckman (1999) do consider how the importance of college price has changed over time for low-income students. Controlling for achievement and parental education, they find that family income does not explain differences in attendance and suggest that the negative relationship between costs and attendance has decreased in recent years because of the increasing availability of aid and the expansion of community colleges. However, this assertion has yet to be tested.

Beyond influencing the enrollment decision, price may also affect *which* college a student chooses to attend. For example, using microdata on high-aptitude students, Avery and Hoxby (2004) find that college choice is sensitive to tuition and room and board costs, and that individuals are attracted to a college when offered grants that reduce the net cost. Furthermore, there is some evidence that the effect of price on the choice of a particular college is becoming increasingly important as the returns to higher education have become more closely tied to the type of school attended. Hoxby and Long (1999) examine reasons for the growing variance in incomes among college-educated workers and find that, of the growth that can be explained, over 40 percent is due to the increasing segregation of students and resources among colleges. Therefore, it is becoming more important to not only attend college, but to attend the best possible college in order to secure a desirable standard of living.

2.2. The importance of other college characteristics

While other college characteristics may be important determinants in college decisions, little has been done to quantify their roles. First, the importance of college quality has largely been ignored in college decision models. The bulk of the focus on college quality is how it relates to the returns to education.[5] However, growing quality competition between colleges and the popularity of college rankings appears to suggest that quality plays an important role in the college choices of students. Hoxby (1997) examines how the transition from a local to nationally integrated market has spurred colleges to compete for students based on the quality of their educational products rather than price. Furthermore, Monks and Ehrenberg (1999) detail how a favorable change in the U.S. News & World Report college rankings positively affects a college's selectivity, applicant yield, and average SAT score for the entering class of freshman.

A second factor that has received little attention is distance to a college. Several studies assume the importance of proximity in college access. For example, Rouse (1995) uses proximity to a 2-year college to instrument for college enrollment. Moreover, Hoxby (1997) suggests that the role of distance has changed with advances in transportation and communication as well as the increased availability of college information. However, no study has estimated the effect of distance to a set of colleges on enrollment or choice.

3. Framework and model

This section outlines the theoretical framework and methodology of the paper. First, I discuss the theoretical framework and how my college choice model accounts for heterogeneity in the market for higher education and approximates the many alternatives facing a prospective student. Then, the data sources and variables for analysis are discussed.

[5] See Behrman et al. (1996), Hoxby (1997), Brewer et al. (1999), and Dale and Krueger (2002).

3.1. Theoretical framework

Assume an individual has J colleges from which to choose. Each school, j, can be described by a vector of characteristics, Y_j, which might include measures of college price, resources for students, and location. In this way, each college can be viewed as a package containing various attributes for a given price. From these characteristics, an individual can infer how much value-added each school will produce for his human capital and the consumption goods the college offers.[6] Another vector, X_i, contains the individual's characteristics such as high-school performance and family income. These factors could affect the individual's demand for education and his opportunity set. For example, prior academic performance will affect the likelihood of being accepted by a highly selective college. An individual's background could also affect the acceptable tradeoffs he perceives between schools. For example, an individual may be more willing to sacrifice school quality for a lower price if he is from a low-income family.

Let the value of the jth college (characterized by the vector Y) to the ith decision maker be given by $\mathsf{U}(Y_j,X_i)$. This value may partly depend on an interaction of the individual's characteristics with those of the school. For example, the cost of public college j for person i may be determined by residence—whether in-state versus out-of-state tuition is charged. Utility may have random elements so that all individuals with X_i are not assumed to have the same tastes:

$$\mathsf{U}(Y_j,X_i) = -(Y_j,X_i) + \varepsilon_{ij}. \tag{1}$$

That is, there are random deviations from the mean valuation _. If we assume that the non-random part of utility is a linear function of individual and college characteristics we get

$$\mathsf{U}(Y_j,X_i) = Z_{ij1}\beta_1 + Z_{ij2}\beta_2 + Z_{ij3}\beta_3 + \cdots + Z_{ijk}\beta_k + \varepsilon_{ij}, \tag{2}$$

where β is a vector of parameters, Z_{ij} are the variables that affect utility, and k is the total number of variables. Z may include variables that describe the elements of Y (i.e. the college) or interact Y and X to form match-specific measures (i.e. a individual-specific tuition charge).

Individuals compare the potential returns to attending different colleges along with the option of *not* enrolling and entering the labor market directly. The college decision is therefore made up of *two* choices. The individual must determine his best college option and concurrently decide whether to attend college at all. The individual chooses the option that maximizes his lifetime utility subject to his budget constraint as shown in (3).

$$\text{choose } Y_k \text{ iff } \mathsf{U}(Y_k,X_i) \geqslant \mathsf{U}(Y_j,X_i)\ \forall k \neq j \text{ with } P_{ik} \leqslant I_i, \tag{3}$$

where I_i is the budget constraint and is related to income.

In this analysis, the enrollment and college choice decisions are treated separately employing conditional logistic and simple logistic models. A description of these models and the justification for estimating the decisions this way are described below.

[6] These two components of human capital accumulation, production and consumption, are an implication of the Becker and Rosen models of human capital. They are denoted as an institution's quality in this study.

3.2. Choosing between colleges: the conditional logistic model

The above framework emphasizes several points that must be addressed in estimating how individuals choose between colleges. First, there is substantial heterogeneity among colleges that must be adequately characterized. Although the literature largely uses the mean tuition level of the public colleges in a person's state, schools differ widely in their costs and the resources they offer. Second, match-specific information is an important part of the model since colleges treat students differently depending upon residence and academic background. A particular college may not suit all students well, and this will depend on how the student's characteristics compare with those of the college's student body. Finally, the framework demands that a model adequately capture the full range of the opportunity set of potential students. In order to understand how students make their college decisions, an empirical model should explore the tradeoffs between the opportunity selected and the alternatives not chosen. The model must be equipped to characterize the thousands of non-specialized, 2- and 4-year schools a student could attend.

The conditional logistic regression model is well suited for the college choice framework since it exploits extensive detailed information on alternatives, can account for match-specific details, and allows for multiple alternatives. Also known as McFadden's choice model (1973), the conditional logit has been used to study the choice of travel mode and occupation. Manski and Wise (1983) acknowledge the advantages of the conditional logit in accounting for the many college alternatives facing a prospective student, but use multinomial logistic regression analysis instead due to computational limitations. Unfortunately, the multinomial logit model forces one to aggregate information about choices and enter them in a restrictive manner making the conditional logit preferable for this application. While the form of the likelihood function of the conditional logit is similar to that of the multinomial logit, the variables of the conditional logit are choice-specific attributes rather than individual-specific characteristics. Therefore, variation in the attributes of the colleges in a student's choice set drives the estimates.

To estimate the conditional logit, the data are organized as pair-wise combinations of each student i with each school j with the observations stratified by individual into groups of j. Using these combinations, the conditional logit model is made up of j equations for each individual i, with each equation describing one of the alternatives. This format allows for match-specific variables based on the interaction of individual i with college j. The conditional logit model then calculates the probability of enrollment at each of the colleges in the stratum relative to all other alternatives with the dependent variable equal to one for the alternative that was chosen. Under the assumption that the ε_{ij}'s are independent and identically distributed with the extreme value distribution, we get the conditional logit functional form

$$\Pr(\mathrm{E}_{ij}) = \mathrm{e}^{Z_{ij}\beta}/(\mathrm{e}^{Z_{i1}\beta} + \mathrm{e}^{Z_{i2}\beta} + \cdots + \mathrm{e}^{Z_{iJ}{}^{\beta}}) \quad (4)$$

where $Z_{ij}\beta = \beta_1 \ (X_{i1}Y_{j1}) + \beta_2 Y_{j2} + \beta_3 Y_{j3} + \beta_4(X_{i4}Y_{j4}) + \cdots + \varepsilon_i$.

The probability of individual i attending college j, Pr (E_{ij}), will be a function of the variables that define Z_{ij}, the characteristics included in vector Y_j and combinations of X_i interacted with Y_j. The format allows for maximum likelihood estimates of β, and the probability of any particular choice can be calculated using the conditional logit specification displayed in (4).

Since the likelihood of attendance at each college is calculated relative to the alternatives within each stratum, there must be variation within the strata for estimation purposes. For this reason, student characteristics cannot be included independently in estimation, and the estimates do not identify the causal effect of a student's attributes on enrollment. [7] Instead, the model indicates how the attributes of a college affect the likelihood that an average individual will choose to enroll at the school. This study reports the results in the form of odds ratios. These should be interpreted as the proportional change in the odds of student i attending college j for a unit increase in the variable, holding all other variables constant.

If the independence of irrelevant alternatives (IIA) condition is met, the estimates will be consistent even if the decision to attend college at all is endogenous. Possible endogeneity in the choice set develops from the fact that colleges available to a student will depend upon the previous decision of where to apply. This application decision is based upon a student's ranking of the colleges, and therefore, the opportunity set that a student will ultimately face is partly endogenous. However, as long as students apply to schools that they determine to be most preferred, estimation will retain good statistical properties due to the IIA property. [8]

The conditional logit model will estimate the choice between colleges, but an individual must also decide whether to attend college at all. For this analysis, the group that did not attend college enters the analysis in a separate model that examines the enrollment decision. This is done for several reasons. First, while ordinarily a nested logit model would be used to estimate the two decisions simultaneously, the conditional logit model is too complex for a simple application of the theory. Second, there is no adequate way to describe the "not attend" option using the conditional logistic model. Describing the not attend option as an alternative with zero tuition cost, zero median test scores, and a distance of zero would bias parameter estimates. Since some individuals would choose not to attend (the option with no cost), the negative effect of tuition price on college choice would be exaggerated. In addition, the decision of whether to attend college is most likely nonlinear relative to choosing between different schools. Therefore, estimates of the model including the "no college" group are unlikely to accurately describe both the college choice and the decision of whether to attend. Treating the enrollment and choice decisions separately is equivalent to deciphering the choices of individuals using backward induction as in game theory.

[7] The j equations within a strata are not independent, and a person's gender, for example, would difference out of all the equations within one stratum since each contains data on only one individual. Therefore, unlike the multinomial regression model, non-college alternatives such as local labor market conditions cannot be included within the model since they are individual-specific.

[8] See Manski and Wise (1983) for further discussion. Also see McFadden (1979).

3.3. Whether to attend: the logistic model

As stated in Eq. (3), an individual will compare his options, including attending college, entering the labor force, or not working at all, and choose the alternative that maximizes his lifetime utility subject to his budget constraint. Given diverse backgrounds, individuals may enjoy the costs, and benefits of higher education differently. For example, low-income individuals may face higher capital costs, while children whose parents did not attend college may lack sufficient information about college preparation or application. In addition, low-ability students may realize smaller gains from a college education if skill and education are complements. As such, these groups are likely to be on the margin of deciding whether to attend college, and therefore, they provide the best picture of how college enrollment decisions have changed over time.[9] To examine the importance of college characteristics on enrollment, the analysis employs a logistic regression model with controls for family background, student achievement, and the unemployment rate of the individual's county.

To approximate the relevant college option facing a potential student, I use the parameters of the conditional logit model to predict the college that is the *most likely* alternative if the individual *does* decide to attend. Although the model contains the characteristics of only one college for a given individual, the college will differ for each person. Therefore, unlike analyses in the previous literature, which just assigned the average college characteristics of the state, this study uses within-state variation in the school of interest based on the exact residence and ability of the student. In this way, this methodology attempts to account for the substantial heterogeneity in the market of higher education and test how the distribution of colleges in a state could affect enrollment decisions rather than mask this variation by using state-level variables.

This strategy of using the predicted most-likely alternative assumes that individuals make decisions between colleges in the same way regardless of whether they actually decide to attend or not. This assumption seems plausible since individuals with similar observable characteristics seem to make different college decisions due to the local higher education options available to them. Several studies conclude that individuals from more disadvantaged backgrounds are more likely to attend if they have a nearby college option with a reasonably low tuition price (Rouse, 1994; Kane, 1995). Therefore, enrollment decisions may be due to the options available and not that the decision-making process differs by background. However, since this assumption may be troublesome, if students who decide not to attend weight their options differently or lack the information needed to make decisions that would maximize their utility, the enrollment decision is also estimated using the characteristics of the closest 2-year, public college. For most students likely to be on the margin of attendance, the option predicted by the conditional logistic choice model is often the nearest 2- or 4-year, public college.

[9] Students with high SAT scores, college-educated parents, and above-average family incomes were likely to attend college in both 1972 and 1992 suggesting that their decisions have changed little during this period.

Table 1
Student summary statistics

	1972 high school graduates (11,060 observations)			1982 high school graduates (7900 observations)			1992 high school graduates (7869 observations)		
	All students	Attended college	Did not attend	All students	Attended college	Did not attend	All students	Attended college	Did not attend
	—	51.23	48.77	—	61.78	38.22	—	72.33	27.67
Female	50.06	48.98	51.21	48.46	51.61	43.36	50.85	52.86	45.57
Black	10.03	9.05	11.05	13.09	11.76	15.24	9.79	9.08	11.62
Hispanic	4.03	3.48	4.62	14.80	13.50	16.89	10.43	9.80	12.08
Asian	1.21	1.85	0.54	3.30	3.97	2.22	7.90	9.01	5.01
Native American	1.08	0.67	1.50	2.05	1.52	2.91	0.97	0.72	1.61
Parent's education (years)	12.96 (2.39)	13.76 (2.44)	12.13 (2.04)	13.52 (2.67)	14.09 (2.77)	12.50 (2.15)	14.44 (2.62)	14.92 (2.66)	13.20 (2.07)
Family income (categorical)	5.62 (2.89)	6.31 (2.89)	4.87 (2.69)	4.13 (1.77)	4.42 (1.69)	3.67 (1.78)	10.37 (2.50)	10.77 (2.39)	9.34 (2.51)
Low-income dummy variable	24.54	16.93	32.54	19.94	14.16	29.28	20.14	15.46	32.38
Composite SAT	840.8 (180.9)	908.1 (187.3)	770.1 (143.3)	838.78 (184.41)	894.64 (182.34)	748.47 (148.46)	878.5 (212.2)	914.0 (209.1)	786.0 (191.3)
High-ability dummy variable	8.49	15.09	1.56	7.77	11.70	1.42	15.28	18.89	5.83

Notes: Numbers are percentages unless otherwise noted. Observations were dropped if the individual did not graduate from high-school in the designated year, went to high school outside of the United States, or enrolled in a specialty college (i.e. art school, nursing college, etc.). "Attended College" is defined as enrolling in any college within 2 years of high school graduation. Income categories are not comparable across data sets due to different group definitions. The SAT composite is defined as the actual score where available; otherwise the ACT score converted or the SAT score was predicted using the student's G.P.A. and demographics. Low-income is defined as having a family income during the senior year of high school below approximately $23,000 to $25,000 in 2000 dollars. For the 1972 sample, if family income information was not available, the SES measure was used to determine if the student was low income.

3.4. Data and variable description

Three data sources provide information on cohorts of high-school graduates from 1972 to 1992. The NLS72, the HSB, and the NELS, 1988 provide information on the backgrounds and college decisions of individuals who graduated from high-school in 1972, 1982, and 1992, respectively. Each data set was collected by the National Center for Education Statistics (NCES) with the explicit purpose of being made to be comparable to each other. Observations were dropped if they were missing crucial background or college information. Moreover, other criteria were imposed to make the samples reflective of traditional high-school graduates considering attendance at non-specialized colleges. [10] Table 1 displays summary statistics for each data set. Due to these changes in the make up of the cohorts, the analysis below takes extra care to standardize the groups that are being compared.

[10] Observations were dropped if the individual did not graduate from high-school in the designated year, went to high-school outside of the United States, or enrolled in a specialty college (i.e. art school, nursing college, etc.).

While the data sets offer a wealth of information about cohorts of high school graduates, they also have several drawbacks. First, since some groups were over-sampled when the data sets were created, the data sets are not nationally representative.[11] Another drawback to the data on students is that they were created using clustered sampling based on high schools. Unfortunately, current statistical programs do not have a correction for this when using the already complex conditional logistic regression model. However, a comparison of the logistic model results with and without the correction for the clustered sampling suggests that this feature of the data does not alter the results much. To address concerns about the clustered sampling, the analysis employs more stringent statistical significance thresholds based on the largest design effects found in the logistic regression analysis as described below.

To approximate the college options of students, information on college costs, quality, and location was garnered from several sources. First, the Integrated Postsecondary Education Data System (IPEDS) and its predecessor, the Higher Education General Information Survey (HEGIS), provided institutional characteristics and financial data. The 1971–1972, 1981–1982, and 1991–1992 schools years were used to describe the characteristics of the colleges, while the student cohorts were in their senior year of high-school (the time during which they were making their college decisions). Specialized colleges, such as nursing or music schools, were dropped as well as graduate institutions, such as law schools. Additional data on median student-body scores was taken from Cass' and Birnbaum's *Comparative Guide to American Colleges* for the 1971–1972 school year and Barron's *Profiles of American Colleges* for the 1981–1982 and 1991–1992 school years.[12] Information on faculty quality is from the College Board's Annual Survey of Colleges data. Summary statistics of the college data sets are outlined in Table 2.

To estimate the conditional logistic model, each student was matched with each possible college. College price, distance, and quality are defined in the following ways. First, the student is assigned the in-state tuition price if a resident of the same state as the college; otherwise the out-of-state price was used. In order to have a better sense of net price, the amount of Federal Pell Grants that a person could expect to receive was subtracted from the list tuition price.[13] Distance was

[11] I avoid using the weights included with the data sets since the sampling frame is such that one is not able to generate a nationally representative sample of high-school graduates or college-going students, particularly after observations with missing values are dropped.

[12] If the test score information was missing for a college but the school was grouped into a competitiveness category, the score assigned is equal to the mean of that category. All 2-year colleges missing score information were assigned 700, a level that qualifies basically all students for attendance.

[13] Other types of scholarships and grants could not be accounted for since information on a student's specific need is unavailable. Furthermore, aid is very individualized, and the award amounts at all possible colleges would be impossible to determine. The lack of information about aid beyond Pell Grants is not a large concern given few students receive substantial grants. Moreover, students who do receive scholarships often do so from only a few schools within their opportunity sets making the awards school-specific and predictions unfeasible.

Table 2
College summary statistics (2000 dollars)

	1971–1972 school year		1981–1982 school year		1991–1992 school year	
	Public	Private	Public	Private	Public	Private
Four-year colleges and universities						
Observations	518	764	485	829	532	1006
In-state tuition	$1963	$7151	$1788	$7550	$2830	$11,029
	(759)	(2370)	(733)	(2795)	(1143)	(4628)
Instructional expenditures	$4396	$4569	$4104	$4017	$7279	$11,037
	(1697)	(2468)	(1577)	(2533)	(2532)	(4618)
Student body SAT %tile	57.5	68.1	56.3	66.8	4765	5169
	(20.6)	(18.1)	(22.9)	(21.2)	(1911)	(3443)
Student–faculty ratio	41.3	42.4	15.0	12.0	56.9	61.4
	(24.7)	(23.9)	(3.5)	(3.6)	(15.5)	(17.9)
% of faculty with a Ph.D.	15.6	12.3	44.5	42.7	53.2	46.2
	(3.4)	(3.6)	(21.0)	(21.7)	(18.8)	(22.3)
FTE student enrollment	6871	1915	8247	2290	7627	1625
	(7186)	(2596)	(8082)	(3317)	(6432)	(1951)
Two-year colleges						
Observations	684	201	655	75	904	232
In-state tuition	$1456	$4942	$946	$4773	$1649	$7235
	(1181)	(2090)	(479)	(1925)	(1050)	(2855)
Instructional expenditures	$3386	$4949	$2596	$2274	$4117	$7238
	(1578)	(2080)	(857)	(913)	(1946)	(2853)
Student body SAT %tile	3526	3242	38.0	40.6	3283	2629
	(1192)	(1776)	(13.0)	(17.1)	(1340)	(1424)
Student–faculty ratio	49.1	40.0	18.4	15.1	21.0	21.0
	(11.6)	(15.9)	(4.0)	(4.1)	(0.0)	(0.0)
% of faculty with a Ph.D.	8.4	17.1	15.9	19.0	8.6	10.1
	(7.7)	(13.2)	(4.2)	(8.3)	(8.7)	(10.5)
FTE student enrollment	2227	478	3652	708	3692	556
	(2340)	(566)	(3598)	(835)	(3748)	(714)

Source: IPEDS and HEGIS for institutional characteristics and financial data. Additional data on median student-body scores was taken from Cass' and Birnbaum's *Comparative Guide to American Colleges* for the 1971–1972 school year and Barron's *Profiles of American Colleges* for the 1981–1982 and 1991–1992 school years. Information on faculty quality is from the College Board's Annual Survey of Colleges data.

computed using the zip code of the college and the zip code of the student's high-school. [14]

To proxy for college quality, several measures were used based on the literature on the returns to college quality on labor market outcomes. First, I use the median

[14] Unfortunately, the HSB does not provide zip code information for each respondent. However, using information provided about each individual's county of residence (unemployment rates, mean per capita income, etc.), I was able to determine approximately 90 percent of the sample's locations within a state. The rest were assigned their state's population-weighted geographic center in order to calculate the distance from potential colleges.

SAT score of the college's student body, the most common measurement of quality in the literature. It has been used by Dale and Krueger (1999) and Brewer et al. (1999) to approximate the ability of a college's student body and the selectivity of a school.[15] Second, I use instructional expenditures per student to proxy for the amount of resources and quality of the educational product offered. Behrman et al. (1996) and Brewer et al. (1999) cite this as a measure, and educational expenditures accounts for a large part of the score used to rank colleges by quality in the U.S. News and World Report (USNWR) Annual Survey.[16] Finally, to measure faculty quality I use the college's student–faculty ratio and the percentage of the faculty with a Ph.D. These measures approximate the amount of contact a student has with the faculty and the skill of the faculty. These measures have also been used or cited by Behrman et al. (1996) and Brewer et al. (1999).[17]

The framework for this model emphasizes the demand aspects of the college decision. However, supply constraints may also be important in the college matching process. Therefore, several variables are used to deal with this concern. First, in order to account for the quality of the match between the student and school, the model includes a variable that measures the difference between the SAT score of the individual and the median score of the college's student body in linear and quadratic form.[18] These variables account for admissions constraints since they prevent the model from predicting that individuals with low test scores will attend more selective schools. Since a difference of 100 points does not mean the same thing at all places in the distribution of scores (e.g. a 400 versus 500 compared to a 1500 versus 1600), test scores were transformed into percentiles to better standardize the measure. A second attempt to account for supply-side influences on college decisions is the inclusion of undergraduate full-time equivalent (FTE) enrollment. This allows for the fact that larger schools have a higher level of visibility and are able to admit greater numbers of students. As a result, a large, public university will be given a greater choice probability than a small, liberal arts college, a reasonable result given enrollment patterns.

Quadratic tuition, instructional expenditure, and distance variables along with a cubic distance variable are also included to allow for nonlinear relationships in estimation. A dummy variable for college level and interactions between this variable and others are also included to account for differing expectations about what students are willing to pay depending on whether a school is a 2- or 4-year college. For example, an

[15] The median SAT of a college's student body is not put explicitly in the model. See the student–college match variables described below.

[16] In the USNWR rankings, total educational and general expenditures accounts for ten percent of the score, the third largest weight, while faculty compensation, an instructional expenditure, accounts for another seven percent of the rating. Total college expenditures are not used since they include many items that do not directly affect education production (i.e. research and infrastructure).

[17] As expected, the measures of college quality are correlated with each other to some degree. For example, for the 1991–1992 college sample, instructional expenditures per student is correlated with in-state tuition price 0.50, college SAT percentile 0.58, percent of the faculty with a Ph.D. 0.54, and the student–faculty ratio −0.49.

[18] This is equivalent to using a college median SAT variable and one defined as the difference between the student's and college's score. However, the specification used in the analysis allows the effect of the difference to differ depending on whether the student's or college's score is higher.

individual may never be willing to pay a tuition price of $15,000 for a 2-year college but would be willing to do so for a 4-year school.

4. The determinants of college choice and enrollment

To illustrate the gains of using the conditional logit model, I first estimated logistic and multinomial logit models traditionally used in the literature. Without the benefit of describing all of a student's alternatives or using match-specific variables, the data had to be aggregated to the state level resulting in variables similar to those in Kane (1995). The results were often statistically insignificant and contrary to the expected sign. The logistic model suggests that there is a negative relationship between enrollment and 2-year college tuition levels, but the estimates are not statistically significant. Results from the multinomial logistic models suggest that higher tuition levels increase the likelihood of enrollment at that type of school, a counterintuitive result likely to be related to the correlation between college price and quality. Without the ability to more accurately describe the choices facing a student, these models provide conflicting and statistically insignificant results. The models below demonstrate the improvements realized by using the conditional logistic choice model instead.

4.1. Estimates of the conditional logit model—which college to attend?

Table 3 displays estimates from the conditional logistic choice model on how individuals choose between colleges during the first 2 years after high-school graduation. For each cohort, the model estimates how the attributes of a college affect the probability of that a person would choose to attend that school conditional on attending any postsecondary institution. As noted above, the standard errors for this analysis remain uncorrected for the clustered sampling of the data given current estimation procedures. Therefore, the benchmarks for statistical significance have been adjusted assuming a design effect of 3.5 (the largest level found in other estimation).[19] However, these more stringent benchmarks do not change most of the results. Figures in italics would have been statistically significant using customary levels.

As expected tuition and distance negatively impact the likelihood of an individual choosing a college, all else equal, while measures of college quality have a positive impact. Moreover, a comparison across cohorts suggests that the importance of these factors has changed over time. For the 1972 cohort, a $1000 (2000 dollars) difference in tuition price would reduce the probability of choosing to attend a particular 4-year college 53 percent, all else equal based on the observable characteristics. However, a 1992 graduate would be only 35 percent less likely to choose a 4-year school under

[19] While *Stata* has a number of adjustments for clustered sampling (the "svy" commands), such a correction has not been developed for use with the conditional logistic model. However, in order to lose statistical significance, the design effects (a measurement of how much survey design affects the variance estimates) for many of the estimated standard errors would have to be quite large (over 150 for the standard error estimated with tuition price for the NLS72 sample). Estimates of the design effects from the logistic analysis discussed below were only 0.99–2.1 for the NLS72 and 1.2 and 3.5 for the NELS suggesting that even if the conditional logistic results were adjusted for clustering, the increase in the size of the standard errors would not change the significance of most of the point estimates.

Table 3
College choice conditional on attendance (conditional logistic model)
Dependent variable: attended college *j* within 2 years of graduation (odds ratios and *Z*-statistics)

	1972 high school graduates		1982 high school graduates		1992 high school graduates	
	Main effect	Two-year interaction	Main effect	Two-year interaction	Main effect	Two-year interaction
College costs						
Tuition price	0.4686**	1.4096**	0.5809**	0.7520**	0.6548**	0.8531**
(per \$1000)	(32.32)	(7.47)	(26.68)	(3.84)	(39.21)	(4.50)
Tuition price^2	1.0485**	0.9678**	1.0328**	*1.0257*	1.0147**	*1.0102*
(per \$1000^2)	(24.87)	(4.93)	(21.98)	*(2.08)*	(31.91)	*(2.94)*
Distance	0.1665**	0.0534**	0.1954**	0.3382**	0.2668**	0.0805**
(per 100 miles)	(65.29)	(31.46)	(60.91)	(17.39)	(64.66)	(35.06)
Distance^2	1.1538**	1.4848**	1.1058**	1.0767**	1.0849**	1.5972**
(per 100^2 miles)	(36.89)	(24.51)	(31.92)	(11.16)	(37.92)	(28.59)
College quality						
Instructional expend.	1.0380	1.1248	1.0303	1.5195*	1.1035**	1.4999**
(per \$1000)	(1.46)	(1.23)	(1.27)	(3.08)	(6.08)	(4.47)
Instructional Expend.^{2}	*0.9960*	*0.9732*	0.9984	*0.9549*	0.9979**	0.9462**
(per \$1000^2)	*(2.64)*	*(2.46)*	(1.42)	*(2.13)*	(3.85)	(5.08)
Student–faculty ratio	0.9980		1.0003		*1.0127*	
	(0.48)		(0.07)		*(2.95)*	
Percent faculty with	1.0050**		1.0048**		1.0060**	
Ph.D. (per 10% pts)	(7.18)		(5.46)		(6.20)	
Students–college match						
Student SAT %tile	0.6525**		0.8662**		0.7129**	
larger (per 10 pts)	(10.26)		(4.64)		(11.26)	
School SAT %tile	0.9950		0.8324**		1.1809**	
larger (per 10 pts)	(0.16)		(5.75)		(4.78)	
College level						
Two-year college	*1.7242*		1.1219		4.8538**	
dummy variable	*(2.49)*		(0.48)		(7.93)	
Strata (individuals)	5666	—	4881	—	5693	—
N (combinations)	12,118,588	—	9,651,768	—	15,011,370	—
Pseudo-R-squared	0.4719	—	0.4400	—	0.4224	—

Notes: *Z*-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college *j* is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. Monetary amounts are in 2000 dollars. Additional controls include: cubic distance, student–college match variables squared, FTE enrollment, and FTE enrollment squared. To address concerns about the clustered sampling, the benchmarks for statistical significance have been adjusted assuming a design effect of 3.5 (the largest level found in other estimation). Figures in italics would have been significant using customary significance levels.

**Statistical significance at the 5 percent level given concerns about clustered sampling (Z-statistic $\geqslant$ 3.61).

*Statistical significance at the 10 percent level given concerns about clustered sampling (Z-statistic $\geqslant$ 3.04).

those circumstances. The declining role of price seems to have been fairly constant during the time period as the estimated effect of tuition on the 1982 cohort is near the midpoint between the other two effects. The role of distance is also estimated to have weakened. All else equal, a 1972 graduate would be 83 percent less likely to choose a 4-year college that was 100 miles farther away, while a 1992 graduate would be only 73 percent less likely.

Except for the percentage of the faculty with a Ph.D., college quality is not found to be an important factor in the college decisions of the 1972 and 1982 cohorts. However, for the class of 1992, an additional $1000 in instructional resources increases the likelihood of attending a 4-year college by 10.4 percent, *ceteris paribus*. The impact on choosing a 2-year college is even larger. Likewise, the importance of the match between the student and college is different for the later cohort. While all cohorts are less likely to attend colleges that have lower median SAT scores, students from the 1992 cohort are more likely to choose schools that have higher scores than their own, a signal of students striving for schools with better students.

Of course many of the differences between the cohorts may be related to the changing makeup of college-bound students for each sample. To help alleviate the concern that the differences are being driven by variation in the characteristics of the students rather than real changes in decision-making patterns, I employ two strategies. The first strategy is to standardize the later cohorts to reflect the 1972 population. This was done by weighting the 1982 and 1992 graduates by the likelihood that they would have enrolled in college in 1972 given their demographic and background characteristics (gender, race, family income, parent's education level, and high-school performance). Therefore, the question becomes "given an individual would have attended college in 1972, how did he choose between colleges in 1992?" Table 4 displays the same models as the previous table using this methodology for the later two cohorts.

The weighted results for the 1982 and 1992 samples do not differ appreciably from those in the previous table. The estimates of the effect of price and distance are slightly less negative while the role of instructional expenditures is a little more positive for the 1992 graduates. Larger differences are found in the estimates of the effect of the student–college match and the percentage of the faculty with a Ph.D. However, the general results remain the same. During the 20-year period, the negative influence of price on college choice fell by a little more than one-third. Likewise, the role of distance declined 17 percent. The importance of educational resources and faculty quality are estimated to have increased for the more recent cohorts. It is also important to note that the role of the 2-year college increased substantially for the 1992 cohort. All else equal, these students were found to be much more likely to attend a 2-year rather than 4-year college. It is likely that this result is tied to the expansion of community colleges in many states during the period.

A second method to reduce differences between the cohorts is to limit the sample to students with specific characteristics. This strategy has a couple of other advantages. First, it allows for more detailed examination of the decisions of groups that are of special interest for policymakers and researchers. Moreover, since the samples are limited, the memory requirements of the analysis are reduced enabling me to pool the different samples in one data set and test for differences across years using interaction

Table 4
College choice conditional on attendance (conditional logistic model) weighed by the probability of attendance in 1972
Dependent variable: attended college *j* within 2 years of graduation (odds ratios and *Z*-statistics)

	1972 High school graduates		1982 graduates (weighted by likelihood of 1972 enrollment)		1992 Graduates (weighted by likelihood of 1972 enrollment)	
	Main effect	Two-year interaction	Main effect	Two-year interaction	Main effect	Two-year interaction
College costs						
Tuition price	0.4686**	1.4096**	0.5953**	*0.8111*	0.6669**	0.8330**
(per $1000)	(32.32)	(7.47)	(23.16)	(*2.06*)	(42.21)	(4.97)
Tuition price^2	1.0485**	0.9678**	1.0318**	1.0068	1.0143**	1.0112*
(per $1000^2)	(24.87)	(4.93)	(19.95)	(0.37)	(36.45)	(3.10)
Distance	0.1665**	0.0534**	0.2139**	0.3215**	0.3056**	0.0809**
(per 100 miles)	(65.29)	(31.46)	(52.45)	(14.52)	(64.71)	(33.08)
Distance^2	1.1538**	1.4848**	1.1025**	1.0698**	1.0777**	1.5745**
(per 100^2 miles)	(36.89)	(24.51)	(27.16)	(9.28)	(37.49)	(25.83)
College quality						
Instructional expend.	1.0380	1.1248	*1.0588*	*1.4880*	1.1243**	1.5965**
(per $1000)	(1.46)	(1.23)	(*2.25*)	(*2.33*)	(8.28)	(4.45)
Instructional expend.^{2}	*0.9960*	*0.9732*	*0.9977*	0.9575	0.9975**	0.9349**
(per $1000)	(*2.64*)	(*2.46*)	(*1.95*)	(1.59)	(5.56)	(5.10)
Student–faculty ratio	0.9980		1.0012		*1.0096*	
	(0.48)		(0.22)		(*2.28*)	
Percent faculty with	1.0050**		1.0066**		1.0080**	
Ph.D. (per 10% pts)	(7.18)		(6.78)		(8.78)	
Students–college match						
Student SAT %tile	0.6525**		0.8388**		0.6290**	
larger (per 10 pts)	(10.26)		(5.11)		(16.81)	
School SAT %tile	0.9950		0.7582**		*1.0682*	
larger (per 10 pts)	(0.16)		(6.74)		(*1.92*)	
College level						
Two-year college	*1.7242*		1.0879		6.5432**	
dummy variable	(*2.49*)		(0.28)		(8.58)	
Strata (individuals)	5666	—	4881	—	5693	—
N (combinations)	12,118,588	—	8,895,488	—	14,990,130	—
Pseudo-R-squared	0.4719	—	0.4228	—	0.3946	—

Notes: *Z*-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college *j* is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. Monetary amounts are in 2000 dollars. Additional controls include: cubic distance, student–college match variables squared, FTE enrollment, and FTE enrollment squared. To address concerns about the clustered sampling, the benchmarks for statistical significance have been adjusted assuming a design effect of 3.5 (the largest level found in other estimation). Figures in italics would have been significant using customary significance levels. To standardize the samples across time, individuals in the 1982 and 1992 samples are weighted by their likelihood of attending college in 1972 based on student characteristics.

**Statistical significance at the 5 percent level given concerns about clustered sampling (Z-statistic $\geqslant 3.61$).

*Statistical significance at the 10 percent level given concerns about clustered sampling (Z-statistic $\geqslant 3.04$).

Table 5
College choice conditional on attendance (conditional logistic model)—low-income students
Dependent variable: attended college *j* within 2 years of graduation (odds ratios and *Z*-statistics)

	Main effect	Two year interactions	1982 interaction	1982* 2-year interactions	1992 interactions	1992* 2-year interactions
College costs						
Tuition price	0.5669**	*1.1623*	1.0908	*0.7834*	1.0690	*0.8959*
(per $1000)	(8.91)	(*1.90*)	(0.98)	(*2.73*)	(0.94)	(*1.78*)
Tuition price^2	1.0340**	0.9869	0.9934		*0.9841*	
(per $1000^2)	(5.91)	(1.38)	(0.90)		(*2.77*)	
Distance	0.0652**	0.2729**	2.1275**	1.1107	3.3781**	*1.4075*
(per 100 miles)	(31.65)	(9.52)	(6.33)	(0.54)	(11.85)	(*2.10*)
Distance^2	1.3423**	1.1322	0.8129**	1.0022	0.8083**	*0.9800*
(per 100^2 miles)	(19.75)	(11.01)	(12.91)	(0.19)	(13.71)	(*1.96*)
College quality						
Instructional expend.	1.1213	1.0502	0.9011	*1.2914*	0.9917	*1.2356*
(per $1000)	(1.57)	(0.37)	(1.10)	(*2.13*)	(0.10)	(*2.44*)
Instructional expend.^{2}	0.9920	*0.9711*	1.0071		1.0058	
(per $1000^2)	(1.55)	(*2.03*)	(1.13)		(1.07)	
Two year interactions						
Two-year college dummy variable	1.4350		0.8815		*0.5192*	
	(1.00)		(0.28)		(*1.76*)	
Strata (individuals)	2530	—	—	—	—	—
N (combinations)	5,687,750	—	—	—	—	—
Pseudo-R-squared	0.4900	—	—	—	—	—

Notes: *Z*-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college *j* is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. Monetary amounts are in 2000 dollars. Low-income is defined as having a family income during the senior year of high-school below approximately $23,000 to $25,000 in 2000 dollars. Additional controls include: cubic distance, student–college match variables squared, FTE enrollment, and FTE enrollment squared. To address concerns about the clustered sampling, the benchmarks for statistical significance have been adjusted assuming a design effect of 3.5 (the largest level found in other estimation). Figures in italics would have been significant using customary significance levels.

**Statistical significance at the 5 percent level given concerns about clustered sampling (*Z*-statistic $\geqslant$ 3.61).

*Statistical significance at the 10 percent level given concerns about clustered sampling (*Z*-statistic $\geqslant$ 3.04).

variables.[20] Table 5 focuses on low-income students, defined as having an income below approximately $23,000–$25,000 in 2000 dollars.[21]

[20] The conditional logistic models run for each cohort in Tables 3 and 4 each require approximately 2 GB of memory, thereby making it impossible to pool the three samples of data and run one model.

[21] Low-income individuals had a family income below $6000 in 1972 for the first cohort, below $12,000 in 1982 for the second cohort, and below $20,000 in 1991 for the third cohort. In terms of 2000 dollars, this translates into $24,643 for the 1972 sample, $21,921 for the 1982 sample, and $25,413 for the 1992 sample.

As found before, tuition and distance have a negative impact on the likelihood that a low-income individual will choose a particular college, ceteris paribus.[22] Moreover, as was found earlier for all students, the role of distance appears to have declined for later cohorts. However, low-income students in later cohorts were still found to be sensitive to price in their decisions between different colleges (none of the year interactions are statistically significant). This suggests that even after 30 years of financial aid programs, price negatively impacted the choices of graduates in 1982 and 1992 in a manner similar to the way it influenced the class of 1972. The interactions also suggest that while the quality of a 4-year institution was not important in the decisions of any cohort, the quality of a 2-year college may have become more influential for the later groups of low-income students (without the more stringent significance levels, the results would be statistically significant).

Table 6 instead focuses on the decisions of high-ability students, defined as having an SAT of at least 1100. Price may have become less important for this group as merit aid has increased and the returns to education for high-ability college graduates have grown. This notion appears to be true as the year interactions suggest that tuition price had less of a negative effect on the choices of the 1992 cohort than the graduates of 1972. Likewise, the role of distance has also fallen over time. Overall, college quality is found to be an important factor in the decisions of high-SAT students. The estimated effect is twice as large as that found in Tables 3 and 4. The reason most likely stems from the fact that ability and school quality are complements. If that is the true, then a high-ability student might expect to realize larger gains from schools with more resources than the average student, and the quality of a college might be more important. However, the role of quality did not become more important over the time period. It is likely that high-aptitude students have always cared about college resources, while average students have only begun to do so more recently.

4.2. *Estimates of the logistic model—whether to attend college?*

The rest of the analysis examines how college characteristics affect the likelihood of individuals enrolling in any college. Tables 7 and 8 display estimates of the effect of college price, distance, and quality on enrollment decisions with controls for background and the county unemployment rate. The relevant college price and distance is approximated in two ways. First, as has been utilized in the literature, I use the characteristics of the closest 2-year, public college. Nearly 17 percent of the 1972 cohort, 20.4 percent of the 1982 cohort, and 17.3 percent of the 1992 attended the closest public, 2-year college making this a somewhat reasonable proxy of the applicable school. However, using the college predicted to be "most-likely" by the conditional logistic models above does a better job predicting the relevant college alternative than using the closest public, 2-year college, and most certainly, the mean public, 2-year tuition of

[22] The literature predicts that low-income students are likely to have a high cost of capital, and therefore, tuition price and distance are expected to have particularly negative effects on their choices. In comparison to other students, the college choices of low-income students did not differ significantly from other students for the 1972 and 1982 cohorts. However, the differences were statistically significant for the 1992 graduates with low-income students be more negatively affected by price and distance.

Table 6
College choice conditional on attendance (conditional logistic model)—high-SAT students
Dependent variable: attended college j within 2 years of graduation (odds ratios and Z-statistics)

	Main effect	Two year interactions	1982 interaction	1982* 2-year interactions	1992 interactions	1992* 2-year interactions
College costs						
Tuition price	0.5169**	1.0851	1.1137	0.8834	1.3214**	*0.8015*
(per $\$1000$)	(12.97)	(0.71)	(1.54)	(0.64)	(5.07)	(2.29)
Tuition price^2	1.0421**	1.0013	*0.9890*		0.9725**	
(per $\$1000^2$)	(10.94)	(0.10)	(*2.26*)		(7.26)	
Distance	0.2505**	0.0480**	1.3430**	0.1385	1.3852**	1.3212
(per 100 miles)	(26.04)	(8.63)	(4.22)	(*2.29*)	(4.89)	(0.80)
Distance^2	1.1048**	1.6710**	0.9574**	*1.1434*	0.9665**	0.9723
(per 100^2 miles)	(15.41)	(9.36)	(5.87)	(*2.07*)	(4.33)	(0.94)
College quality						
Instructional expend.	1.2572**	*2.9406*	*0.8651*	*0.5750*	0.9683	0.9450
(per $\$1000$)	(4.85)	(*2.73*)	(*2.44*)	(*1.88*)	(0.66)	(0.33)
Instructional expend.^{2}	0.9919*	*0.8727*	*1.0058*		*1.0045*	
(per $\$1000^2$)	(3.44)	(*2.90*)	(*2.01*)		(*1.84*)	
Two year interactions						
Two-year college	*0.2232*		*11.8770*		8.3449*	
dummy variable	(*1.70*)		(*2.78*)		(3.24)	
Strata (individuals)	2455	—	—	—	—	—
N (combinations)	5,745,506	—	—	—	—	—
Pseudo-R-squared	0.3877	—	—	—	—	—

Notes: Z-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college j is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. Monetary amounts are in 2000 dollars. High-SAT students have an SAT greater than or equal to 1100. Additional controls include: cubic distance, student–college match variables squared, FTE enrollment, and FTE enrollment squared. To address concerns about the clustered sampling, the benchmarks for statistical significance have been adjusted assuming a design effect of 3.5 (the largest level found in other estimation). Figures in italics would have been significant using customary significance levels.

**Statistical significance at the 5 percent level given concerns about clustered sampling (Z-statistic $\geqslant 3.61$).

*Statistical significance at the 10 percent level given concerns about clustered sampling (Z-statistic $\geqslant 3.04$).

a state. Among the class of 1972, 24 percent enrolled in the most-likely college, while 23 and 18 percent of the 1982 and 1992 cohorts did so, respectively. Therefore, I use the characteristics of these most-likely colleges as a second specification. Often these schools were either the closest 2- or 4-year public college. Although this specification still only contains the attributes of one college for each individual, the full distribution of colleges within a state is taken into account rather than masking the significant within-state heterogeneity in postsecondary institutions by relying on state-aggregated measures or only public 2-year institutions without regard to student ability. The standard errors and Z-statistics have been adjusted to account for the clustered sampling of the student data.

Table 7
The college enrollment decision (logistic model)—all students
Dependent variable: attended any college within 2 years of graduation (odds ratios and Z-statistics)

	1972 Graduates		1982 Graduates		1992 Graduates	
	Closest 2-year public	Most-likely college	Closest 2-year public	Most-likely college	Closest 2-year public	Most-likely college
College costs						
Tuition price	0.8533**	0.8491**	1.0698	0.8913*	0.9671	1.0009
(per $1000)	(2.38)	(3.05)	(0.75)	(1.74)	(0.83)	(0.03)
Tuition^2	1.0231**	1.0152**	0.9902	1.0078*	1.0017	1.0000
(per $\$1000^2$)	(2.17)	(2.36)	(0.62)	(1.80)	(0.50)	(0.04)
Distance	0.3946**	0.4985**	0.5585**	0.6283*	0.4638**	0.8939
(per 100 miles)	(2.31)	(1.97)	(2.47)	(1.70)	(2.68)	(1.00)
Distance^2	2.7880*	1.8285	1.3414*	1.0438	1.2856	1.0177*
(per 100^2 miles)	(1.89)	(1.33)	(1.86)	(0.22)	(1.18)	(1.78)
College quality						
Instructional exp.	0.9283	0.9994	0.9912	1.0242	0.9161	1.0004
(per $1000)	(0.91)	(0.01)	(0.06)	(0.47)	(0.93)	(0.01)
Instructional exp.^{2}	1.0075	0.9993	1.0146	0.9973	1.0086	1.0000
(per $\$1000^2$)	(0.81)	(0.09)	(0.67)	(0.64)	(0.92)	(0.03)
Labor market						
County unemp.	0.9975	0.9943	0.9956	0.9996	1.0232*	1.0175
rate	(0.14)	(0.31)	(0.48)	(0.04)	(1.94)	(1.54)
Observations	8821	8821	7007	7007	7539	7563

Notes: Z-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college j is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. Since both the student data have clustered sampling based on high schools, this is taken into account using survey procedures ("svy" commands in *Stata* were used to compute the regression coefficients). Figures are in 2000 dollars. Additional controls include: student SAT percentile, family income, parent's highest level of education, and race and gender dummy variables, FTE college enrollment, FTE college enrollment squared, and distance cubed.

**Statistical significance at the 5 percent level.

*Statistical significance at the 10 percent level.

Table 7 shows the results for all students. For the class of 1972, tuition price had a negative effect on attendance. Individuals were 15 percent less likely to attend any college for each additional $1000 in price using the most-likely college as the relevant option (an elasticity of 0.34). When instead using the closest, 2-year public college, the estimated effect is slightly less negative. Ten years later, the role of price appears to be less negative. For the 1982 cohort, the price of the closest public 2-year does not explain who does and does not enroll in college. However, using the characteristics of the most-likely college, individuals are found to be 11 percent less likely to attend any college for an additional $1000 in price. This translates into an elasticity of 0.233, a little larger than that found by Kane (1995) using state tuition means for the same

Table 8
The college enrollment decision (logistic model)—marginal students
Dependent variable: attended any college within two years of graduation (odds ratios and Z-statistics)
Sample: marginal students (family income $< \$24,000$ or SAT < 900 or parents do not have a college degree)

	1972 Graduates		1982 Graduates		1992 Graduates	
	Closest 2-year public	Most-likely college	Closest 2-year public	Most-likely college	Closest 2-year public	Most-likely college
College costs						
Tuition price	0.8190**	0.8195**	1.0906	0.8766*	0.9834	1.0102
(per \$1000)	(2.99)	(3.50)	(0.90)	(1.80)	(0.37)	(0.27)
Tuition^2	1.0282**	1.0215**	0.9822	1.0106**	1.0001	0.9998
(per \$1000^2)	(2.72)	(2.80)	(1.09)	(2.05)	(0.02)	(0.15)
Distance	0.3704**	0.3349**	0.3810**	0.4190	0.5515*	0.8617
(per 100 miles)	(2.25)	(2.61)	(2.36)	(1.51)	(1.90)	(1.12)
Distance^2	3.0434*	3.4112*	2.5502*	1.3835	1.1322	1.0182
(per 100^2 miles)	(1.80)	(1.88)	(1.79)	(0.32)	(0.55)	(1.64)
College quality						
Instructional exp.	0.9490	1.0504	0.9853	1.0825	0.8835	0.9551
(per \$1000)	(0.58)	(0.58)	(0.09)	(1.34)	(1.16)	(1.07)
Instructional exp.^{2}	1.0059	0.9927	1.0140	0.9907*	1.0122	1.0040
(per \$1000^2)	(0.56)	(0.92)	(0.61)	(1.66)	(1.15)	(1.53)
Labor market						
County unemp.	1.0032	0.9986	0.9991	1.0038	1.0320**	1.0262**
rate	(0.16)	(0.07)	(0.09)	(0.40)	(2.42)	(2.11)
Observations	6971	6971	5284	5284	5027	5048

Notes: Z-statistics are reported in the parentheses to denote statistical significance. Effects are interpreted as the multiple by which the probability favoring attendance at college j is multiplied with a one-unit increase in that variable. Figures greater than one are considered positive effects. This is the group most likely to be on the margin of going to college as predicted by theory and shown in the data: only 42.9 percent of marginal students in 1972 attended college while 83.4 percent of the non-marginal students did. Similarly, 63.7 percent of the marginal students in 1992 attended college while 90.21 percent of the non-marginal students did. Since both the student data have clustered sampling based on high schools, this is taken into account using survey procedures ("svy" commands in *Stata* were used to compute the regression coefficients). Figures are in 2000 dollars. Additional controls include: student SAT percentile, family income, parent's highest level of education, and race and gender dummy variables, FTE college enrollment, FTE college enrollment squared, and distance cubed.

**Statistical significance at the 5 percent level.

*Statistical significance at the 10 percent level.

period. For the most recent cohort, however, the price of the most relevant college option does not explain differences in enrollment at all. This suggests that the role of price has fallen in terms of access. Similar statements could also be made about proximity. While the class of 1972 was 61 or 50 percent less likely to attend any college for each addition 100 mile, the effect is mixed for the 1992 cohort. The distance to the closest 2-year public college still seems to be important but not when using the characteristics of the most-likely college. Educational resources do not appear to play a role in determining whether an individual from any cohort chose to attend college.

While one might expect to see some changes among all students in their enrollment decisions, students on the margin of attendance are the ones most likely to be impacted by the major developments in higher education during the last thirty years. Table 8 limits the sample to this group. Individuals from families with an income below $24,000 (2000 dollars), an SAT score below 900, or parents who did not attend college are defined as the marginal group, and enrollment patterns from the data support this definition.[23] As expected, compared to the estimates in Table 7, these students appear to have been more negatively impacted by college price and distance. However, similar trends can be found. The role of price appears to have declined across the cohorts. While individuals from the class of 1972 were 18 percent less likely to attend any college for an additional $1000 in the cost of the relevant college (an elasticity of 0.46), the effect of price was not statistically significant for the most recent cohort. Likewise, the negative effect of distance also declined over the period. Again, the quality does not appear to influence the enrollment decision for any cohort.

The determinants of college attendance appear to be much different for the class of 1992 than those for the previous generations. Except for perhaps the distance to the closest public, 2-year college, no college characteristic is estimated to have statistically significant effects on enrollment for this group. This result may suggest that capital constraints are no longer a major deterrent in college decisions as asserted by Cameron and Heckman (1999) and Ellwood and Kane (2000). Given the changes estimated between the generations of high-school graduates, it is possible that the expansion of financial aid has aided in this result or that the benefits of a college education have grown to a level that paying the cost is perceived as a good investment. It is also possible that factors such as academic preparation, high-school resources, or college information have become increasingly important since 1972. However, one factor found to make a difference in the enrollment of students from the class of 1992 is the county unemployment rate. A 1-percent rise in unemployment is estimated to have increased the likelihood of enrollment by 3 percent for this group while having no effect on the earlier cohorts. This relationship may suggest that college decisions have become more closely linked to developments in the economy.

5. Conclusion

How individuals make decisions about college has significant implications for what the best policies are to increase access. Likewise, by looking at the past 30 years, changes in the considerations of college students reflect how past policies and changes in higher education have impacted access and choice. With the introduction of several

[23] For the class of 1972 high-school graduates, only 42.9 percent of marginal students attended college while 83.4 percent of the non-marginal students did. Likewise, for the class of 1992 graduates, 63.7 percent of the marginal students attended college, while 90.2 percent of the non-marginal students did.

major financial aid programs, the increasing segmentation of the market by resources and student ability, and the increasing return to a college degree, it is not surprising to find differences in the ways more recent cohorts of students made decisions about college compared to students decades ago.

The role of college costs in the enrollment decision was found to have decreased over the time period. For the class of 1972, a \$1000 increase in tuition costs is estimated to reduce the likelihood of enrollment by 15 percent. This translates into an elasticity of 0.34. The effect is slightly smaller for the 1982 cohort (the estimated elasticity was 0.23). However, for the class of 1992, tuition price did not help to explain differences in who did and did not enroll in college. The lack of statistical significance in the enrollment models for the class of 1992 suggests that factors beyond capital constraints play a role in deterring students from college attendance. Therefore, more research is needed to establish the importance of issues such as academic performance, high-school preparation, and family resources beyond income and how they might interact with monetary conditions. On the other hand, local labor market conditions, as measured by the county unemployment rate, were an important determinant of attendance for the most recent group while it was not estimated to impact the decisions of the 1972 and 1982 samples.

The role of price is not limited to enrollment decisions. All of the models suggest that price plays an important role in how students choose *between* colleges. In 1972, a student was 53 percent less likely to choose a college that was \$1000 more expensive, all else equal based on observable characteristics. By 1992, this effect had decreased by one-third. The exceptions to this trend are low-income students. The point estimates from the conditional logistic choice models suggest that low-income students did not make the same gains as other students and remained as negatively influenced by price in 1992 as they were in 1972 when choosing between colleges. This lack of progress is somewhat surprising given the growth in state and federal need-based aid programs during the period although most of these programs are designed to increase access, not choice. However, given that resources vary substantially among colleges and that they are positively correlated with price, policymakers should be concerned now with not only *if* students are attending college but also *where* they are enrolling. To quote McPherson and Schapiro (1990): "Differences in the allocation of resources across institutions imply very different educational experiences for different groups." Furthermore, the returns to higher education have become more closely tied to the type of school attended (Hoxby and Long, 1999). Therefore, even if price were no longer a deterrent to enrollment, college costs will continue to play an important role in determining who receives the greatest benefits from their college investments and how income is distributed in this country.

For all groups, distance became a less important factor in choosing between colleges. Meanwhile, college quality became more important during the time span. Although educational resources were not found to significantly influence the college choices of students from the classes of 1972 and 1982, an additional \$1000 in instructional expenditures increased the likelihood of choosing a particular college 7.5 percent for the 1992 cohort. The importance of faculty quality and the student–college match also became more important. Among the 1992 graduates, individuals were more likely to choose a college with a median SAT score above their own suggesting that many strive to be with better students.

Another contribution of this paper is its demonstration of the conditional logistic choice model as an improvement in estimating the determinants of college choice. The model is able to exploit the incredible heterogeneity within the market for higher education while accounting for numerous college options and match-specific information between potential students and schools. The estimates allow one to explore many factors that could impact choices between difference college options. In addition, using the conditional logistic choice model does a better job predicting the relevant college alternative than using the closest public, 2-year college or the mean public, 2-year tuition in a state.

6. Further reading

The following references may also be of interest to the reader: McPherson and Schapiro, 1998; Peterson's Educational Center, 2000.

Acknowledgements

The author gratefully acknowledges valuable comments from Caroline Hoxby, Larry Katz, Claudia Goldin, Richard Murnane, John Willett, and Judith Singer.

References

Avery, C., Hoxby, C., 2004. Do and should financial aid packages affect students' college choices. In: Hoxby, C. (Ed.), College Decisions: How Students Actually make Them and How They Could, University of Chicago Press, Chicago.

Barron's Educational Series, various years. Barron's Profiles of American Colleges, Barron's Educational Series, Various Editions. Hauppauge, New York.

Behrman, J., Rosensweig, M., Taubman, R., 1996. College choice and wages: estimates using data on female twins. Review of Economics and Statistics 78 (4), 672–685.

Brewer, D.J., Eide, E.R., Ehrenberg, R.G., 1999. Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. Journal of Human Resources 34 (1), 104–123.

Cameron, S., Heckman, J., 1999. Can tuition policy combat rising wage inequality? In: Kosters, M.H. (Ed.), Financing College Tuition: Government Policies and Educational Priorities. The American Enterprise Institute Press, Washington.

Cass, J., Birnbaum, M., various years. Comparative Guide to American Colleges, Various Editions. Harper and Row, New York.

Dale, S., Krueger, A., 2002. Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables. Quarterly Journal of Ecnonomics 117 (4), 1491–1528.

Ellwood, D., Kane, T., 2000. Who is getting a college education? Family background and the growing gaps in enrollment. In: Danziger, S., Waldfogel, J., (Eds.), Securing the Future. Russell Sage, New York.

Hoxby, C., 1997. The changing market structure of U.S. higher education. Harvard University Mimeo, Unpublished.

Hoxby, C., Long, B.T., 1999. Explaining rising income and wage inequality among the college-educated. National Bureau of Economic Research Working Paper No. 6873, January.

Kane, T., 1995. Rising public college tuition and college entry: How well do public subsidies promote access to college? National Bureau of Economic Research Working Paper No. 5164, July.
Leslie, L., Brinkman, P., 1989. The Economic Value of Higher Education. Macmillan Publishing Company, New York.
Manski, C., Wise, D., 1983. College Choice in America. Harvard University Press, Cambridge.
McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York.
McFadden, D., 1979. Quantitative methods for analyzing travel behavior of individuals: some recent developments. In: Hensher, D., Stopher, P., (Eds.), Behavioural Travel Modelling. Croom-Helm, London.
McPherson, M., Schapiro, M.O., 1990. Selective Admission and the Public Interest. College Board, New York.
McPherson, M., Schapiro, M.O., 1998. The Student Aid Game. Princeton University Press, Princeton.
Monks, J., Ehrenberg, R., 1999. U.S. news & world report's college rankings: Why they do matter. Change, November/December.
Peterson's Educational Center, 2000. College and University Almanac 2001. Thomson Learning, Stamford.
Rouse, C., 1994. What to do after high-school: the two-year versus four-year college enrollment decision. In: Ehrenberg, R.G. (Ed.), Choices and Consequences: Contemporary Policy Issues in Education. ILR Press, Ithaca.
Rouse, C., 1995. Democratization or diversion? The effect of community colleges on educational attainment. Journal of Business Economics and Statistics 13 (2), 217–224.