Volume Title: Annals of Economic and Social Measurement, Volume 5, number 4

Volume Author/Editor: Sanford V. Berg, editor

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/aesm76-4

Publication Date: October 1976

Chapter Title: An Empirical Investigation of Factors Which Influence College-Going Behavior

Chapter Authors: Meir G. Kohn, Charles F. Mansk, David S. Mundel

Chapter URL: http://www.nber.org/chapters/c12702

Chapter pages in book: (391 - 419)

Annals of Economic and Social Measurement, 5/4, 1976

AN EMPIRICAL INVESTIGATION OF FACTORS WHICH INFLUENCE COLLEGE-GOING BEHAVIOR

BY MEIR G. KOHN, CHARLES F. MANSKI, AND DAVID S. MUNDEL*

This report describes a theoretical and empirical model of student behavior that will help forecast enrollment patterns. In the model, actual college enrollments are the result of decisions made both by college administrators and by prospective students. The administrators determine a set of feasible alternatives for the students, who then select a "best" college. The student's decision problem is separated into three successive stages: (1) for each available college, the choice of whether to commute or to live on campus, should that college ultimately be chosen; (2) the choice of the "best" college available, given the residency decision; (3) the choice of whether to enroll at this "best" college or not at all. At each stage of the student's decision problem, we assume that he maximizes a utility function defined over the relevant alternatives. We use McFadden's conditional logit maximum likelihood procedure to estimate the parameters of these utility functions. Prior to estimation, several data-related problems were resolved. Our results regarding the impact of price and academic quality on student decisions indicate that both tuition and room and board charges had a lower effect on the decisions of students from higher-income families. We also found that the effect of parental education on college-going behavior declines with increasing family income.

I. INTRODUCTION

Discussion of higher education policy has been hampered in the past by an inability to predict with much confidence the effects on student behavior of proposed policies. How do federal and state programs of institutional or student support affect a prospective student's decision of whether to enroll and where to enroll? How will the location of new colleges affect these decisions? What might the impact be of proposed tuition increases in public institutions or the expected closure of particular colleges and universities? This report describes our efforts to develop a theoretical and empirical model of student behavior that will help to answer these and similar questions.

In our model, actual college enrollments are the result of decisions made by both college administrators and by prospective students. The administrators, through offers of admission and of financial aid, determine a set of feasible alternatives for the students, who then select a "best" college. Enrollment follows if this best alternative is more attractive than the various possibilities other than college, such as technical education, the armed services, or immediate employment. Colleges offer admission on the basis of relative academic merit among

* The research treated in this report was supported by grants from The Ford Foundation, the College Entrance Examination Board, the U.S. Department of Health, Education, and Welfare, and the National Science Foundation in a grant to The Rand Corporation. The conclusions of the report are not necessarily those of The Rand Corporation or the National Science Foundation.

The authors of this report are Rand consultants. M. G. Kohn is a Lecturer, Department of Economics, Faculty of Agriculture at The Hebrew University in Rehovoth; C. F. Manski is presently Assistant Professor of Economics, School of Urban and Public Affairs, Carnegie-Mellon University; and D. S. Mundel is Assistant Professor of Public Policy, John F. Kennedy School of Government at Harvard University.

The views expressed are those of the authors and do not, in any way, represent those of the sponsoring organizations. The authors wish to thank Stephen Carroll of Rand: Stephen Hoenack, the University of Minnesota; and Daniel McFadden, the University of California at Berkeley for their helpful comments on and suggestions for our work.

those that apply and financial aid on the basis of relative need among those admitted.[1] The student's evaluation of a given college will be based on what he perceives to be its benefits and costs. Our specification recognizes that college is both an investment and a consumer good and that the taste for college may vary with individual background. We estimate a model of college behavior, and use this to impute the set of feasible college alternatives for each student. The model of college choice is then estimated by taking the actual alternative chosen as the one preferred to all others in the feasible set. Finally, we estimate a model of college-going behavior in which the student compares the "best" college available with the alternative of not going to college at all. Our results, in general, corroborate existing beliefs about the selection of students by college administrators and the selection of colleges by students. More important, they bring us closer to our ultimate goal of being able to forecast the effect of proposed federal and state policies on the number and composition of enrollments and on the distribution of students among institutions.

II. Previous Research on the Demand for Higher Education

The early empirical research in this area included a number of longitudinal studies of high school students.[2] These studies confirmed that the important factors influencing enrollment are family income, parental education, high school peer relationships (and tracking), and the proximity of a college to the student's home. Unfortunately, these studies had nothing to say on the absolute or relative magnitude of the various effects. For instance, L. L. Medsker and J. W. Trent found that a low-cost, nearby college was an important stimulator of enrollment, but they did not define or quantify "low-cost," "nearby," or "important."

Later studies by economists[3] focused mainly on the effect of income and price; Stephen Hoenack, however, did consider the effects of admission policy. The major deficiencies of all of these studies were the high level of aggregation, with resulting misspecification of crucial variables, and the lack of attention given to admission policy.

[1] Other factors, such as regional distribution of students and student heterogeneity, also enter the college admission and financial processes. In our model, however, we ignore these decision criteria for reasons of simplification.

[2] See L. L. Medsker and J. W. Trent, *Beyond High School: A Study of 10,000 High School Graduates*, Center for Research and Development in Higher Education, University of California, Berkeley, 1967; J. Flanagan, *et al.*, *The American High School Student*, Final Report of Coop Research Project No. 635, U.S. Office of Education, University of Pittsburgh, 1964; and U.S. Bureau of the Census, *Factors Related to High School and College Attendance: 1967*, Series P-20, No. 185, July 11, 1969.

[3] See R. Campbell and B. Siegal, "The Demand for Higher Education in the United States, 1919–1964," *American Economic Review*, Vol. 57, No. 3, July, 1967, pp. 482-494; H. Galper and R. M. Dunn, Jr., "A Short-Run Demand Function for Higher Education in the United States," *Journal of Political Economy*, Vol. 77, No. 5, September-October 1969, pp. 765–777; P. Feldman and S. Hoenack, "Private Demand for Higher Education in the United States," *The Economics and Financing of Higher Education in the United States*, The Joint Economic Committee, 1969; A. Corazzini, D. J. Dugan, and H. Grabowski, "Determinants and Distributional Aspects of Enrollment in the U.S. Higher Education," *Journal of Human Resources*, Vol. 7, No. 1, Winter 1972, pp. 39–50; and G. W. Barnes, E. W. Erickson, W. Hill, Jr., and H. S. Winokur, Jr., "Direct Aid to Students: A 'Radical' Structural Reform," Inner City Fund, June 1972.

The most ambitious study to date has been done by R. Radner and L. S. Miller.[4] The general approach and specification of their study are similar to our own, and their report has given important direction to our work. But several data weaknesses and assumptions limit the potential utility of their results. First, although Radner and Miller used the same student data source as we used (the SCOPE sample), they had access only to a small subset of the available data. Second, the institutional alternatives were specified as broad categories of institutions (public four-year college, public two-year college, etc.) rather than as the individual schools themselves. The resultant averaging of institutional variables may have weakened their analysis. Third, their treatment of nonenrollment—assigning this option a price of zero and a Scholastic Aptitude Test (SAT) score of 344.4 (i.e., the average of nonenrollees)—is not very appealing, given the implicit assumption that "nonenrollment" is simply an institutional alternative very similar to a low-cost, low-quality school. Fourth, Radner and Miller assumed that within a given distance all students were commuters. Fifth, the only variables that entered into their analysis were student ability and income, institutional price, and average student ability. This prevented them from addressing the effect of family background—e.g., parental education—on college-going behavior. Overcoming the constraints imposed by each of the weaknesses we perceive in the Radner and Miller effort has been an important source of direction in our own effort.

III. A Model of College Choice

The allocation of students to colleges is the result of a process of mutual selection. This process is influenced by the actions of the government at various levels as well as by the secondary school systems that prepare the prospective student and provide him with information and guidance in choosing a college. We have not attempted to model this very complex system in its entirety; rather, we have concentrated on the behavior of the individual student himself. The behavior of colleges was treated only insofar as it was needed to provide inputs for our model of student decisionmaking.

Conceptually, the student's decisionmaking problem may be broken down into three successive stages: (1) for each available college, the choice of whether to commute or to live on campus, should that college finally be chosen; (2) the choice of the best college available, given the residency decision; (3) the choice of whether to enroll at this best college or not at all. A college is available to the student if he would be admitted to that college were he to apply. The effective cost of attending an institution is determined by the institution itself, by government, and by private groups through the setting of tuition and living costs and through the distribution of financial aid. Students vary in ability, location, income, and family background so that the colleges available, the costs and benefits of a given college, and the alternatives to going to college will be different for each student.

[4] Reported in R. Radner and L. S. Miller, "Demand and Supply in U.S. Higher Education: A Progress Report," *American Economic Review—Papers and Proceedings*, Vol. 60, No. 2, May 1970; and L. S. Miller, *Demand for Higher Education in the United States*, Working Paper No. 34, Economic Research Bureau, State University of New York, May 1971.

The limitations of existing theory, particularly with respect to college-going behavior, and the need for empirical tractability force us to adopt a number of simplifying assumptions. The mutual selection process of colleges and students is assumed to be recursive rather than simultaneous: all colleges make offers of admission and financial aid in advance of the student's decision, and these offers remain in effect throughout the decisionmaking period. The student then selects a college, and this selection is binding: no recontracting is allowed and, typically, the market does not clear. We assume that the student has perfect information about colleges. We do not consider part-time enrollment or simultaneous enrollment at more than one institution.

Clearly, the student will not actively consider all possible alternatives, and he may actually apply to very few colleges. There is an all important element of self-selection in this process: the student will not apply to those colleges that he considers inferior, too expensive, or unlikely to admit him. In our model, we include in the feasible set all those colleges to which the student *might* have applied and to which he would have been admitted; the process of self-selection and active selection are treated as one.

The structure of our behavioral model is really quite simple. Given a feasible set of college alternatives, the student ranks these in order of preference and compares the best of them with the alternative of not entering college at all. Enrollment follows if the best college alternative is preferred to not entering college. In the following two subsections, we present the formal structure of the model and discuss its estimation.

The Formal Structure

Let A_i be the set of college-residency alternatives available to a given student, i; and let B_i be his set of alternatives other than college. His choice set C_i is defined as the union of A_i and B_i. We assume that choice is rational, so that if c is the alternative actually chosen from C_i, then there is no c' in C_i strictly preferred to c.

Each student is described by a vector of characteristics Z_i, each college alternative by a vector of characteristics X_a, and each other alternative by a vector of characteristics Y_b. We assume that the student's behavior is consistent with his possessing a continuous stochastic utility function over the elements of his choice set.[5] Hence, the utility to student i of college alternative $a \in A_i$ is given by

(1) $$U_{ia} = U_A(Z_i, X_a, \theta, \varepsilon_{ia}),$$

where θ is an n-vector of parameters, and ε_{ia} a vector of random elements; likewise, the utility of $b \in B_i$ is given by

(2) $$U_{ib} = U_B(Z_i, Y_b, \Psi, \varepsilon_{ib}),$$

[5] The stochastic nature of the utility function might be attributed to omitted variables or variations in taste. For a fuller discussion, see Charles F. Manski, "Analysis of Qualitative Choice," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1973, Chapter 1.

where Ψ is a vector of parameters and ε_{ib} a vector of random elements. The choice of c from C_i implies

$$U_{ic} \geq U_{ic'} \qquad \text{for all } c' \in C_i. \tag{3}$$

Estimation

Our ability to estimate the parameters of the utility function is limited by the nature of our observations. For each student, we know whether or not he went to college and, if so, which college he chose. We are also able to impute to him a set of feasible college alternatives A_i. We do not, however, have any information on the student's set of alternatives other than college, B_i. Because of this deficiency, we have separated the student's decision into two stages: in the first stage, he chooses the best alternative in each of the sets A_i and B_i; and, in the second, his choice between the two "winners" determines whether or not he goes to college.

The college-choice model. Since we have observations on the choice among alternatives in A_i (for those students that did go to college), we are able to estimate the parameters of U_A directly.

We assume that U_A is linear in parameters with an additive disturbance:

$$U_{ia} = V_1(Z_i, X_a) \cdot \theta_1 + V_2(Z_i) \cdot \theta_2 + \varepsilon_{ia'} \tag{4}$$

where V_1 is an m-vector valued-function, V_2 is an $(n-m)$-vector-valued function; θ is partioned into θ_1 and θ_2 accordingly; and ε_{ia} is a scalar random variable. The element $V_2(Z_i) \cdot \theta_2$ represents the utility derived from college in general by individual i in a way that does not depend on the qualities of a given college. The choice of a from A_i implies

$$\begin{aligned} &V_1(Z_i, X_a) \cdot \theta_1 + V_2(Z_i) \cdot \theta_2 + \varepsilon_{ia} \\ &\quad \geq V_1(Z_i, X_{a'}) \cdot \theta_1 + V_2(Z_i) \cdot \theta_2 + \varepsilon_{ia'}, \qquad \text{for all } a' \in A_i \end{aligned} \tag{5}$$

or

$$(V_1(Z_i, X_a) - V_1(Z_i, X_{a'})) \cdot \theta_1 \geq \varepsilon_{ia'} - \varepsilon_{ia}, \qquad \text{for all } a' \in A_i. \tag{6}$$

It should be clear from this that the choice *among* colleges does not enable us to identify and estimate the vector of parameters θ_2.

In order to estimate the parameters θ_1 by the maximum likelihood principle, it is necessary to specify a joint probability distribution for the random variables ε_{ia}. Unfortunately, this specification must be made on the grounds of computability since only one probability distribution is known to lead to a likelihood function of any simplicity. That distribution is the Weibull distribution:

$$\text{Prob}\,(\varepsilon \leq T) = e^{-\alpha e^{-\beta T}}, \ \alpha > 0, \ \beta > 0 \text{ constants.} \tag{7}$$

if ε_{ia} and $\varepsilon_{ia'}$ are independent and identically distributed with this distribution, their difference has a logistic distribution with parameter β:

$$\text{Prob}\,(\varepsilon_{ia} - \varepsilon_{ia'} \leq T) = \frac{1}{1 + e^{-\beta T}}. \tag{8}$$

Generalizing (8) to consider the joint distribution of the differences $(\varepsilon_{ia} - \varepsilon_{ia'})$ for all $a' \in A_i$ and recalling (6), we obtain

(9) Prob (a chosen from A_i) =

$$\text{Prob}\,(\varepsilon_{ia'} - \varepsilon_{ia} \leq [V_1(Z_i, X_a) - V_1(Z_i, X_{a'})] \cdot \theta_1 \text{ for all } a' \in A_i)$$

$$= \frac{1}{1 + \sum_{a' \in A_i, a' \neq a} \exp(-\beta[V_1(Z_i, X_a) - V_1(Z_i, X_{a'})] \cdot \theta_1)}$$

This type of model—called a conditional logit model—was developed by Daniel McFadden, who has shown that an estimate of the parameters that maximizes the likelihood of the observed choices is under certain quite general conditions consistent and asymptotically normal.[6]

The college-going model. The set of alternatives other than college faced by individual i, B_i, is not available to us, but we can capture something of a student's decision on whether or not to go to college in the following way. Consider the function

$$S(Z_i, \phi, \delta) = \sup_{b \in B_i} U_{ib}. \tag{10}$$

S is a sort of "envelope" function giving the level of utility attainable by individual i when he chooses the best of the noncollege alternatives. It is assumed that S is linear in parameters with additive disturbance:

$$S(Z_i, \phi, \delta_i) = W(Z_i) \cdot \phi + \delta_i, \tag{11}$$

where W is a vector-valued function; ϕ is a vector of parameters; and δ_i is a scalar random variable.

If a^* is the best college-residency combination (the chosen element of A_i), then individual i will enroll at a^* if

$$U_{ia^*} \geq S(Z_i, \phi, \delta_i) \tag{12}$$

or

$$V_1(Z_i, X_{a^*}) \cdot \theta_1 + V_2(Z_i) \cdot \theta_2 + \varepsilon_{ia} \geq W(Z_i) \cdot \phi + \delta_i \tag{13}$$

or

$$V_1(Z_i, X_{a^*}) \cdot \theta_1 + [V_2(Z_i) \cdot \theta_2 - W(Z_i) \cdot \phi] \geq \delta_i - \varepsilon_{ia^*}. \tag{14}$$

Before discussing the stochastic specification and estimation of the parameters of the college-going model, we must point out three difficulties with equation (14). First, the functions $V_2(Z_i)$ and $W(Z_i)$ are both defined over the same variables Z_i so that the coefficients θ_2 and ϕ are not identified with respect to one another. We therefore replace the second expression on the left-hand side of equation (14) with the reduced form $Y(Z_i) \cdot \lambda$. Second, the best element of A_i is not known for students who do not go to college so that $\hat{a}$, the predicted best college, must be used for these observations. Using a^* for students who do go to college would cause a systematic measurement bias (since $V_1(Z_i, X_{\hat{a}}) \cdot \theta_1 \geq$

[6] D. McFadden, "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1973.

$V_1(Z_i, X_{a^*}) \cdot \theta_1)$ so that in order to avoid this type of bias $\hat{a}$ is used for all observations. Therefore, $V_1(Z_i, X_{a^*})$ is replaced by $V_1(Z_i, X_{\hat{a}})$ in (14). Third, θ_1 is not known, but we do have an estimate $\widehat{\beta\theta}_1$ of $\beta\theta_1$. Thus, inequality equation (14) is replaced by

$$\gamma \hat{U}_{1i} + Y(Z_i) \cdot \lambda \geq \eta_i, \tag{15}$$

where $\hat{U}_{1i} = V_1(Z_i, X_{\hat{a}}) \cdot \widehat{\beta\theta}_1$; γ is introduced to allow for the scale factor β in the estimate of θ_1; and η_i is a composite disturbance term.

By assuming a logistic distribution for the disturbance, it is possible to estimate the coefficients γ and λ using a logit model and a maximum likelihood estimation procedure.[7]

IV. Empirical Specification and Results

In this section we describe the empirical application of the theoretical framework developed above. A discussion of our procedure for imputing sets of feasible college alternatives is followed by a presentation of the results of estimating the college-choice model and the college-going model.

Imputing the Set of Feasible College Alternatives

Our principal source of data, the SCOPE survey,[8] provided us with information on whether or not graduating high school seniors went on to college; and, if so, it gave the college of enrollment. The SCOPE survey does not tell what other colleges the student was admitted to or might have been considering. Clearly, without such information no inference can be drawn about student preferences over colleges. We were forced to make up this deficiency as best we could.

We made use of a number of auxiliary data sources in trying to reconstruct the set of feasible college alternatives for each student. These were the "Institutional Research File" of the American Council on Education,[9] the "Manual of Freshman Class Profiles (1965–67)" of the College Entrance Examination Board,[10] and a file of geographical data we gathered ourselves.

The procedure for imputing the feasible set involved three steps: (1) rejecting as many colleges as possible a priori; (2) predicting for the remaining colleges whether or not a student would have been admitted; (3) predicting the level and composition of financial aid. Each step is described below. The section ends with a discussion of a number of general problems arising in the estimation of this kind of choice model.

Rejection a priori. Since fewer than 18 percent of the SCOPE college-goers attended colleges further than 200 miles from their homes, we felt that little

[7] It can be shown that in the case of dichotomous choice, the stochastic specification is relatively unimportant and that a logit model will give a good approximation so long as the functional specification is sufficiently flexible.

[8] *School to College: Opportunities for Postsecondary Education*, The Center for Research and Development in Higher Education, University of California at Berkeley.

[9] J. Creager and C. Sell. *The Institutional Domain of Higher Education: A Characteristics File for Research*, ACE Research Reports, Vol. 4, No. 6, American Council on Education, Washington, D.C., 1969.

[10] College Entrance Examination Board, *Manual of Freshman Class Profiles* (1965–1967). New York, 1968.

information would be lost by excluding such colleges.[11] Furthermore, since very few students were observed to commute further than 60 miles, it seemed reasonable to exclude colleges with no facilities for residency on campus that were 60 miles or more away.

In these calculations, we used the straight-line distance between the student's high school and the college in question. Since students do not live at their high schools and since the straight-line distance may be a poor measure of the actual shortest route, our calculated distance is probably quite inaccurate for short distances but more reasonable for longer distances. We excluded from the sample all high schools having a high percentage of boarders because the high school–college distance would not serve as a proxy for home to college distances for boarders.

Admission. For each student, we took the set of all colleges not excluded on a priori grounds and simulated for each college in this set its admission decision regarding that student. In order to do this, we needed an estimate of the probability of admission of a given student to a given college. The results of our estimation are presented below. Using this estimated probability, it was possible to simulate the college's decision in the following way: a random number on the unit interval was generated; if this number was less than the calculated probability, the college was included in the feasible set; otherwise it was not. In this way we ensured that if the calculated probability of admission was, say, 0.8, we included the college in the feasible set with probability 0.8.[12]

The equations predicting the probability of admission were estimated using data from the CEEB's "Manual of Freshman Class Profiles (1966–67)." This provides records of percentages of applicants accepted by ability based on their Scholastic Aptitude Test (SAT) score and class rank, by sex, and by high school control categories for over 400 colleges. We based our estimation of admission probability on the 66 colleges whose tables show explicitly the interaction between SAT score and class rank in determining the frequency of admission.[13] We made no effort to distinguish open enrollment schools from those with discretionary admission policies.

We assumed that the frequency of admission was a function of the deviation of the applicant's SAT score from the median SAT score for all students enrolled at the college, of his high school class rank, of the applicant's sex, and of the type of high school he attended. In order to have a flexible specification, the continuous relative-SAT-score variable was introduced in a piecewise-linear form and interacted with the class rank categories.[14]

[11] This process rejected clearly inferior colleges from the student's choice set. If, however, the student attended a more distant school, the attended school was included in the choice set.

[12] The procedure we finally used was slightly different from this. It is discussed in Sec. IV.

[13] These 66 colleges (predominantly private four-year colleges) were not representative of the whole spectrum of American colleges. In order to test whether admission procedures differ systematically across colleges, we used less detailed data to estimate separate equations for over 150 colleges and then compared coefficients with college type. Some definite systematic variation was found, with public colleges relying less heavily on SAT scores for admission than private colleges. However, the differences were small enough to warrant use of the equation for the 66 colleges to forecast admission at all colleges.

[14] Although we realize that a probit or logit specification would have been superior, programming for such an algorithm had not been completed at this stage of the project.

The results of the estimation are presented in Table 1 and in Figure 1. The probability of admission increases monotonically with relative SAT score and with class rank. The appearance of the graphs in Figure 1 indicates that SAT score and class rank do interact other than additively: the interaction is greater for

TABLE 1

THE ADMISSION EQUATION

Variable			Coefficient	Standard Error
Class Rank-SAT Score Interaction				
Class Rank Quintile	Relative SAT Score[a]			
I	−1,200		−2.98	0.18
	−50		0.60	0.01
	50		0.77	0.01
	1,200		1.36	0.18
II	−1,200		−3.16	0.18
	−50		0.47	0.01
	50		0.65	0.02
	1,200		0.99	0.32
III	−1,200		−2.78	0.20
	−50		0.30	0.02
	50		0.52	0.02
	1,200		0.60	0.43
IV	−1,200		−2.49	0.28
	−50		0.13	0.02
	50		0.32	0.03
	1,200		1.12	0.57
V	−1,200		−1.93	0.35
	−50		0.00	0.03
	50		0.15	0.05
	1,200		1.26	1.00
Student Sex, SAT Type, and Sex of College Population				
Sex	Test Type	College Type		
Male	Verbal	All Male	0.11	0.02
Male	Verbal	Coed	0.18	0.01
Female	Verbal	All Female	0.07	0.01
Female	Verbal	Coed	0.08	0.01
Unknown	Verbal	Coed	0.13	0.01
Male	Math	All Male	0.00	0.02
Male	Math	Coed	0.09	0.01
Female	Math	All Female	0.13	0.01
Female	Math	Coed	0.08	0.01
Unknown	Math	Coed	0.00	(normalization)
College control				
Private college			0.02	0.00
Public college			0.00	(normalization)

Notes: The range of values for all variables is [0, 1]. R^2 (transformed variables) = 0.79561; $F = 813.57887$. Standard error of the regression (transformed variables) = 1.03070.

[a] Student SAT minus average SAT.

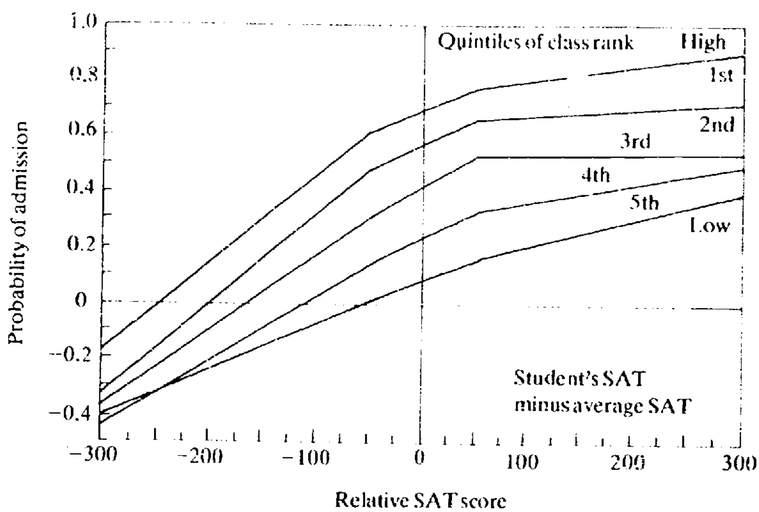

Figure 1 Probability of admission versus relative SAT score (using normalized dummy variables, by class rank in quintiles)

students of lower class standing. The regression lines for the first and second quintile students are almost parallel, indicating an equal class rank effect for all SAT levels. However, for the lower quintiles, class rank appears to exercise a much greater effect on admission probability when the student's SAT is close to the college's median score.

The asymmetry of the admission-probability curves about the origin may be of some interest. With our piecewise-linear specification, all five curves are convex rather than S-shaped. It seems that increments of SAT score have a greater effect on admission probability when the student is below the college's median than when he is above it. This may be a reflection of a skewness to the right in the ability distribution of the applicants to any college.

Financial aid. In order to evaluate the actual cost of attending a college, we attempted to predict the level and composition of the financial aid that would be offered to each student by each college. There are two quite different ways of doing this: (1) by predicting the expected level of aid using a regression model; (2) by using a discrete probability model to find the probability of receiving aid, estimating the level of aid conditional on aid having been received, and then using a simulation procedure to determine the aid offered to each student by each college. We were unsuccessful in our attempt to obtain accurate forecasts of financial aid using either approach.

Our model of aid determination assumes that colleges made awards on the basis of a student's ability and income, relative to that of its median applicant for admission. The coefficients of our estimated equations were both small and generally insignificant. Among the possible causes of this result are: (1) general data inadequacies, (2) lack of a good specification of the aid distribution process, and (3) the possibility that colleges may have acted capriciously.

Given our failures, we did not include a financial aid variable in our final specification for the college-choice model, although we did attempt to observe some of the possible effects of financial aid through the interaction of tuition and student income, and this is reported below.

Some problems associated with estimation. The imputed choice set of feasible alternatives will usually differ from the actual choice set, and this may be the source of a problem. The inclusion of colleges absent from the true choice set but inferior to the chosen college will have no adverse effect on estimation since the choice would have been the same even if they had been present. On the other hand, the inclusion of superior or preferred colleges that do not appear in the true choice set will make it seem that a college with less of the desired qualities is chosen over one with more; this will impart a negative bias to the coefficients of these qualities. For instance, if our admission simulator "admits" the student to a college of high academic quality and the student is observed to choose a college of low academic quality, it will seem that he dislikes academic quality; whereas, in fact, he might have preferred the college of high academic quality had it really been a feasible alternative.

This problem is aggravated if, as we believe, students have a positive preference for academic quality up to some point but then begin to dislike schools in which academic levels are too far above their own. It then becomes almost impossible to distinguish between such an effect and the bias introduced by the simulation of admission. The steps we have taken in an attempt to minimize this bias will be described in our discussion of the estimation of the college-choice model.

The consequences of choice-set imputation, both on the estimated coefficients of the utility function and on the associated reported standard errors, are still not fully understood. An alternative and theoretically sounder procedure would be to maximize a likelihood function in which the choice set, as well as choice from a given set, is probabilistic. That is, let a^* be student i's chosen college; $\mathscr{A}$, a universe of possible choice sets, each including a^*; and $P_i(A)$ the probability that $A \in \mathscr{A}$ is the choice set faced by student i. Then, the marginal probability that a^* is chosen is

$$p_i(a^*) = \sum_{A \in \mathscr{A}} P_i(A) \operatorname{Prob}(U_{ia^*} \geq U_{ia}, \text{all } a \in A), \tag{16}$$

and the likelihood function is

$$L = \prod_i p_i(a^*). \tag{17}$$

We considered estimation of θ through maximization of L but rejected it as a practical procedure because of its prohibitive computational expense. If L is accepted as the correct likelihood function, then the function we did maximize should be viewed as a quasi-likelihood function, one in which a "certainty-equivalent" choice set replaces the actual distribution of possible sets.[15]

[15] Lack of data regarding choice-set composition is a common occurrence in current data sources. A formal analysis of the consequences of choice-set imputation procedures and other means of handling the problem would be a useful contribution.

Another problem arises because the conditional logit model assumes independently distributed disturbances. To understand the problem, consider a choice between a private university and a community college, with the probability of choosing the former, 0.6, and that of choosing the latter, 0.4. Now introduce a second identical community college into the choice set. In the conditional logit model the probability of choosing the university becomes 0.43; for choosing each community college, 0.29. Thus, the total probability of choosing *some* community college now exceeds that of choosing the university—the reverse of the original situation.

The source of the problem in the above example is an improper assumption of independent disturbances in the stochastic utility function. Given that the two community colleges are identical, their disturbances (which are due to omitted variables) should be identical. In practice, alternatives are rarely identical but may be "similar," in which case the problem persists if their similarity extends to unobserved dimensions. In order to minimize the amount of unobserved similarity among college alternatives, and hence to make the assumption of independent disturbances plausible, we augmented the set of college attributes with a set of college-type dummy variables. It was hoped that by using dummies to pick up choice-relevant features that each college type shares (which were not included among our explicit college attributes), the remaining unobserved attributes would be independently distributed.

Our imputed choice-sets contained a large number of colleges (most included from 50 to 150 schools) so that computational considerations forced us to find some way of decreasing the number. We decided to select ten colleges at random from the imputed choice set of each student and to estimate using this random subsample plus the college actually chosen as the choice set. We found that changing the randomization (i.e., drawing different subsample choice sets for each student) had almost no effect on the estimation results.

Estimating the Residency-Choice Model

Each college in a student's choice set actually embodies two alternatives, corresponding to the two residency choices of commuting or living on campus.[16] Recall the college utility function given in equation (4);

$$U_{ia} = V_1(Z_i, X_a) \cdot \theta_1 + V_2(Z_i) \cdot \theta_2 + \varepsilon_{ia}.$$

In this equation "a" should be taken as a college-residency combination, not simply as a college. Because for a given college many X attributes, and consequently many V_1 functions, are invariant to the residency option, it becomes feasible to separate estimation of the part of the utility function relevant to residency choice from that of the remainder of the function. Prior estimation of the residency model was performed; it was useful because it allowed us to forecast a student's hypothetical residency choice at each college and hence to trim the size of the choice set in later estimations.

[16] In some cases, more than two such choices may be available. However, for college freshmen, this is rarely so.

The residency-choice model was formulated as a dichotomous logit model and was estimated using a subsample of the SCOPE survey (students attending a college at which they could have chosen either alternative). The results are presented in Table 2 and Figure 2.

The first four coefficients in Table 2 are for the interaction of distance and income (measured in log to base 10). If X is a distance value and Y is an income value (measured in log to base 10), then the following formulas are used to determine a probability of campus residency for that distance-income combination. Variable $(1)=(100-X)(5-Y)/500$; $(2)=X(5-Y)/500$; $(3)=(100-X)Y/500$; and $(4)=XY/500$. It is difficult to interpret the coefficients of a piecewise-linear specification, but Figure 2 should make this easier. Tracing the probability of campus residency as a function of distance for three income levels, the graph shows that the probability increases with distance and is higher at each distance for the student with higher income.

TABLE 2
ESTIMATED COEFFICIENTS OF THE RESIDENCY CHOICE MODEL

Variable		Range of Values	Coefficient	Standard Error
Distance-Log Income Interaction				
Distance	Log (income)			
0	0	0–1	−16.59	1.73
100	0	0–1	−8.680	0.90
0	5	0–1	−0.6277	0.40
100	5	0–1	10.81	0.70
Percentage Dormitory Capacity[a]		0–100	0.01929	0.0023
Student Sex (if female)		0–1	−0.04789	0.11
Residency Preference[b] (1 = on-campus)		0–1	1.470	0.16

[a] This variable may reflect the campus character of the school. It also may be interpreted statistically as a mixture probability, where the kernel distribution is the choice probability conditional on college and student attributes and the mixing distribution is the distribution of such attributes.

[b] The SCOPE question asked the student what his residency decision would be if money were not a problem. Thus, it should capture the pure preference for living style.

The interpretation of the remaining coefficients is straightforward. The probability of campus residency increases with percentage dormitory capacity, is very slightly higher for males than for females, and is higher for those students who, all else equal, would prefer to live on campus.[17]

[17] It is, of course, not surprising that choices should reflect student's pure preference for residency mode. It is interesting, though, how small a role pure preference plays compared to the economic variables of distance and income.

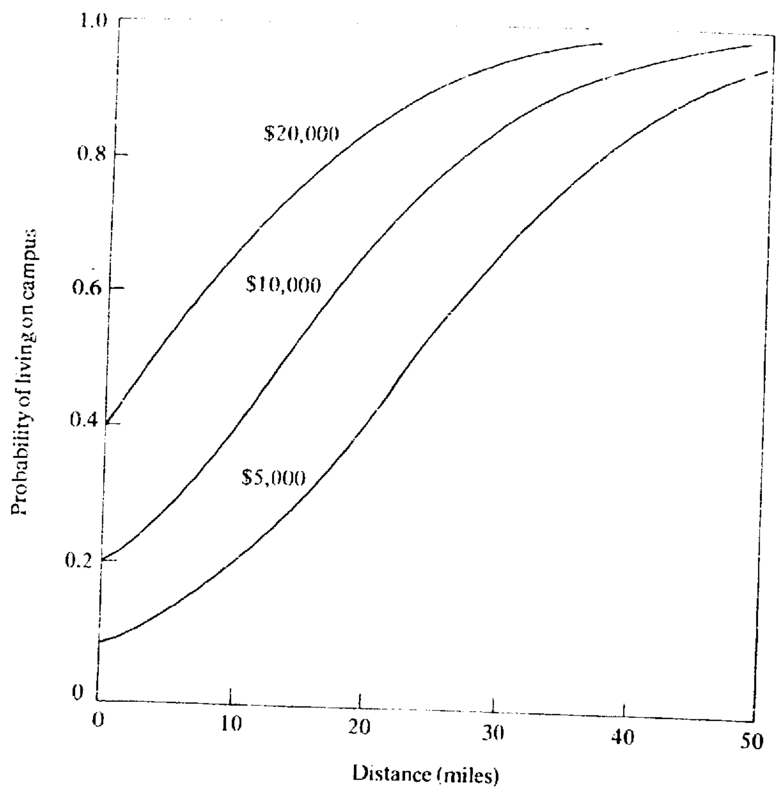

Figure 2 Probability of living on campus (for males attending a college with 50 percent dormitory capacity and who prefer to live on campus)

This model of residency choice may be of some interest beyond its use here, since an understanding of this aspect of student behavior is needed in resolving certain issues in higher education policy. As an example, one could cite the current controversy over the location of community colleges. It is often said that such colleges should be built in the city to provide more readily available education for the urban poor. The argument put forward for this policy—one of bringing the college to the student—is that commuting over short distances results in lower out-of-pocket costs than campus residency. Therefore, it is asserted, the availability of nearby community colleges will encourage enrollment by low-income students. There is, however, another possible solution; to build campus-type community colleges on low-cost, non-urban land and to subsidize the on-campus living costs of poor students. If it is found that students prefer on-campus residency to commuting, then the latter alternative, not the former, would provide the better solution to the college location problem.

Estimating the College-Choice Model

The college-choice model, apart from the residency-choice component, was estimated using a subsample of the SCOPE survey that included students graduating from Illinois high schools who went on to enroll in some college. This subsample contained some 3,000 observations. These observations were divided into three income strata of approximately equal size, and the model was estimated separately for each stratum. This allowed full interaction between income and all other variables.

The first subsection below describes the income variable used to carry out this stratification. The next four subsections deal with the four types of variable included in the specification of this model: (1) variables relating to the cost of a college, (2) those relating to academic quality, (3) those relating to the "quality of life," and (4) dummy variables for college type.

The model was reestimated on a similar subsample of seniors graduating from North Carolina high schools to see whether behavior was reasonably uniform in different states and hence whether our estimated coefficients would be of any use in predicting behavior outside the sample. The results are described in the final subsection.

Income variable. The difficulties associated with the reporting of family income in the SCOPE survey have been discussed at length in an unpublished paper by L. S. Miller. We decided to use family income reported by parents minus a deduction for family size as our income variable. We were forced to predict family income on the basis of other variables, however, because of the large number of observations for which these data were missing. Because our income predictor is quite similar to Miller's[18] and is of little intrinsic interest, we do not report the details here. The deduction for family size was made on the basis of the family size allowance used by the College Scholarship Service (CSS). The CSS bases its allowance on the impact of additional children on the "moderate income level" of the Bureau of Labor Statistics. We used this deduction to control for effects of children on disposable income available for college expenses.

Cost variables. We view the student as comparing different colleges on the basis of their costs and benefits, and our coefficients are intended to capture the revealed tradeoff between these costs and benefits.

The principal cost variable is college tuition. Although financial aid should be deducted from tuition to give *net* cost, our failure to find a satisfactory method of predicting financial aid prevented us from doing this.[19] As a substitute, we introduced a term quadratic in tuition, the rationale being that the burden of tuition would rise less than linearly with the amount, particularly for the poor students, since high-tuition colleges are also the ones most likely to offer financial aid. The coefficients of variables 1 and 2 in Table 3 seem to bear this out. Figure 3 shows the disutility of tuition for the three income levels: the curves become lower and flatter as income rises, exhibiting decreasing curvature as well.

If the student "chooses"[20] to live at home and to commute to a particular college, then variable 3 is set equal to the distance from home to college;

[18] The only major difference is our use of an additional variable: the median reported income for parents of students in the particular high school. Income was predicted even for those observations where it was reported because we felt that the predicted value was probably a better measure of permanent income. The use of this variable as the basis for a crude stratification rather than as a continuous variable in the equation itself reduces the importance of precise prediction.

[19] It also prevented us from comparing the value of different types of aid (fellowship, loan, work-study), something we had hoped to do.

[20] The result of our simulation of the residency decision.

TABLE 3

ESTIMATED COEFFICIENTS OF THE COLLEGE-CHOICE MODEL

(Illinois data)

Variables	Units (approximate range of values)	Coefficients, by Income Stratum <\$7,860	\$7,860–11,470	>\$11,470
Cost				
1. Tuition, \$	100	−0.397	−0.285	−0.105
	(0–40)	(0.0265)[a]	(0.0247)	(0.0196)
2. (Tuition)^2, \$	$(100)^2$	0.00843	0.00620	0.00242
	(0–1600)	(0.000882)	(0.000907)	(0.000560)
3. Distance from home to college, if commuting	Miles	−0.0314	−0.0532	−0.0254
	(0–200)	(0.00749)	(0.00946)	(0.0124)
4. Room and board, if living on campus, \$	100	−0.186	−0.136	0.00589
	(0–30, approx.)	(0.0143)	(0.0139)	(0.0150)
Academic Quality				
5. Average student ability at college	100 SAT points	1.08	1.02	1.63
	(2–8)	(0.123)	(0.113)	(0.107)
6. Ability difference[b]	(100 SAT points)^2	−0.263	−0.415	−0.298
	(0–36)	(0.0545)	(0.0544)	(0.0498)
7. College revenue per student, \$	1,000	−0.0197	−0.0694	−0.0529
	(1–4)	(0.0337)	(0.0293)	(0.0254)
8. Breadth of offering, index		0.125	0.183	0.0934
	(1–13)	(0.0254)	(0.0261)	(0.0242)
Quality of Life				
9. Coeducational college		−0.593	−0.789	−0.698
	(0, 1)	(0.183)	(0.149)	(0.127)
10. Dormitory capacity, %		−0.00109	−0.00628	−0.00108
	(0–100)	(0.00266)	(0.00231)	(0.00206)
Dummies				
11. Private university[c]	(0, 1)	0	0	0
12. Private 4-year college		−2.33	−1.77	−1.55
	(0, 1)	(0.203)	(0.166)	(0.145)
13. Private 2-year college		−4.06	−2.26	−0.843
	(0, 1)	(0.444)	(0.382)	(0.374)
14. Public University		−1.65	−1.19	−0.0579
	(0, 1)	(0.207)	(0.123)	(0.144)
15. Public 4-year college		−2.42	−1.83	−0.675
	(0, 1)	(0.232)	(0.200)	(0.187)
16. Public 2-year college		−1.29	−0.361	0.0279
	(0, 1)	(0.332)	(0.322)	(0.342)
Number of observations		997	990	1,028
Maximum log-likelihood		−1,410	−1,650	−1,770
Log-likelihood for $\hat{\theta} = 0$		−2,280	−2,310	−2,420

[a] Numbers in parentheses are asymptotic standard errors.
[b] (Average SAT score at college-student SAT)^2.
[c] Normalization.

otherwise it is set equal to zero. Similarly, if the student chooses to live on campus, variable 4 is set equal to the cost of room and board at that college; otherwise it is set equal to zero. Thus, a college is "debited" with transportation costs, as represented by the distance, if the student commutes, and with the cost of room and board if he lives on campus.

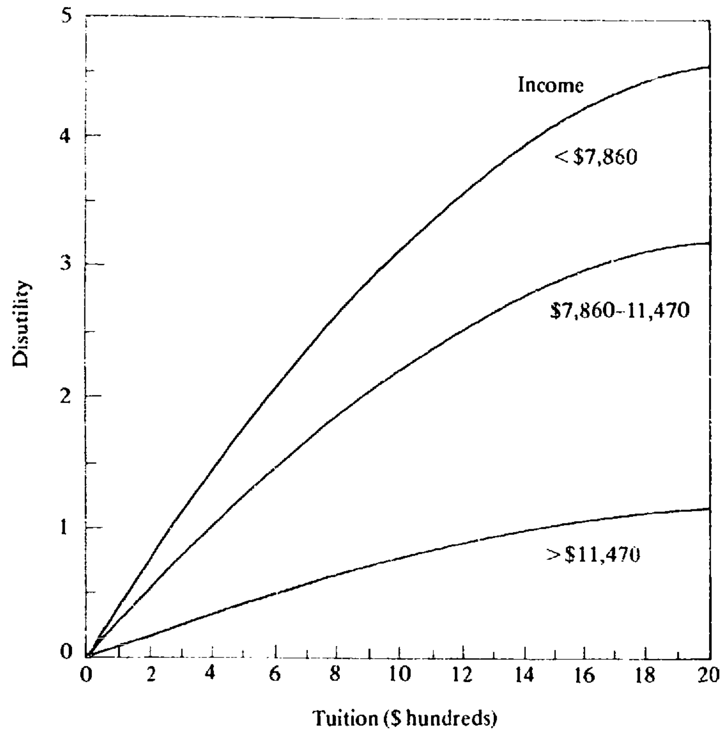

Figure 3 Disutility of tuition

The room-and-board variable may also be interpreted as an indicator of "quality of life" in addition to its role as a cost variable: Higher charges for room and board may, in some cases, indicate better living conditions. The coefficients seem to support this conclusion since the effect of a $100 increase in room-and-board fees has less impact than a similar increase in tuition (except at high levels of tuition). Furthermore, the coefficient is positive for the highest income stratum.

By comparing the changes in utility resulting from marginal changes in tuition and in the distance from home to college, it is possible to calculate the implicit evaluation, in money terms, of a mile traveled. For the low-income stratum this is about $0.05 per mile; for the middle-income stratum, about $0.11, and for the high-income stratum, about $0.17.[21]

Academic quality variables. The average Scholastic Aptitude Test (SAT) score for students attending a college was used as one measure of academic quality (variable 5). Although we believed that students would prefer colleges with higher average SAT scores, it seemed possible that a student would not wish to attend a school where the average ability was too far above his own. In order to capture this effect, we included variable 6, a measure of the distance of the student's ability

[21] This is calculated on the basis of 150 round trips per year. The marginal disutility of tuition taken at $1,000.

It should be noted that distance also enters as an explanatory variable in the residency choice model so that its full impact on student behavior is divided between the two models.

from that of the average student at the college. Thus, the total impact of average SAT is obtained by combining its effects through variables 5 and 6.

As we mentioned in our discussion of the problems associated with imputed choice sets, it is impossible to isolate this aversion for schools of too high an academic level from the bias caused by our simulation of college admission. We did, however, conduct some experiments to determine the sensitivity of the results. This was done by excluding altogether from the choice set those schools with a predicted admission probability below a certain level. These experiments had little effect on the coefficients of variable 5, but the coefficients of variable 6 fell somewhat in absolute value as the cutoff level was raised from 0 to 0.25 and almost disappeared as the cutoff level was raised again to 0.5. Other coefficients were minimally affected by these experiments. The coefficients reported in Tables 3 and 4 are for an admission cutoff level of 0.25, which we felt was probably sufficient to counter this bias. (This is, of course, only a guess.)

We had hoped that variable 7, college revenue per student, would be a proxy for educational expenditure and so an indicator of academic quality. However, it proved to be of little statistical or economic significance, and the coefficient had the "wrong" sign.

Variable 8, breadth of offering, is an index constructed by us from the list of fields, for each college, in which a bachelor's degree is offered. As expected, this variable had a positive coefficient showing that students preferred schools offering a wider choice of possible specializations. This preference seems to be stronger in the middle-income stratum than in the high and low strata.

Quality-of-life variables. Since we believe that students view college at least in part as a consumption good, we tried to capture the value of a college in this respect through variables representing the "quality of life" at the school.

Variable 9 is a dummy variable set equal to unity if the college is coeducational. Contrary to our expectations, the coefficients proved to be negative. Variable 10 is the dormitory capacity of the college as a percentage of enrollment; it was included as an indicator of the degree to which the school was a campus rather than a community institution. This variable had little influence on decisions. As we mentioned above, variable 4, fees for room and board, may also be interpreted as a quality-of-life variable.

College-type dummy variable. Variables 11 through 16 are a set of dummy variables for college control and type. Any behavioral interpretation of their coefficients is problematic because they capture the influence of a combination of unobserved variables. Our use of college dummies is similar to the use of mode-specific variables in transportation-choice studies.

Reestimation with North Carolina data. The coefficients of the college-choice model estimated from North Carolina data are presented in Table 4. Although there are differences from the Illinois results—most notably for variable 3—and while a classical statistical test for equality of the two sets of coefficients would fail, the similarity between the two sets of behavioral coefficients (variables 1 through 10) indicates that there may be considerable uniformity in the behavior of students in quite different geographical areas. It also suggests that our results are rather better than might have been anticipated in view of all the simulation and imputation of crucial variables.

TABLE 4

ESTIMATED COEFFICIENTS OF THE COLLEGE-CHOICE MODEL
(North Carolina data)

Variables	Units (approximate range of values)	Coefficients, by Income Stratum		
		<\$7.860	\$7.860–11.470	>\$11.470
Cost				
1. Tuition, \$	100	0.250	−0.174	−0.0952
	(0–40)	(0.0722)[a]	(0.0503)	(0.0440)
2. (Tuition)^2, \$	$(100)^2$	−0.0237	0.00250	0.00176
	(0–1600)	(0.00479)	(0.00271)	(0.00159)
3. Distance from home to college, if commuting	Miles	0.0239	0.0239	0.00758
	(0–200)	(0.00419)	(0.0100)	(0.0144)
4. Room and Board, if living on campus, \$	100	−0.162	−0.141	−0.0987
	(0–30, approx.)	(0.0121)	(0.0199)	(0.0192)
Academic Quality				
5. Average student ability at college	100 SAT points	0.837	1.89	2.13
	(2–8)	(0.0639)	(0.116)	(0.119)
6. Ability difference[b]	(100 SAT points)^2	−0.615	−0.549	−0.428
	(0–36)	(0.0406)	(0.0629)	(0.0574)
7. College revenue per student, \$	1,000	−0.0920	−0.112	−0.0786
	(1–4)	(0.0377)	(0.0542)	(0.0513)
8. Breadth of offering, index		0.0844	0.134	0.102
	(1–13)	(0.0153)	(0.0276)	(0.0277)
Quality of Life				
9. Coeducational college		0.795	0.634	−0.0299
	(0, 1)	(0.156)	(0.171)	(0.148)
10. Dormitory capacity, %		−0.00936	−0.0108	−0.00337
	(0–100)	(0.00152)	(0.00215)	(0.00219)
Dummies				
11. Private university[c]	(0, 1)	0	0	0
12. Private 4-year college		−1.19	−0.596	−0.0951
	(0, 1)	(0.276)	(0.353)	(0.354)
13. Private 2-year college		−1.31	0.724	1.26
	(0, 1)	(0.343)	(0.499)	(0.509)
14. Public university		0.000711	0.124	1.03
	(0, 1)	(0.247)	(0.265)	(0.267)
15. Public 4-year or 2-year college		−0.00576	0.244	0.778
	(0, 1)	(0.271)	(0.348)	(0.359)
Number of observations		1,623	749	760
Maximum log-likelihood		−2,810.	−1,250	−1,240
Log-likelihood for $\hat{\theta} = 0$		−3,860.	−1,790	−1,820

[a] Numbers in parentheses are asymptotic standard errors.
[b] (Average SAT score at college-student SAT)^2.
[c] Normalization.

Estimating the College-Going Model

The college-going model was estimated from a subsample of the SCOPE survey consisting of all Illinois high school graduates—both those who did enroll in some college (these were also used to estimate the college-choice model) and those who did not. The subsample contained about 7,000 observations, which

TABLE 5

ESTIMATED COEFFICIENTS OF THE COLLEGE-GOING MODEL
(Illinois data)

Variables	Approximate Range of Values	Coefficients, by Income Stratum: <$7,860	$7,860–11,470	≥$11,470
General				
1. Utility of best college	(−8 to +8)	1.08	0.876	0.973
		(0.0627)[a]	(0.0585)	(0.0688)
Father's education				
2. Some grade school	(0, 1)	−0.0346	−0.0487	0.588
		(0.213)	(0.339)	(0.770)
3. Finished grade school	(0, 1)	−0.0975	0.148	0.974
		(0.210)	(0.319)	(0.636)
4. Some high school	(0, 1)	0.363	0.400	1.21
		(0.209)	(0.310)	(0.596)
5. Finished high school	(0, 1)	0.601	0.537	1.46
		(0.230)	(0.313)	(0.592)
6. Some college (or other post high school)	(0, 1)	0.805	0.577	1.80
		(0.358)	(0.373)	(0.596)
7. Finished college	(0, 1)	2.79	1.14	1.84
		(0.839)	(0.453)	(0.621)
8. Master's degree	(0, 1)	2.33	1.53	1.79
		(1.04)	(1.06)	(0.628)
9. Doctor's degree	(0, 1)	−0.406	−0.780	0.279
		(0.312)	(0.407)	(0.689)
10. Not reported[b]	(0, 1)	0	0	0
Mother's education				
11. Some grade school	(0, 1)	0.588	0.935	−0.476
		(0.283)	(0.564)	(1.15)
12. Finished grade school	(0, 1)	0.685	0.910	−0.912
		(0.270)	(0.490)	(1.09)
13. Some high school	(0, 1)	0.698	1.29	0.0163
		(0.273)	(0.485)	(1.07)
14. Finished high school	(0, 1)	1.26	1.83	0.624
		(0.290)	(0.490)	(1.08)
15. Some college (or other post high school)	(0, 1)	1.35	1.94	0.640
		(0.443)	(0.550)	(1.08)
16. Finished college	(0, 1)	1.97	1.34	0.756
		(0.881)	(0.662)	(1.13)
17. Master's degree	(0, 1)	0.216	0.453	1.11
		(1.53)	(1.50)	(1.57)
18. Doctor's degree	(0, 1)	0.360	0.627	4.18
		(0.395)	(0.597)	(1.17)
19. Not reported[b]	(0, 1)	0	0	0
Student sex				
20. Male[b]	(0, 1)	0	0	0
21. Female	(0, 1)	0.448	−0.263	−0.0422
		(0.0883)	(0.102)	(1.40)
22. Constant	1	−6.00	−5.10	−8.80
		(0.387)	(0.562)	(1.33)
Number of observations		3.436	2.177	1.493
Maximum log-likelihood		−1,540	−1.150	−661
Log-likelihood for $\hat{\theta}=0$		−2,380	−1,510	−1.030

[a] Number in parentheses are asymptotic standard errors.
[b] Normalization.

were divided according to the same income strata used in the college-choice model. The estimated coefficients are presented in Table 5.

Recall the structure of this model, described by equation (15) and rewritten below:

$$\gamma\hat{U}_{1i} + Y(Z_i) \cdot \lambda \geq \eta_i.$$

The principal variable, variable 1 in Table 5, is the utility of the best college available—that is, $\hat{U}_{1i}$. This is taken to be the highest utility of any college in the imputed-choice set when the utility is calculated using only the behavioral coefficients (i.e., variables 1 through 10 in Table 3). We tried to calculate the highest utility using the college-type dummy variables as well. The likelihood was increased, but not by very much. Consequently, we excluded the dummies since their behavioral significance was doubtful and their contribution here small. Variables 2 through 22 are the Y functions of equation (15).

Variables 2 through 10 in Table 5 are a set of dummy variables representing the education of the student's father. Variables 11 through 19 are a set of dummy variables representing the education of the student's mother. The effect of the father's education seems to be greater than that of the mother's. There seems to be a jump in the probability of the student attending college if the father has completed college (or had some college for the high-income group) and if the mother has completed high school. There is a decrease in the importance of parental education as family income rises.

Variable 22 is a constant, and the differences between the values of its coefficient for the three income strata represent "pure" income effects. Variable 1, the utility of the best college, already accounts for all of the effect of family income on the availability and attractiveness of college alternatives. The "pure" income effect here represents the effect of a change in family income when all other variables, including the utility of the best college, are held constant.

The size of the difference in the constant coefficients depends on the normalization of the other dummy variables: The values in Table 5 are for male students reporting neither father's nor mother's education. If we normalize on male students whose fathers have finished high school and whose mothers have finished grade school, the constants become −4.7, −3.7, and −7.3. The same kind of differences occur in the constant as income rises, regardless of the normalization: The probability of going to college rises as we go from the low-income stratum to the middle-income stratum, and then falls sharply as we go on to the high-income stratum. This result does not contradict the fact that a higher proportion of students from high-income families go to college than those from low-income families. It does imply, though, that this pattern among enrollment rates is largely explained by the existence of more attractive college alternatives for students from high-income families. This relative attractiveness depends, in part, on the lower disutility of tuition for higher-income students. If faced with equally attractive college alternatives (in the subjective sense of equal utility), the high-income student will be less likely to go to college. Furthermore, in the light of our previous theoretical discussion, the result is not really surprising since high-income students probably face a more attractive set of noncollege alternatives.

The college-going model was reestimated with similar North Carolina data. The pattern of results did not differ substantially, except that the parental education effects were weaker, and the mother's education seemed more important than the father's (see Table 6).

V. A Simple Example of the Use of the Results in Forecasting

The purpose of our study has been to investigate and estimate the impacts of student and institutional attributes on college-going behavior and to describe the use of these estimates in predicting the effect of alternative higher education policies. All too often, investigators arrive at a set of estimates of coefficients and leave the use of their results entirely in the hands of others. In most instances, this leaves the estimates either unutilized or misused. To avoid this problem, and to fulfill partially the second goal of our study, this section presents a simplified simulation of the results of the choices facing a student. In our simulation, we forecast the behavior of a student facing a choice among three alternatives: choosing a public university, choosing a public two-year college, and not enrolling. The simulation follows the three stages of the estimation process. The public policy question is where to locate the two-year college given the already existent university.

The student in our simulation is a male with a Scholastic Achievement Test (SAT) score of 500, which places him approximately at the median for high school graduates. His family income is $6,000; his father graduated from high school; and his mother has some high school education. He prefers living on campus to living at home and commuting. He is an Illinois resident. The role of these various student characteristics is described by the appropriate coefficients in Tables 2, 3, and 5. The attributes of the two colleges are as follows:

Variable	Public University	Public Two-year College
Tuition	$800	$200
Room and board	$1,000	—
Breadth of offering	9	4
Average revenues	$1,300	$1,000
Average SAT score	600	500
Coed	Yes	Yes
Dormitory capacity	50%	0%
Distance from student	Fixed at 0 miles	(Variable)

The first stage in the simulation is to estimate the probability that the student will live on campus given the various distances from home to college. Using

TABLE 6

ESTIMATED COEFFICIENTS OF THE COLLEGE-GOING MODEL
(North Carolina data)

Variables	Approximate Range of Values	Coefficients, by Income Stratum: <$7,860	$7,860–11,470	>$11,470
General				
1. Utility Index of best college	(−8 to +8)	0.936	1.03	0.841
		(0.039)[a]	(0.0705)	(0.0659)
Father's education				
2. Some grade school		0.278	0.348	0.191
	(0, 1)	(0.178)	(0.593)	(1.35)
3. Finished grade school		0.282	1.22	0.168
	(0, 1)	(0.123)	(0.539)	(1.76)
4. Some high school		0.876	1.30	0.00563
	(0, 1)	(0.122)	(0.525)	(1.23)
5. Finished high school		0.994	1.57	0.429
	(0, 1)	(0.145)	(0.530)	(1.23)
6. Some college (or other post high school)		0.889	1.88	0.576
	(0, 1)	(0.217)	(0.557)	(1.23)
7. Finished college		1.26	1.69	0.500
	(0, 1)	(0.397)	(0.663)	(1.26)
8. Master's degree		0.246	3.30	0.742
	(0, 1)	(0.728)	(2.56)	(1.28)
9. Doctor's degree		−0.0953	0.354	−0.868
	(0, 1)	(0.158)	(0.599)	(1.34)
10. Not reported[b]	(0, 1)	0	0	0
Mother's education				
11. Some grade school		0.0324	1.21	−1.95
	(0, 1)	(0.180)	(0.808)	(1.50)
12. Finished grade school		0.319	−1.38	−0.772
	(0, 1)	(0.167)	(0.743)	(1.10)
13. Some high school		0.870	−0.770	1.06
	(0, 1)	(0.164)	(0.734)	(1.04)
14. Finished high school		1.52	−0.345	1.29
	(0, 1)	(0.184)	(0.742)	(1.05)
15. Some college (or other		2.05	0.364	1.57
post high school)	(0, 1)	(0.230)	(0.754)	(1.05)
16. Finished college		1.54	0.457	2.04
	(0, 1)	(0.467)	(0.998)	(1.12)
17. Master's degree		NE	0.120	1.68
	(0, 1)		(1.49)	(1.43)
18. Doctor's degree		−0.208	−1.75	0.845
	(0, 1)	(2.45)	(0.860)	(1.20)
19. Not reported[b]	(0, 1)	0	0	0
Student sex				
20. Male[b]	(0, 1)	0	0	0
21. Female		−0.053	−0.0741	−0.394
	(0, 1)	(0.0668)	(0.135)	(0.171)
22. Constant	1	−6.58	−9.35	−8.93
		(0.250)	(0.925)	(1.42)
Number of observations		6,389	1,453	1,096
Maximum log-likelihood		-0.281×10^4	-0.694×10^3	-0.457×10^3
Log-likelihood for $\hat{\theta} = 0$		-0.443×10^4	-0.101×10^4	-0.760×10^3

Note: NE = not estimable.

[a] Numbers in parentheses are asymptotic standard errors.

[b] Normalization.

Table 2 and equation (16), we arrive at the following estimate of the student's residency choice:

Variable	Public University	Public Two-year College
Probability of living on campus	0.108	0[22]

Using these estimates, we assign the student to commuter status in each institution for the remainder of the simulation.

In the second stage of the simulation, we estimate the attractiveness (or utility) of each institution for the particular student. This is done using the equation $\hat{U} = V \cdot \theta$, where θ is the appropriate vector of coefficients from the college-choice equation (Table 3) and V is the vector of institutional attributes. The results of the utility calculation are shown in Table 7.

TABLE 7
UTILITY OF COLLEGE ALTERNATIVES

Distance (miles)	Utility: Public University	Utility: Public Two-year College
0	3.80	4.35
10	NA	4.04
20	NA	3.72
30	NA	3.41
40	NA	3.09
50	NA	2.78
60	NA	2.47
70	NA	2.15

Note: NA = not appropriate.

If we were concerned with the enrollment effects of the two-year college alone (i.e., assuming either that the university did not exist or that the student would not be admitted to it), the third stage of the simulation would involve transformation of the utility estimate into an enrollment probability. This would have the logistic form

$$P\,(\text{enrollment}) = \frac{e^{w}}{1+e^{w}},$$

where $w = \hat{\gamma}\hat{U} + Y\hat{\lambda}$. The results of this calculation are shown in Figure 4. The third stage of our simulation (of the effect of the two-year college location decision on enrollment rates and on a student's behavior pattern when he lives next door to a university for which he is eligible) has two parts. In the first part, the enrollment rate is calculated in the same way as in the single-college model, except that the

[22] For all distances, given zero dormitory capacity.

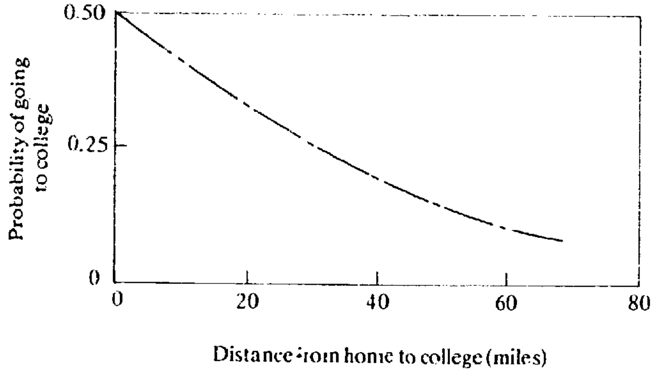

Figure 4 Probability of enrollment versus distance for student facing two-year college

best college (i.e., the one with the highest utility) is entered into the probability of enrollment equation. Consequently, the two-year college stimulates enrollment only when it is at a distance at which its utility exceeds that of the close-by university. In the case of our example, the utility of the two-year college exceeds that of the university when the two-year institution is within 20 miles of the student. The result of the enrollment rate calculation is shown in Figure 5.

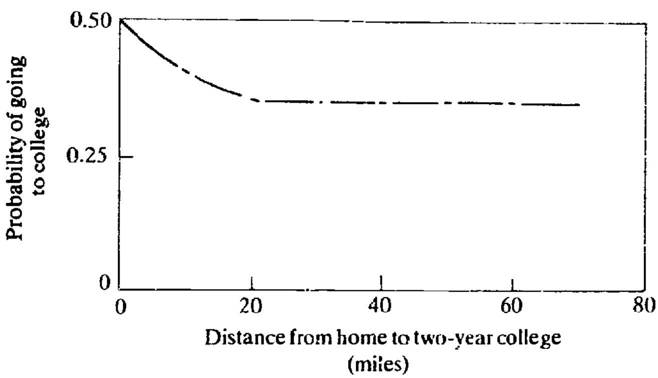

Figure 5 Probability of enrollment versus distance to two-year college for a student living near a public university

Even though the two-year institution stimulates additional enrollment only if it is within 20 miles of the student, it does influence the distribution of enrollment between the two institutions at all distances. In our simulation, we estimate that the likelihood of the student enrolling in a particular institution, if he enrolls, is given by the following

$$P(a) = e^{\hat{U}_a} \div (e^{\hat{U}_a} + e^{\hat{U}_{a'}}),$$

where a and a' are the two institutions. Using the estimated utilities of the two institutions (Table 7), we arrive at the division of the student's enrollment between the two institutions shown in Figure 6.

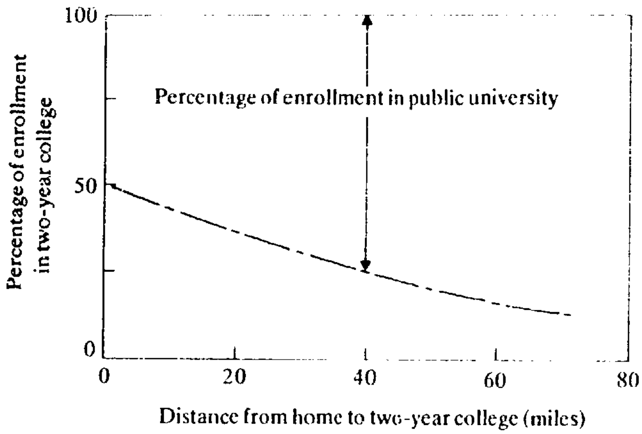

Figure 6 Distribution of enrollment versus distance to two-year college

Another way of looking at the distribution of students is from the point of view of the university, which faces competition for students from the proposed two-year college. The effect of the new institution on the enrollment of the existing university can be estimated using Figures 4 and 5. The estimate is shown in Figure 7.

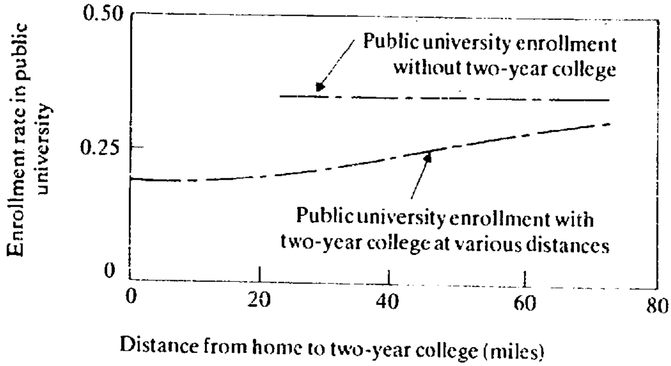

Figure 7 Public university enrollment versus two-year college location

VI. Some Thoughts on Next Steps

In their current state, the utility of our models of college choice and college going in the policymaking process is somewhat restricted. There is a need for both improved specification and improved data in order to increase their utility. An important limitation of our current efforts is our inability to specify variables—especially financial aid offers—which are important (and, at this time, highly controversial) policy instruments. Improving the reliability of our financial aid estimates in order to examine the effect on student behavior will require better data and a better understanding of institutional behavior, the key determinant of financial aid offers. Without improved understanding of this determination process, the reliability and consequently the utility of our models will remain somewhat limited.

Another important constraint on our model is the lack of currency in the data. Many observers have noted that student choice and college-going behaviors have changed dramatically in recent years. Whether these changes are the result of modified desires on the part of students or the result of an altered demographic distribution of high school graduates or of changes in the alternatives students face remains in doubt; but, without more recent data, our estimations remain less than thoroughly convincing in the current policy debate.

There is also a further need to investigate the similarity of student behavior among students from different states. Although our coefficients derived from an initial run with data on North Carolina students are essentially like those derived from data on Illinois, they differ in important respects. These differences limit the acceptability of using one state's coefficients to predict behavior for students in another state. The similarity does, however, lead us to believe that more careful specification may result in patterns of coefficients that are similar enough that one state's model will prove useful in predicting behavior in another state.

Another important limitation of our model is our crude handling of noncollege alternatives. It is clear that the noncollege alternatives of one student will not be the same as those of another, although we implicitly appear to assume that they are. The characteristics of noncollegiate choices are also important aspects of the on-going policy debate, and our model provides essentially no information for these considerations. With the end of the military draft and the new eligibility for federal financial aid awards of students in proprietary and public vocational-technical schools, the availability and desirability of noncollegiate choices has probably changed dramatically. Our model must be refined to include the impact of these changes.

Finally, improvements will result from application of the model to the real, rather than simulated, policy problems facing post secondary decisionmakers. An effort to achieve these improvements is currently under way as we introduce the model into the post secondary education policymaking process in Florida. This effort will involve gathering state data and building simulation routines that are specific to the state's higher education environment.

Annotated List of Data Sources

The following annotations describe the data sources we used for this study.

The College Entrance Examination Board, *Manual of Freshman Class Profiles (1965–67).*

This source describes the admission decisions of 419 colleges in relation to the ability of the applicants, accepted and enrolled freshmen students. These data were the principal input into the admission prediction model. Published biennially, the book contains tables with data relevant to the analysis of the institutional decision process. Although the level of detail and arrangement of tabular material is not uniform for the set of colleges covered, the tables are organized under a small number of categories within which presentation is uniform.

American Council on Education, *Institutional Research File.*

This source provides data on institutional characteristics for 2,319 higher education institutions. The college information contained in this file (median student ability, educational expenditures per student, institutional affluence, tuition, financial aid outlays, institutional control and type, and range of academic fields) is a primary source of data on college attributes.

As a source of consistent and accurate data, the ACE file has a number of disadvantages. First, while ACE data for 4-year colleges and universities deal with academic year 1966–67, data for community colleges are for the previous academic year. Second, some data are presented in a form which renders them only marginally usable for our study. Third, the interpretation of a number of pieces of information is hampered by the vagueness of the questions asked, which results in a lack of consistency of the reported data for different colleges.

Institutional Location Data

As a complement to the ACE Institutional Research File, we have developed a data file containing the latitude and longitude of almost every college and university.

The SCOPE Survey

SCOPE is a comprehensive survey of 1966 high school freshmen and seniors. Approximately 33,000 students in each grade level were surveyed and tested. The students' parents were then surveyed, and the seniors who could be located in a college were subsequently followed-up one year after high school graduation. All students except seniors were resurveyed annually and one year after high school graduation if they were enrolled in a college. The students surveyed came from 305 high schools in four states—California, Illinois, Massachusetts, and North Carolina.

The SCOPE data include aptitude and achievement scores, parental income and education (from student and parent reports), student's career plans, college enrollment, source of funds for college expenses (from student and parent reports), college residence type, etc.

We used the survey of 1966 seniors. This survey has the following response pattern:

1966 high school seniors	33,000
1966 parental responses	11,700
1967 students attending college	17,200
1967 college respondents	10,600

References

Campbell, R. and B. Siegal, "The Demand for Higher Education in the United States, 1919–1964," *American Economic Review*, Vol. 57, No. 3, June 1967, pp. 482–494.

Corrazini, A., D. J. Dugan, and H. Grabowski, "Determinants and Distributional Aspects of Enrollment in the U.S. Higher Education," *Journal of Human Resources*, Vol. 7, No. 1, Winter 1972, pp. 39–50.

Feldman, P. and S. Hoenack, "Private Demand for Higher Education in the United States," *The Economics and Financing of Higher Education in the United States*, The Joint Economic Committee. Washington, D.C., 1969.

Flannagan, J., *et al.*, *The American High School Student*, Final Report of Coop Research Project No. 635, U.S. Office of Education, University of Pittsburgh, 1964.

Galper, H. and R. M. Dunn, Jr., "A Short-Run Demand Function for Higher Education in the United States," *Journal of Political Economy*, Vol. 77, No. 5, September–October 1969, pp. 765–777.

McFadden, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, 1973.

Medsker, L. L. and J. W. Trent, *Beyond High School: A Study of 10,000 High School Graduates*, Center for Research and Development in Higher Education, University of California, Berkeley, 1967.

Miller, L. S., *Demand for Higher Education in the United States*, Working Paper No. 34, Economic Research Bureau, State University of New York, May 1971.

Radner, R. and L. S. Miller, "Demand and Supply in U.S. Higher Education: A Progress Report," *American Economic Review—Papers and Proceedings*, Vol. 60, No. 2, May 1970, pp. 326–334.

U.S. Bureau of the Census, *Factors Related to High School Graduation and College Attendance: 1967*, Series P-20, No. 185, July 11, 1969.