Including Transfer-Out Behavior in Retention Models: Using the NSLC Enrollment Search Data

Presented at the 1999 North East Association of Institutional Research conference, Newport, Rhode Island

Copies of this paper may be found at
http://www.wam.umd.edu/~srporter/NEAIR99.pdf

Stephen R. Porter, Ph.D.
Office of Institutional Studies
2119 Main Administration Building
University of Maryland, College Park
College Park, MD 20742

Telephone: (301) 405-5608
Fax: (301) 314-9443
Email: sporter@accmail.umd.edu

Abstract

Almost all studies of retention inappropriately combine stopouts with transfer-outs due to a lack of data. The National Student Loan Clearinghouse has created a new database that tracks students across institutions. These data in combination with institutional databases now allow researchers to take into account both stopout and transfer-out behavior. Using NSLC data for the University of Maryland, College Park, the paper analyzes one-year retention with dichotomous and multinomial logit under two specifications: the traditional binary retained/not retained dependent variable and a three-outcome dependent variable where students are coded as retained, transferred to another institution, or stopped out. Taking into account transfer-out behavior affects not only the statistical significance of the explanatory variables but also their substantive interpretation.

Introduction

Studies of student retention at the college level are numerous and heterogeneous, taking into account various combinations of academic, financial, institutional and social factors (e.g. Bean, 1980, Manski & Wise, 1983, St. John, 1996, Tinto, 1993). All of these studies, however, have one thing in common: they view the student's decision to reenroll as a binary yes/no decision. This formulation masks the larger set of choices faced by students. After beginning college, students can decide to remain at their current institution, transfer to any number of other postsecondary institutions, or stop out and discontinue their postsecondary education altogether. The binary formulation biases any statistical results, because students who wish to finish their degrees elsewhere are inappropriately combined with students who have decided not to finish their education.

Traditional studies have combined the transfer and stopout choices together due to a lack of information. College databases only record registration and graduation activities. If a student does not appear in the database at a certain point in time, they are assumed to have stopped out or transferred and assigned that category for analysis. Tracking students who do not enroll and determining if and where they transferred is a difficult task for many institutions. Although some public university systems have developed tracking databases, these often exclude private institutions within the state and cannot track students to out-of-state institutions.

The National Student Loan Clearinghouse (NSLC) has developed a transfer student database that should revolutionize the study of post-secondary student behavior. Their Enrollment Search database allows researchers to:

1. Determine which of their students have transferred.
2. Identify the name and FICE number of the transfer institution.

3. Identify when the student first enrolled there.

By combining the NSLC data with college and university databases institutional researchers are now able to study retention in ways previously impossible.

The paper consists of five sections. The first section describes the NSLC data, their collection procedures and coverage. The second discusses traditional retention models and how they can be revised using Enrollment Search data to include the transfer-out option. The third section discusses other possible ways of obtaining transfer data and how to appropriately analyze discrete data with more than two outcomes. The fourth section estimates models of retention using both the traditional two-outcome and a three-outcome variable that includes the transfer-out choice and discusses the results. The last section is a summary and conclusion with a discussion of possible future research using this data.

Enrollment Search data

The NSLC acts as a central reporting agency for colleges and lenders and assists both with various aspects of student loans, such as tracking and confirming the deferment status of borrowers[1]. Member institutions periodically report enrollment information to the NSLC. Because some students may receive loans at one institution and then appear at another institution and not receive any loans, institutions report enrollment information on *all* students, not just those students receiving financial aid. The resulting data is used for their Enrollment Search program.

In the Enrollment Search program participating institutions submit the names, birth dates and dates of last attendance of students who fail to reenroll during a given semester. The NSLC takes this information and searches their database for a match among other participating institutions.

[1] See http://www.nslc.org/ for more information.

If a match is found, information about when and where the student transferred is provided to the home institution. Data provided by the NSLC for each student found include the name and FICE code of their new institution, school type (two-year versus four-year), and transfer term begin date. As of May 1999, the NSLC had enough colleges participating (or planning to participate) that approximately 80% of the enrolled students nationwide were covered (National Student Loan Clearinghouse, 1999).

The current status of the Enrollment Search procedure is somewhat uncertain. In its previous iteration as “Transfer Track”, institutional data requests included Social Security numbers that were then used to match with students in NSLC databases. This procedure now appears in violation of FERPA regulations and the current Enrollment Search procedure will not allow the submission of Social Security numbers for most data requests of interest to institutional researchers (Ward, 1999). NSLC believes it will achieve a very high match rate based on name, birth date and dates of enrollment, so the data will still be a very valuable resource for studies of student persistence. As of this writing the NSLC had not conducted any studies comparing match rates under the two systems. The data used in this paper were obtained last year through the former Transfer Track program and students were matched based on their Social Security numbers.

Expanding choice sets in retention models

Numerous statistical models of persistence have been estimated over the past several decades, focusing on such varied factors as student integration and goal commitment (Allen & Nora, 1995, Cabrera, Nora, & Castaneda, 1993, Okun, Benin, & Brandt-Williams, 1996, Pascarella & Terenzini, 1980, Tinto, 1993), financial aid (Nora, 1990, St. John, 1994;St. John, 1996, St. John et al., 1990), human capital (Manski & Wise, 1983), and organizational attributes (Bean, 1980;Bean, 1983, Berger & Braxton, 1998, Nora et al., 1996). The standard approach for constructing dependent variables in these studies tracks student registration behavior from one year to the next and codes students as re-enrollees or stopouts based on registration activity. Alternatively, some researchers have used a dependent variable based on student survey responses (Berger & Braxton, 1998, Braxton et al., 1995). For example, Berger and Braxton (1998) used a five-point Likert scale ranging from “likely to reenroll” during the next fall semester to “extremely unlikely” in a survey administered to new freshmen. In both cases retention outcomes are viewed as two possibilities along one dimension: stay versus go.

Transferring to another institution is a second dimension of retention that researchers have for the most part ignored. Many students whom we treat as stopouts are actually transfer-outs. By leaving their home institutions, transfer-out students make a much different decision compared with stopouts. Transfer-outs still wish to continue their education, but for some reason they decide that finishing at another institution would help them better achieve their educational goals than remaining where they matriculated. Conversely, true stopouts decide their educational goals are best met by discontinuing their education altogether. If this is indeed the case, transfer-outs and stopouts must be treated separately in any statistical analysis. If not, combining them into one category as has traditionally been done should not pose a problem.

Research on transfer students tells us how similar these two groups of students are. Unfortunately this research has focused almost exclusively on students transferring from two-year to four-year institutions rather than students transferring out from four-year institutions. Although the student populations are quite different (Dougherty, 1992), they are analogous. Community college students who eventually earn a bachelor’s degree must transfer to and complete their education at a four-year institution; similarly, at the four-year level transfer-outs leave and complete their education at another institution. Community college students who do not earn a bachelor’s degree have for some reason declined to further pursue their education; stopouts at the four-year level also do not pursue their education and fail to finish their degree.

Community college students who either express an intent to transfer or who actually transfer and complete a bachelor's degree are quite different from those who do not. They come from higher socioeconomic backgrounds and do better in high school and community college (Kinnick & Kempner, 1988, Kraemer, 1995, Nora & Rendon, 1990, Pascarella et al., 1986). In addition, a study of multiple transfers (many of whom had transferred between four-year institutions) shows that they also come from higher socioeconomic backgrounds and have high academic ability (Kearney et al., 1995). Clearly the explanatory variables used in retention models will have different impacts on transfer-outs and stopouts. Therefore researchers *must* take into account the different choices faced by students when studying persistence.

Data and methodological concerns

Obtaining good data

Knowing that the choice sets of students should be expanded is of little use if the data measuring such choices is unavailable. Registration data and beginning student surveys can only provide data on whether or not the student is retained (or is planning to return) during a given semester. Researchers have tried to circumvent this problem in three ways.

The first solution uses state higher education agencies to track student movement between public two-year and four-year institutions (DesJardins & Pontiff, 1999, Ronco, 1996), but students who transfer to in-state private or out-of-state institutions are treated as stopouts (although some states track students in all institutions regardless of their public/private status).

The second solution uses an "intent to transfer" question on exit surveys of graduating students (Kraemer, 1995), but this works only at the community college level where such surveys can be made part of the graduation process. Students who do not graduate at the community college level and students who leave at the four-year level can also be surveyed. Given differences in socioeconomic background of transfer-outs and stopouts, and that the probability of survey response is often correlated with socioeconomic background, unless a high response rate is achieved such data would be of questionable use.

The third solution involves examining transcript requests and calling all institutions where a student has submitted a transcript to verify enrollment (Kraemer, 1995). Of the three this approach offers the cleanest data, but the costs can be high for larger institutions and may not be practical for many institutional researchers.

The NSLC Enrollment Search data provides a fourth solution. Member institutions can submit lists of student stopouts and for a fee obtain information about when and where they transferred. As with all data there will be some error: due to lack of complete coverage some transfers will not appear and will be coded by the researcher as stopouts, and some stopouts may be mistakenly identified as transfers. But compared to the traditional approach where only institutional data is used and all transfers are erroneously treated as stopouts, the inclusion of Enrollment Search data results in much cleaner data. Depending on the type of institution the Enrollment Search data will also be much cheaper and easier to obtain.

Statistical approach

A more complicated choice set requires a more complex statistical approach than is typically used. Discrete choice models are a class of maximum likelihood techniques that are commonly used in the social sciences to model choice behavior where the outcome, or dependent variable, is discrete rather than continuous. The familiar *logistic regression* (or *logit*), for example, is used when the dependent variable has only two outcomes, such as the traditional measure of student persistence. There are other types of discrete choice models that allow analysis of more complex educational behavior. Because many textbooks and researchers use different names for the same methodology, a brief review is in order[2].

Ordered logit models are used when the dependent variable has more than two discrete outcomes, and these outcomes can be ranked in some fashion (i.e. the data is ordinal). Bond ratings are the common example in economics research, while in the field of education opinion surveys would be another. In this approach we assume that one outcome can be ranked above another, but we know nothing about the distance between outcomes. For example, in an opinion survey there may be three responses such as "very satisfied", "somewhat satisfied", and "not satisfied at all". We know the first response can be ranked above the second in terms of satisfaction, and the second response ranked above the third, but we cannot be sure that the distance between the first and second responses is equal to the distance between the second and third. Multiple regression makes this assumption of common distance, rendering it theoretically unsuitable for such data[3].

There are two additional techniques that allow analysis of dependent variables with more than two discrete outcomes, but these are used when the outcomes cannot be ranked in any meaningful way (i.e. the data is nominal). The technique used depends on the data being analyzed. In the field of economics information about choices is very common. For example, analyses of commuter choice behavior will use datasets in which information varies over the commuting choices of bus, car or train. This information may take the form of cost of the commuting choice per mile, or the time of commute for each choice. These models are known as *conditional logit* models and have often been used to model educational choice after high school (e.g. Fuller et al., 1982).

The other technique for nominal data is known as *multinomial logit*, and is used when only

[2] Much of the following discussion is taken from Chapter 19 of Greene, 1997). Although his textbook is very technical the chapter on discrete choice models has a very clear narrative and is a must-read for anyone working with these techniques.
[3] Of course, in practice there may not be much difference between multiple regression and ordered logit for many applications.

individual-specific (versus choice-specific) data is analyzed. Using the commuter example, we may only have access to data such as income, education and occupation of the individual commuter (as well as their commute choice). Data from public opinion surveys is often analyzed using multinomial logit. Examples of this technique in the field of education include work by Keil and Partell (1999), Ordovensky (1995) and Weiler (1987, 1989).

The main drawback to multinomial logit is a restrictive assumption known as the independence of irrelevant alternatives (IIA). These models assume that if one of several alternatives was suddenly removed from the choice set, the probability of an individual choosing the remaining alternatives increases proportionally. For example, if transferring to another institution suddenly were no longer an option, the probability of transferring would be distributed equally to the options of reenrolling and stopping out. This is somewhat unrealistic, because we would assume that students who could no longer transfer would not be evenly distributed between reenrolling and stopping out; instead, most would choose to reenroll as they would wish to continue their postsecondary education.

One solution to this problem is a procedure known as *nested multinomial logit.* It is similar to regular multinomial logit except for how the choice process is viewed: simple multinomial logit treats the choice made as one among a group, while nested multinomial logit breaks the choices into branching sequential subgroups (such as enroll or stop out; if enroll, remain at home institution or transfer, etc.) (see Ordovensky, 1995, Weiler, 1987 Weiler, 1996). SuchWeiler, 1996 nesting avoids the independence of irrelevant alternatives (IIA) problem. Unfortunately this procedure demands data on attributes of the choices, such as tuition or distance, which are not available given the formulation of the data used in this study.

However, use of the IIA assumption may not be problematic for these types of studies. Weiler (1987) calculated models of educational choice using both regular and nested multinomial logit models. The substantive results for the two models were generally similar, although occasionally the size of the coefficients differed quite a bit. His study, while only suggestive, indicates that simple multinomial logit should yield fairly robust results.

One confusing aspect of multinomial models for the uninitiated is the generation of multiple sets of coefficients. For example, in this analysis there will be two sets of coefficients rather than one. This results from the nature of the dependent variable. In the binary case the coefficients are usually estimated in the form of measuring the impact of an independent variable on the probability of the yes outcome versus the no outcome. The multinomial case is exactly the same: the coefficients measure the impact of an independent variable on the probability of *one* outcome versus a *base* outcome. Since there are three outcomes and one outcome is treated as the base (or “excluded”) outcome, the result is two sets of coefficients. In the context of this study the natural base category is reenrolling after one year. Note that changes in probability remain the same no matter which outcome is excluded; however, the coefficients themselves will change depending on the excluded category.[4]

Analysis

The paper analyzes one-year retention for the Fall 1996 cohort of new first-time full-time degree-seeking freshmen at the University of Maryland, College Park. In addition to the standard two-outcome enroll/not enroll dependent variable, this study uses a three-outcome variable derived from institutional databases and the Enrollment Search data. Based on their Fall 1997 registration behavior students are coded as reenrolled at UMCP, transferred to another institution, or stopped

[4] The probabilities do not change because different formulas are used for different outcomes depending on which outcome is excluded. See Greene (1997) p. 875.

out[5]. This choice set captures the some of the complexity involved in student decision-making while remaining simple enough for a rigorous statistical analysis.

There is an extensive literature on the decision after high school to begin work on a baccalaureate degree (e.g. Fuller et al., 1982, Ordovensky, 1995, Weiler, 1987 Weiler, 1987). This decision is similar to the decision students face after one year in college and the same theoretical and statistical tools can be used. The theoretical model is a human capital approach, where students are assumed to view their educational choices as investment decisions (Becker, 1975). Simply put, students compare the costs and benefits of obtaining an education at a particular institution versus other institutions and immediately participating in the labor market and make the choice that will maximize their utility, generally conceived as their lifetime earnings[6]. Students' choices will differ because individual attributes of the students will affect both the return and the costs of their educational investment.

Explanatory variables are divided into four groups: demographics, human capital, uncertainty, and costs (see Table 2; descriptive statistics are given in Table 3). Demographic variables are simply used as controls and include the student's age, gender, minority group status and international student status.

The human capital variables measure the amount of "capital" the students have to invest by obtaining a baccalaureate degree. Students with greater capital will earn higher returns from attending college. In the case of one-year retention these students should prefer continuing their education to stopping out, so students with greater academic ability should be more likely to be retained. Five variables capture various aspects academic ability: Scholastic Aptitude Test scores, high school grade point average, the number of college credits at matriculation, living on campus

[5] Note that these are *presumed* stopouts, because we have no knowledge of their educational behavior in Fall 1997.

during the first semester and participation in an honors program.

The inclusion of living on campus and honors participation may appear controversial because these variables are often treated as “safety net” programs that directly affect student behavior. Implicit in this formulation, however, is the assumption that students who participate are no different than those who do not, so any differences in behavior between the groups are due to the effect of participation. This is clearly not the case. Admission to an honors program is dependent on academic aptitude, and studies have shown that students who choose to live on campus have higher socioeconomic status and higher high school grade point averages (Levin & Clowes, 1982). These variables are more measures of student background than program impacts and are treated as such.

As with any decision, students are somewhat uncertain as to the exact benefits a post-secondary education will bestow. Students with greater certainty about the benefits should be more likely to be retained. While direct measures of uncertainty are not available, the number of days between the date of application and the first day of class in Fall 1996 can be used as a proxy. Students who are more certain that they wish to pursue a bachelor’s degree and that the University of Maryland offers the best return on their investment compared to other alternatives should tend to apply earlier than those who are not.

Finally, the benefits accrued from higher education must be greater than the costs, so students facing higher costs should be more likely to pursue alternatives (either working or attending a less costly institution) and less likely to be retained. Four variables measure the costs faced by students. Indirect costs such as lack of family support are measured by whether the student was a first generation college student. Other indirect costs such as being far away from family and friends are proxied by the student’s residency status, in-state versus out-of-state. The direct costs of

---

[6] National surveys indicate that students indeed “view higher education less as an opportunity and more as a means to increase their incomes” Bronner, 1998. In the 1998 HERI survey of first-time, full-time freshmen, 74.9% of respondents

attending the university are measured by the amount of unmet need (the amount of money needed by the student after their financial aid package has been taken into account), and the total amount of debt taken on by the student. Because not all students apply for financial aid, an indicator variable is included to measure possible differences between the two groups.

The purpose of this analysis is simple: does the expansion of students' persistence choice set add to our understanding of persistence behavior? Taking into account the transfer-out option requires more data and more complex statistical tools. If our understanding of retention remains the same then nothing is gained. The remainder of the paper attempts to answer this question.

Which model is "better"?

Table 4 presents the results for the two retention models. The first column lists the coefficients and standard errors for the traditional binary retention model where students are classified as retained or not retained as of the Fall 1997 semester. Note that for comparison purposes the values of the dependent variable have been reversed, so the model is estimating the probability of a student *not* being retained instead of the usual being retained. The next two columns list the results for the multinomial logit model of retention. The excluded or base outcome is retained in Fall 1997, so results are given for two outcomes: stopping out and transferring. With these formulations the coefficients are comparable across the models.

We need some sort of criteria to decide between the two approaches to modeling retention. At least two criteria are relevant: predictive ability and explanatory power. Predictive ability is the ability of the model to correctly predict the outcomes of the dependent variable. Explanatory power, on the other hand, has a different connotation in the context of this paper. Explanatory power refers to what the model tells us about student behavior (*not* "what percentage of the variance is

---

cited "to be well off financially" as their educational goal.

explained.”). Are students who live on campus during their first semester more likely to return to the university after a year? Models that can answer these types of questions can be said to have good explanatory power. Obviously explanatory power, unlike predictive ability, cannot be measured directly and is more of a judgement call.

The distinction between the two criteria is important because models can have high predictive power and little explanatory power, and vice versa. A simple example makes this clear. Suppose two analysts estimate dichotomous logit models on a dataset where the overall retention rate is 80%. The first analyst uses a typical group of variables such as demographics, SAT scores, etc., while the second uses only a constant.

Next, an evaluation committee examines the models to determine which one should be used for policy-making purposes. They discover that the standard retention model correctly predicts student retention outcomes only 45% percent of the time, while the constant model predicts correct outcomes 80% of the time (this follows from the construction of the model, because all students are predicted to be retained and 80% actually are retained). The committee rejects the first model and decides to use the second model for their decision-making because of its superior predictive ability. They ask the second analyst, “What does your model tell us about student behavior?” The answer, of course, is nothing, because the model consists only of a constant. The first model, although a poor predictor of retention, nonetheless can offer interesting information about the impact of various variables on student behavior. This example illustrates the difficulty in relying on predictive power for these types of models, because one can easily develop highly predictive models with little explanatory power.

Predictive ability

From the likelihood ratio indices at the bottom of Table 4 we can conclude that the multinomial model appears to fit the data better than the dichotomous model.[7] However, if some type of intervention system for at-risk students is under consideration, the real measure of predictive ability is the proportion of outcomes correctly predicted. An institution does not want to waste intervention resources on students who are likely to stay, and they also do not want to miss applying the intervention to those at-risk students who are likely to stop out. Here the multinomial model performs poorly, because the sample used is what Greene (1997, p. 892) terms “unbalanced”. An unbalanced sample has cases that are not evenly distributed across outcomes. This poses a problem because the base probability for an outcome for every individual will be the relative frequency of that outcome. If the relative frequency is very high or low, then only an extraordinary number of regressors could cause the predicted probability of this outcome to shift above or below the predicted probabilities of the other outcomes.

Because of the unbalanced sample, predicting outcomes in the multinomial model is difficult. Like the dichotomous case, a predicted probability for each individual student and each outcome can be derived from the model coefficients. We can use two different decision rules for predicting outcomes based on these probabilities. First, the outcome with the highest predicted probability can be declared the predicted outcome. Unfortunately with this sample every student is predicted to be enrolled all three semesters, because the predicted probability for this outcome is always in the 70%-90% range, much larger than all the other outcomes. Second, we can compare the predicted probability of each outcome with the actual relative frequency for each outcome. For example, if the predicted probability of stopping out for a student is 8%, this student is assigned this outcome because 8% is greater than the actual relative frequency (or sample mean) of 7.51%. Unfortunately for many students in the sample *two* outcomes are predicted using this decision rule. That is, one outcome has a reduced probability, and since the sum of the probabilities for the three outcomes must sum to 1, this probability is often shifted to two other outcomes rather than just one. The result is ambiguous predictions for many individuals in the sample. Unfortunately the multinomial approach does not seem very useful for actually predicting student outcomes; however, in a more balanced sample the multinomial approach might prove superior to dichotomous logit.

Explanatory power

What the model tells us about student behavior is the second criteria by which to judge the two approaches. Here the differences between the two models are quite interesting. In the dichotomous model four variables have a statistically significant impact on the probability of not enrolling. Students with higher grade point averages, who live on campus and who applied early are more likely to reenroll after one year, while students with unmet need are less likely to reenroll.

When the choice of not reenrolling is broken down into not reenrolling by stopping out and not reenrolling by transferring, the results are quite different. As in the dichotomous case, two variables still have a significant impact on both stopping out and transferring: application time and unmet need. Students who applied late and students with large unmet need are both more likely to either stopout or transfer. High school grade point average, however, only affects stopping out. In addition, three variables insignificant in the dichotomous model are now significant. First generation college students are less likely to stopout, while in-state residents and participants in the Honors program are less likely to transfer.

---

[7] The likelihood ratio index is calculated as 1 – (log likelihood of the full model / log likelihood of a model estimated with only a constant) and is bounded from zero to one (Greene 1997, p. 891). It can be thought of as representing the

The substantive meaning of these results can be seen in Table 5, which presents the change in probability of an outcome occurring given a change in an independent variable[8]. Changes in probability were calculated from the model coefficients as follows. The predicted probability of reenrolling was calculated using the sample means for all independent variables except the variable for which the change is calculated. That variable is constrained to the value indicated. The process was repeated using the second value of the independent variable and the difference between the two probabilities was taken. For example, the impact of housing on retention was estimated by calculating the predicted probability with the on campus variable set to zero rather than the sample mean; this was repeated with on campus set to one and the difference taken.

The probabilities of enrolling for the two models are listed in the first two columns and are similar, as expected. The one major difference is that the multinomial changes are all slightly smaller than the dichotomous logit changes.

Because of a fundamental axiom of probability theory, when the probability of reenrolling increases by a certain amount, the probability of not reenrolling must decrease by the same amount. This can be seen in the third and fourth columns of Table 5, which list the changes in the probability of stopping out or transferring. Note that the differences (which are bolded) in these two columns sum to the negative probability of reenrolling in the second column. Here the advantage of using of the Enrollment Search data combined with the multinomial logit model can be seen. The impact of changes in the explanatory variables on the overall probability of not reenrolling can be "broken out" into two parts: the effect on the decision to stopout and the effect on the decision to transfer. In doing so we can now distinguish between factors that affect the decision to *discontinue* post-secondary education and the decision to *continue* by attending another institution.

increase in the log likelihood due to the addition of explanatory variables.
[8] These changes are sometimes referred to a "delta-p's" (Petersen, 1985).

Changes in high school grade point average illustrate this point. When grade point average increases from 3.0 to 4.0 the probability of reenrolling increases about seven percentage points; as theorized, students with greater academic ability are more likely to be retained. The third and fourth columns of Table 5 show that the impact is not the same on the decision to stopout and the decision to transfer. The impact of high school grade point average is much larger for the stopout alternative than the transfer alternative. Living on campus is similar, while the impact of honors program participation is more evenly split between the two alternatives (although not significant for the stopout alternative). The results indicate that academic ability chiefly affects a student's decision to continue with their educational investment, and has little to do with their decision to transfer.

The effect of uncertainty as measured by application time and direct costs as measured by unmet need appear fairly similar for both stopping out and transferring. Application time captures both decision-making aspects faced by students. Students who know they want a college degree will tend to apply earlier, and students who believe that UMCP will provide the best education for them compared to alternative institutions will both tend to apply early. Similarly, as the direct cost of education rises some students will react by deciding that investing in a college degree is not worth the cost. Others will decide it is worth the cost, but their site of investment is too costly in comparison with alternative post-secondary institutions.

The impacts of first generation college student and residency are quite different when taking into account the transfer-out option. In the dichotomous case neither variable is significantly related to persistence, but in the multinomial case first generation status significantly affects the probability of stopping out and residency is significantly related to the probability of transferring.

The effect of being a first generation college student on stopping out is counter-intuitive. First generation college students are more likely to be retained, not less likely as most people would expect, and this is related to the decision to stopout. These results are confirmed by the raw data.

The one-year retention rate for these students is 94% compared with 87.4% for the entire cohort. There are two possible explanations for this result. The first involves the application process. It is possible that the applications of students who identify themselves as first generation college students are carefully evaluated to make sure that these students possess the ability to succeed. If such filtering takes place then the variable would tend to be a proxy for those factors listed on their application or in their essay that are associated with successful students but that are not recorded in institutional databases. The second is that these students are flagged as at-risk students and receive more advising than the average student.

Students from out of state are three percentage points more likely to transfer than students who are Maryland residents. From a human capital perspective this result makes sense. Students from out of state are farther away from home and face greater psychological and monetary costs associated with distance, such as separation from family and travel expenses. In addition, out of state students generally have one or more lower priced educational alternatives in their home state.

The estimated results in general agree with the predictions of a human capital model of student persistence behavior. Uncertainty and direct costs affect both the decision to continue and the decision to transfer. Academic ability affects whether a student continues to pursue their degree, but not whether they transfer. Residency status affects the decision to transfer to another institution only.

Conclusion

Students in higher education face many decisions while pursuing their degree. Two of the most fundamental are whether to finish, and whether to finish at the institution where they matriculated. Only by disentangling these decisions can institutional researchers hope to gain a greater understanding of persistence behavior. The results presented here indicate that NSLC's

Enrollment Search data in combination with internal databases are a practical alternative to the traditional binary outcome approach. Taking into account transfer-out behavior affects not only the statistical significance of the explanatory variables but also their substantive impact.

Many researchers build extremely complex models of retention that completely overlook transfer behavior. Given the difference in results when using the three-outcome persistence variable, these researchers must begin to consider transfer-out behavior when estimating their models. Failure to do so will result in biased estimates and the wrong conclusions about what affects student behavior. Given that over a quarter of students who students who begin their post-secondary education at a four-year institution transfer to another (McCormick & Carroll, 1997), transfer-out behavior cannot be ignored.

Future research in this area should focus on expanding student choice sets even further. Besides facing decisions about continuing their education and staying at their home institution, students must make other decisions. Should I get a four-year degree or settle for an associate's degree? Should I attend an institution in my home state or transfer to an out-of-state institution? Such decisions can easily be analyzed using the Enrollment Search data and a multinomial logit model.

References

Allen, D., & Nora, A. (1995). An empirical examination of the construct validity of goal commitment in the persistence process. Research in Higher Education, 36(5), 509-533.

Bean, J. P. (1980). Dropouts and turnover: the synthesis and test of a causal model of student attrition. Research in Higher Education, 12, 155-187.

Bean, J. P. (1983). The application of a model of turnover in work organizations to the student attrition process. Review of Higher Education, 60, 155-182.

Becker, G. (1975). Human Capital. New York: National Bureau of Economic Research.

Berger, J. B., & Braxton, J. M. (1998). Revising Tinto's interactionalist theory of student departure through theory elaboration: examining the role of organizational attributes in the persistence process. Research in Higher Education, 39(2), 103-119.

Braxton, J. M., Vesper, N., & Hossler, D. (1995). Expectations for college and student persistence. Research in Higher Education, 36(5), 595-612.

Bronner, E. (1998 January). College students value money over mind, survey finds. The New York Times.

Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College persistence: structural equations modeling test of an integrated model of student retention. Journal of Higher Education, 64(2), 123-139.

DesJardins, S. L., & Pontiff, H. (1999). Tracking institutional leavers: an application. Association of Institutional Research Professional File, 71, 1-14.

Dougherty, K. J. (1992). Community colleges and baccalaureate attainment. Journal of Higher Education, 63(2), 188-214.

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

Greene, W. H. (1997). Econometric Analysis. 3rd ed.,). Upper Saddle River, NJ: Prentice Hall.

Kearney, G. W., Townsend, B. K., & Kearney, T. J. (1995). Multiple-transfer students in a public urban university: background characteristics and interinstitutional movements. Research in Higher Education, 36(3), 323-344.

Kinnick, M. K., & Kempner, K. (1988). Beyond "front door" access: attaining the bachelor's degree. Research in Higher Education, 29(4), 299-318.

Kraemer, B. A. (1995). Factors affecting hispanic student transfer behavior. Research in Higher Education, 36(3), 303-322.

Levin, B. H., & Clowes, D. A. C. (1982). The effect of residence hall living at college on attainment of the baccalaureate degree. Journal of College Student Personnel, 99-104.

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

McCormick, A. C., & Carroll, C. D. (1997). (Report No. NCES 97-266). Washington, DC: National Center for Education Statistics.

National Student Loan Clearinghouse. (URL http://www.nslc.org/doc/part_state.htm [1999].

Nora, A. (1990). Campus-based aid programs as determinants of retention among Hispanic community college students. Journal of Higher Education, 61(3), 312-331.

Nora, A., Cabrera, A., Hagedorn, L. S., & Pascarella, E. (1996). Differential impacts of academic and social experiences on college-related behavioral outcomes across different ethnic and gender groups at four-year institutions. Research in Higher Education, 37(4), 427-451.

Nora, A., & Rendon, L. I. (1990). Determinants of predisposition to transfer among community college students: a structural model. Research in Higher Education, 31(3), 235-255.

Okun, M. A., Benin, M., & Brandt-Williams, A. (1996). Staying in college: moderators of the relation between intention and institutional departure. Journal of Higher Education, 67(5), 577-596.

Ordovensky, J. F. (1995). Effects of institutional attributes on enrollment choice: implications for postsecondary education. Economics of Education Review, 14(4), 335-350.

Pascarella, E. T., Smart, J. C., & Ethington, C. A. (1986). Long-term persistence of two-year college students. Research in Higher Education, 24(1), 47-71.

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. Journal of Higher Education, 51(1), 60-75.

Petersen, T. (1985). A comment on presenting results from logit and probit models. American Sociological Review, 50(1), 130-131.

Ronco, S. L. (1996). How enrollment ends: analyzing the correlates of student graduation, transfer and dropout with a competing risks model. Association of Institutional Research

Professional File, 61, 1-16.

St. John, E. P. (1994). The influence of student aid on within-year persistence by traditional college-age students in four-year colleges. Research in Higher Education, 35(4), 455-480.

St. John, E. P. (1996). The nexus between college choice and persistence. Research in Higher Education, 37(2), 175-220.

St. John, E. P., Kirshstein, R. J., & Noell, J. (1990). The impact of student financial aid on persistence: a sequential analysis. Review of Higher Education, 14, 383-486.

Tinto, V. (1993). Leaving college: rethinking the causes and cures of student attrition. 2nd ed.,). Chicago: University of Chicago Press.

Ward, J. (1999 November). Personal communication.

Weiler, W. C. (1987). An application of the nested multinomial logit model to enrollment choice behavior. Research in Higher Education, 27(3), 273-282.

Weiler, W. C. (1996). Factors influencing the matriculation choices of high ability students. Economics of Education Review, 15(1), 23-36.

Table 1. One-Year Persistence of Fall 1996 Freshmen Cohort

Student group	Fall 1997 outcome	%	N
Entire cohort	Enrolled	87.4	3,105
	Not enrolled:		
	Unknown outcome (stopouts)	7.5	267
	Transferred to:		
	Maryland 4-year	0.5	17
	Maryland 2-year	1.4	51
	Out of state 4-year	2.0	70
	Out of state 2-year	1.2	43
		100.0	3,553
Only not enrolled	Unknown outcome (stopouts)	59.6	267
	Transferred to:		
	Maryland 4-year	3.8	17
	Maryland 2-year	11.4	51
	Out of state 4-year	15.6	70
	Out of state 2-year	9.6	43
		100.0	448

Source: NSLC and University of Maryland, College Park databases.

Table 2. Variable Names and Descriptions

Variable type	Variable name	Description
Demographics	Age	Age at time of matriculation (in years)
	Female	Coded 1 if female, 0 otherwise.
	Nonwhite	Coded 1 if the student was a minority or international student, 0 otherwise.
	International	Coded 1 if the student was not a U.S. citizen or permanent resident, 0 otherwise.
Human capital	Combined SAT	Combination of the highest math and verbal Scholastic Aptitude Test scores submitted by the student.
	HS GPA	High-school grade point average.
	Credits	Number of credits brought by the student at matriculation.
	On campus	Measures whether the student resided on campus their first semester, coded 1 if so, 0 otherwise.
	Honors	Coded 1 if student participated in the university Honors program, 0 otherwise
Uncertainty	Application time	Number of days between the first day of classes and the date of the student's application.
Costs	First generation	Taken from the student's application, coded 1 if student indicated s/he was first in family to attend college, 0 otherwise.
	MD residency	Residency based on tuition status, coded 1 if Maryland state resident, 0 otherwise.
	Unmet need	Amount of money needed by the student to cover costs of attending the university during FY 1997. Positive amounts indicate need, negative amounts indicate no need. Students who did not apply for financial aid have missing data for this variable; they are assumed to have zero unmet need and are coded 0.
	Total debt	Total amount of debt accrued by the student during FY 1997
	Aid flag	indicator variable coded one if student did not apply for financial aid, 0 otherwise.

Table 3. Independent Variables – Descriptive Statistics

Variable	Mean	Standard deviation	Minimum	Maximum
Age	18.178	0.954	16	46
Female	0.486	0.500	0	1
Nonwhite	0.350	0.477	0	1
International	0.017	0.131	0	1
SAT combined	119.280	14.425	57	160
HS GPA	3.450	0.495	1.84	5.05
Credits	0.224	1.009	0	12
On campus	0.810	0.392	0	1
Honors	0.358	0.479	0	1
Application time	262.035	45.809	1	599
First generation	0.023	0.151	0	1
MD residency	0.641	0.480	0	1
Unmet need	-1754.999	8705.377	-80746	16612
Total debt	2261.384	3361.137	0	18573
Aid flag	0.214	0.410	0	1

Table 4. Dichotomous and Multinomial Logistic Regression Estimates

	Dichotomous	Multinomial	
	P(not enrolling)	P(stopping out)	P(transferring)
Age	-0.0401	-0.0174	-0.1503
	(0.0485)	(0.0507)	(0.1280)
Female	0.0886	0.0704	0.0879
	(0.1095)	(0.1385)	(0.1636)
Nonwhite	-0.0972	-0.0945	-0.0883
	(0.1269)	(0.1582)	(0.1929)
Foreign	-0.4698	0.1240	-1.7417
	(0.4268)	(0.4699)	(1.0488)
SAT combined	0.0029	0.0105	-0.0094
	(0.0048)	(0.0060)	(0.0074)
HS GPA	-0.7970***	-1.1580***	-0.2567
	(0.1353)	(0.1709)	(0.2016)
Credits	-0.0218	0.0071	-0.0989
	(0.0544)	(0.0626)	(0.1011)
On campus	-0.4446***	-0.6955***	0.0291
	(0.1397)	(0.1648)	(0.2394)
Honors	-0.3912*	-0.2217	-0.6171*
	(0.1627)	(0.2077)	(0.2461)
Application time	-0.0049***	-0.0053***	-0.0044**
	(0.0011)	(0.0013)	(0.0017)
First generation	-0.8342	-2.0053*	0.0345
	(0.4742)	(1.0157)	(0.5363)
MD residency	-0.1949	0.2207	-0.7322***
	(0.1172)	(0.1541)	(0.1725)
Unmet need	0.000032***	0.000032**	0.000031**
	(0.000008)	(0.000011)	(0.000012)
Total debt	0.000011	0.000029	-0.000013
	(0.000017)	(0.000021)	(0.000024)
Aid flag	0.0991	0.0974	0.0955
	(0.1374)	(0.1747)	(0.2030)
Intercept	2.9523*	2.2336	3.5953
	(1.2195)	(1.3791)	(2.6828)
N	3,553	3,553	
Log likelihood	-1253.73	-1525.46	
Model chi-square	184.91***	245.90***	
Likelihood ratio index	0.069	0.075	

Note: standard errors in parentheses; * p<.05, ** p<.01, *** p<.001.

Table 5. Change in Probability of Retention Outcomes for Significant Independent Variables

	Dichotomous	Multinomial		
	P(enrolling)	P(enrolling)	P(stopping out)	P(transferring)
High school GPA = 3.0	85.5%	86.2%	9.4%	4.5%
High school GPA = 4.0	92.9%	93.1%	3.2%	3.7%
Difference	**7.4%**	**6.9%**	**-6.2%**	**-0.7%**
Resided off campus	85.5%	86.3%	9.8%	3.9%
Resided on campus	90.2%	90.7%	5.1%	4.2%
Difference	**4.7%**	**4.3%**	**-4.7%**	**0.3%**
Not enrolled in Honors program	88.0%	88.7%	6.2%	5.1%
Enrolled in Honors program	91.6%	92.0%	5.2%	2.9%
Difference	**3.6%**	**3.3%**	**-1.0%**	**-2.2%**
Applied six months	84.9%	85.7%	8.6%	5.7%
Applied twelve months	93.2%	93.6%	3.6%	2.8%
Difference	**8.3%**	**7.9%**	**-5.0%**	**-2.9%**
Unmet need = $20,000	80.9%	82.0%	10.5%	7.4%
Unmet need = $0	88.9%	89.5%	6.1%	4.4%
Difference	**8.0%**	**7.5%**	**-4.4%**	**-3.1%**
Not first generation college	-	89.8%	6.1%	4.1%
First generation college	-	94.6%	0.9%	4.5%
Difference	-	**4.8%**	**-5.2%**	**0.4%**
Out-of-state resident	-	88.5%	5.0%	6.5%
Maryland resident	-	90.5%	6.3%	3.2%
Difference	-	**2.0%**	**1.4%**	**-3.3%**

Note: probabilities calculated using coefficients from Table 3 and sample means.