DOCUMENT RESUME

ED 445 650 HE 033 340

AUTHOR	DesJardins, Stephen L.; Moye, Melinda J.
TITLE	Studying the Timing of Student Departure from College. AIR 2000 Annual Forum Paper.
SPONS AGENCY	Association for Institutional Research.
PUB DATE	2000-05-00
NOTE	54p.; Paper presented at the Annual Forum of the Association for Institutional Research (40th, Cincinnati, OH, May 21-23, 2000). The AIR Grants Program receives funding from the National Science Foundation and the National Center for Education Statistics.
CONTRACT	99-104-0
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Persistence; *Attrition (Research Studies); College Students; *Dropout Research; Factor Analysis; Higher Education; Longitudinal Studies; Measurement Techniques; Methods Research; Models; Outcomes of Education; Research Design; Research Methodology; *School Holding Power; Statistical Analysis; *Student Attrition; *Time to Degree; Trend Analysis; Withdrawal (Education)
IDENTIFIERS	*AIR Forum; High School and Beyond (NCES); National Center for Education Statistics

ABSTRACT

This study describes the use of event history modeling as a tool for understanding student departure from college. Using data from the National Center for Education Statistics High School and Beyond (HS&B)/Sophomore Cohort longitudinal study a detailed history of individual students' college careers was constructed using a regression-like methodology often called "survival analysis" or "hazard modeling". The model used customized event history files that included student files (e.g., transcripts), course files (college courses, year-specific college credits, grade point averages, and cumulative college credits), and institution-specific information (type and selectivity of institution) to create indicators of enrollment status, yearly performance data, time to graduation. Other factors considered by the analysis included financial aid and grade point average. It was concluded that the study demonstrated the potential of using HS&B data to study student outcomes. And it also demonstrates how event history modeling can be applied to longitudinal data to study events that take place over time. A third outcome of the study is the information provided to researchers and policymakers on the factors that affect student departure from college and how these effects change over time. Four figures and 11 data tables are included. (Contains approximately 53 references.) (CH)

Studying the Timing of Student Departure from College

AIR Grant Invited Paper

Stephen L. DesJardins and Melinda J. Moye

The University of Iowa
Planning, Policy and Leadership Studies
491 Lindquist Center North
Iowa City IA 52242-1529

(319) 335-6413 phone
(319) 384-0587 fax
email: stephen-desjardins@uiowa.edu

May 22, 2000

The research reported in this paper was supported by a grant from the Association for Institutional Research (AIR Grant # 99-104-0) which receives funding for its "AIR Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education). The opinions reflect those of the authors and do not necessarily reflect those of the funding agencies.

We would like to thank Clifford Adelman, Dennis Ahlburg, Cynthia Barton, Dennis Carroll, Brian McCall, and Terry Russell for their assistance on this project.

Studying the Timing of Student Departure from College

Abstract

This study describes the usefulness of event history modeling as a tool for understanding student departure from college. Using the National Center for Education Statistics' postsecondary transcript file of the High School and Beyond/Sophomore Cohort and event history modeling we have constructed a detailed history of individual students' college careers. Methodologically, by using a procedure specifically designed to study temporal processes, we are able to remedy many of the analytic problems that have plagued earlier studies of student departure. Additionally, our findings address important issues for educational policy makers.

THIS PAGE INTENTIONALLY LEFT BLANK

Studying the Timing of Student Departure from College

INTRODUCTION

Understanding student attrition in institutions of higher education continues to be an important policy issue for institutional researchers and policy analysts (Bean, 1980; Cabrera, Nora, & Castaneda, 1992; Pascarella, 1986; Tinto, 1987). When individuals leave college before finishing a degree, costs are not only imposed on the individual, but also on the college or university and on society in general (Pascarella & Terenzini, 1991). Given the costs associated with failing to complete a college degree, educational policy makers must continue to search for better ways to understand how social, environmental, and individual factors affect student academic decision-making.

What matters most for college administrators, state legislators, parents, and students Adelman (1999) argues is completing a degree, not retention to the second year or persistence without a degree. There are many monetary and non-pecuniary benefits for those individuals who do graduate from college. College graduates earn twice as much as high school graduates and six times as much as high school dropouts (Murphy & Welch, 1993). Their financial assets are two and one-half times those of high school graduates and five times higher than high school dropouts (Diaz-Jiminez, Quadrini, & Rios-Rull, 1997). In addition to the financial rewards that accrue to the college graduate, others benefit as well. For example, the spouses of college graduates tend to be more highly educated, and their children do better in school and are less likely to disobey the law (Murphy & Welch, 1993; Jencks & Edlin, 1995). Despite the benefits and financial rewards of earning a college degree, the probability of completing a degree once college is undertaken is unacceptably low in the United States (DeBrock, Hendricks, & Koenker, 1996). Lederman (1991) reported that of all students enrolled in Division-I universities in 1984, only 48 percent had graduated by August 1989.

Statistics compiled more recently continue to indicate that a large number of students who enroll in higher education do not complete a degree. According to data presented in the Digest of Education Statistics (National Center for Education Statistics, 1997) 28.3 percent of the students who enrolled in a four-year institution in 1989-1990 had left college by 1994 without a degree. The dropout figures are especially troubling for certain minority groups and those who come from low and middle socioeconomic backgrounds (See Table 1).

Given that there continues to be a low completion rate for students who enroll in college, we hope to shed light on the timing of student departure by analyzing data from the National Center for Education Statistics. In the publication, *Answers in the Toolbox*, Adelman (1999) used longitudinal data from the National Center for Education Statistics' postsecondary transcript file of the High School and Beyond/ Sophomore Cohort database in order to examine the effects of academic intensity and attendance patterns on bachelor's degree attainment. However, Adelman did not employ a methodology that allowed him to explore the factors that might influence the *timing* of departure from college. In this paper, we use the same longitudinal data set as Adelman, however we apply *event history modeling techniques* in order to investigate time to bachelor's degree completion.

Table 1: National Completion/Attrition Rates

Student Characteristics	Attained B.A.	No Degree/Not Enrolled
	Status in 1994	
Total	45.8 %	28.3 %
Male	41.3 %	30.9 %
Female	50.3 %	25.7 %
Race/Ethnicity		
White, non-Hispanic	48.1 %	27.0 %
Black, non-Hispanic	34.3 %	36.8 %
Hispanic	32.4 %	36.6 %
Asian/Pacific Islander	46.8 %	25.5 %
Socioeconomic Status		
Low (bottom 25 percent)	22.1 %	51.8 %
Middle (middle 50 percent)	38.9 %	33.5 %
High (top 25 percent)	52.9 %	22.3 %

Note: These statistics are for those first-time students starting college during the 1989-1990 academic year.
Source: U.S. Department of Education, National Center for Education Statistics, Beginning Postsecondary Student Longitudinal Survey, 1994.

LITERATURE REVIEW

Two theoretical perspectives have dominated research into the factors that influence student attrition and persistence in college. The first is the Student Integration Model. This perspective, developed primarily by Spady and Tinto, emphasizes the predictive validity of *precollege* variables (Spady, 1970, 1971; Tinto, 1975, 1982, 1987, 1988, 1993). The second viewpoint is referred to as the Student Attrition Model, and it centers on aspects of student life that are external to the institution (Bean, 1978; 1980; 1981, 1982, 1983, 1992; Price, 1977).

The Student Integration Model emphasizes students' academic and institutional commitments. These commitments are influenced by the extent to which there is congruence between the motivations and academic ability of the student, and the institution's academic and social characteristics (Cabrera, et al., 1993). The Student Attrition Model emphasizes the importance of the intention to remain enrolled or to depart from college. The sequencing of events leading to departure is hypothesized to be the following: beliefs shape attitudes, and attitudes affect the intent to remain enrolled or to drop out. These two models can be distinguished primarily by the relative importance attributed to factors external to the institution. Specifically, the Student Attrition Model posits a much more important role for factors external to the institution in influencing student attitudes and enrollment decisions. While the Student Integration and Student Attrition Models have often been viewed as competing frameworks, Cabrera and associates (1993) have shown that this is not the case. Important components of the models overlap, and other aspects of the models are complementary. Cabrera and his colleagues offer an integrated model that yields a different understanding of the persistence process. Emphasis is placed on the structural specification of the psychological and sociological processes underlying persistence behavior.

In the publication, *Answers in the Toolbox* (henceforth simply "Toolbox"), Adelman (1999) describes what contributes the most to long-term bachelor's degree completion for students who attended four-year colleges. This particular database tracks individuals from the time they are in 10th grade in 1980 until they are approximately age 30 in 1993. This longitudinal database allows a long period for individuals to enter higher education, attend a four-year institution, and complete a bachelor's degree. Adelman incorporates high school and college transcript records, test scores, and information from surveys into his ordinary least squares and logistic regression equations. Adelman finds that the variable, Academic Resources, which is a composite of high school curriculum, test scores, and class rank, and whether or not an individual was continuously enrolled, accounts for the majority of his model's explanatory power in determining whether or not individuals are likely to complete a bachelor's degree.

Existing views of student departure have been effective at *describing* student departure but quite ineffective in *explaining* this process (Tinto, 1993). By shedding light on the *longitudinal* nature of the student departure process, we hope to provide an approach that will better enable us to describe *and* explain student departure from college. Although it is widely appreciated that the process of student departure from college is dynamic, previous studies have ignored the timing of the various ways students can exit higher education (stopout, dropout, transfer, academic dismissal, or graduation). Early studies of student departure have focused on a convenient time frame (year-to-year tracking), have examined departure before and after an arbitrarily chosen point in time (four or five years), or have focused on the student's first-year only. Event history models allow us to more exactly model the *timing* of student departure thereby permitting a more appropriate utilization of longitudinal data.

DATA SOURCES

In 1980, the U.S. Department of Education began collecting detailed information such as socioeconomic status and personal characteristics from a stratified random sample of 14,799 sophomores in United State's high schools. Follow-up interviews were conducted in 1982, 1984, 1986, and 1992 in order to obtain information on such variables as cognitive test scores, college enrollment and transcripts, type of institution attended, and labor force participation.

The sample used to study the timing of student departure was derived from the postsecondary transcript file of the High School and Beyond/Sophomore Cohort (here after HS&B) longitudinal study (NCES CD #98-135). This data file provides detailed event histories of students' higher education careers and outcomes. Unfortunately, not all of the variables used in the Toolbox study are on this CD. We determined that there was a newer version of the CD being processed that contained additional information used in the Toolbox study. We subsequently obtained a pre-release version of this CD and incorporated information from it as well (given we used a pre-release copy, the CD has not been assigned a name or number). In order to construct year-specific financial aid variables in our study, we use information from yet another CD, the HS&B Sophomore Cohort 1980-1992 (NCES CD #95-361). (All of these CD's are restricted data files and require a license obtainable from NCES).

As one author has recently noted, at present "this is the only data source in the nation that can answer the basic question about long term degree completion rates in recent years" (Adelman, 1998). These data files were developed by the National Center for Education Statistics (NCES) with the specific intent of providing longitudinal information about individual students' academic careers and outcomes. Thus, when studying student outcomes, such as graduation and dropout, the use of this data file is preferable to other nationally available data files because one can track *individuals* and can do so *over time*. Being able to track individual students is an improvement over procedures that use institutional-level data because aggregation often results in a loss of information. Using longitudinal data is also an improvement over cross-sectional designs because student departure is a process that takes place over time, and static designs do not properly incorporate the temporal dimension.

METHODOLOGY

The Technique

Event history models are a regression-like methodology that have long been used in other disciplines but have only recently been used to study educational issues. Event history models, which are often used in sociology (Rossi, et al., 1980), demography (Michael & Tuma, 1985), medical studies (Crowley & Hu, 1977), and economics (Kiefer, 1988) are a general class of statistical models used to study the occurrence and timing of events. Much of the terminology is a function of its roots in the fields of demography and biostatistics. Thus, biostatisticians usually call this method "*survival analysis*" since their interest is in studying how long patients survive under a treatment regime. Other disciplines use different nomenclature to describe this method. Economists often call the technique *hazard modeling* because they are interested in the *hazard* or risk of event occurrence (e.g., leaving unemployment) at a particular point in time. Since the fundamental outcome of interest in event history models is the amount of time that it takes until an event occurs, some disciplines call this method *duration analysis*. In engineering, event history models are known as *failure time* or *reliability* analysis since they are often used to study the failure rates of electronic components, physical structures, and machinery. Even though there may be technical differences among these methods, they all share a common approach to analyzing longitudinal events: studying not only the occurrence but also the timing of events.

The Hazard Rate

Before proceeding, it is imperative that we discuss a number of key concepts used in event history modeling. The first is the concept of the *hazard rate*. The hazard rate (and a related measure the *survival rate*) is "an unobserved variable, yet it controls both the occurrence and timing of events and it is the fundamental dependent variable in an event history model" (Allison, 1984, p.16). Even though the hazard rate (often shortened to the *hazard*) is unobservable, we make inferences about it, and about individuals' choices, by observing the "movement of persons between states" (Lancaster, 1990, p.5). For instance, when a student makes a decision to remain enrolled or drop out of college the student is making a choice. Utility theory suggests that individuals will leave college when the future benefits fail to outweigh the future costs of attendance. Thus, by studying students' transitions from enrollment to not being enrolled, we are indirectly examining students' benefit/cost calculations with regard to college attendance. That students make college continuation decisions on internal optimality conditions is an often overlooked, but important theoretical point. Tinto's model (1975) assumes that "a person will tend to withdraw from college when he perceives that an alternative form of investment of time, energies, and resources will yield greater benefits, relative to costs, over time than staying in college" (p.97-98). Lancaster also notes that "it is this choice component that distinguishes the econometric analysis of transition data from standard applied statistical analysis of survival and transition data and gives a richness but also an added complexity to econometric work" (Lancaster, 1990, p. 6).

The hazard rate "expresses the instantaneous risk of having the event at time *t* given that the event did not occur before time *t*" (Yamaguchi, 1991, p. 9). The hazard rate is calculated by dividing the number of students who experienced the event in question (i.e., graduation) in a particular time period (year) by the number enrolled in that time period. For instance, a hazard rate of .48 (see Figure 1) indicates that students who survive until year four have about a 48 percent chance of graduating in that year. Another measure often cited in event history research is the survival rate. This statistic is defined as the probability of *not having* the event of interest (graduation) prior to time *t* (in this case, year four). The survivor function is a non-increasing function of time (most often it declines) indicating that the event of interest (i.e., graduation) has not occurred and that the person is still enrolled (for more on the mathematics of these functions see Blossfeld, Hamerle, and Mayer, 1989).

[INSERT FIGURE 1 ABOUT HERE]

The Risk Set

The *risk set* is typically defined as the "set of individuals who are at-risk of event occurrence at each point in time" (Allison, 1984, p.16). In this study, the initial at-risk pool consists of students who ever set foot in a four-year institution. As time passes, the at-risk pool diminishes as some students have the event (graduation) or another *terminating event* (dropout). With regard to the latter, when studying dropout, graduation is a terminating event, because students who graduate are no longer subject to dropping out of the institution. Thus, in any given time period, enrolled students are subject to having a terminating event and are therefore the relevant risk set. When studying dropout and graduation the at-risk pool can become very small in later years, since many students have some terminating event in earlier years. When the at-risk pool is very small, the occurrence of a few events will result in "spikes" in the hazard function (see Figure 1, years 10 & 11).

Censoring

Standard analytic methods, like linear and nonlinear regression techniques and structural equation models do not adequately handle *censored* observations. Censoring takes place when the outcome or event of interest is not determinable for each individual within the observation period. In studies of student dropout or graduation, for instance, some students remain enrolled beyond the end of the observation period (in this study the observation period is 12 years). When this is the case it is impossible to determine if, or when, a given student had the fundamental outcome of interest—the time to the event. This phenomenon is known as *right censoring* and has been shown to cause estimation problems such as severe bias or loss of information when standard regression techniques are used to estimate longitudinal events (Allison, 1984).

When attempting to use standard regression techniques to model longitudinal processes, researchers typically deal with right censored observations by excluding them from the sample. However, if there are a large number of these cases substantial estimation bias will result (Sorensen, 1977). Another strategy often employed is to assign the maximum value of duration to individuals who are right censored thereby keeping them in the sample. However, this "data fix" causes an underestimate of the time to the event.

Though there are problems trying to deal with censored observations when applying standard statistical techniques to longitudinal data, event history methods are not hampered by this limitation. Information about right censored cases is easily incorporated and therefore the technique is preferred over methods like conventional linear regression or path analytic techniques when analyzing longitudinal data (Yamaguchi, 1991). Inclusion of right censored observations eliminates any bias due to removing censored observations and more accurately estimates the time to the event than when censored cases are given the maximum duration value. A more appropriate handling of right censored cases should, therefore, provide researchers with a more accurate picture of the temporal process being studied.

Time-Varying Variables and Coefficients

When analyzing longitudinal data, the values of some variables used to explain the event of interest may change over time. In the application provided below, we include variables whose values can change over time. We include year-specific grade point average and financial aid factors in an effort to demonstrate how these variables, and their effects, can vary over time. It is important to distinguish between *variables* whose *values* can change over time and regressors whose *effects* are time varying. The former means that we observe different measurements for variables over time. For instance, one's grade point average may change from 3.00 to 2.50 from one year to another. In the model presented herein, the values that these time-dependent variables can take are permitted to change by year. This is a function of the way the data were constructed—year by year.

Allowing the *effects* of variables to change over time is, however, a function of the statistical properties of the event history model used. In effect, a separate regression is estimated for each year, thereby providing us with different coefficient estimates for each regressor for each academic year. Thus, the model allows *all* of the regressors included to have *time-varying coefficients*, and therefore *time-varying effects* for each year. Allowing the betas to change over time permits variables often thought to be time-invariant (i.e., gender) to display *effects* that vary over time. This happens because we obtain a different coefficient estimate for gender in each year. This is an important distinction and one that readers can find more information about in Lancaster (1990).

For instance, factors that affect student departure certainly have differential effects over time. As students' academic careers progress, a host of factors may become increasingly (or decreasingly) important indicators of whether students remain enrolled. Techniques typically used to study student departure are not easily adapted to include time-varying covariates, and problems in estimation may ensue. According to one author, "there is simply no satisfactory way of incorporating time-varying explanatory variables in a multiple regression predicting time of an event" (Allison, 1984, p.11). Since event history models explicitly allow for inclusion of *time-dependent covariates*, it is preferred to standard regression and path analytic techniques when studying longitudinal events.

Controlling for Unobserved Population Differences

Controlling for unobserved or unmeasured factors is especially important when dealing with complex behavioral processes like student departure because specifying all possible explanations of such a process is an impossible task. Even if all *observed* explanatory variables are included in a model there may still be unknown or unobserved factors that affect the outcome. In general, not controlling for unmeasured or unknown sources of variability in statistical models will cause biased coefficient estimates (Trussell & Richards, 1985).

Unobserved or omitted variable problems can be particularly troubling in event history modeling. In the presence of *unobserved heterogeneity* (henceforth UH), it is difficult to determine if the estimates of time to the event of interest (i.e., graduation) are providing a true representation of the hazard (of graduating). The hazard rates we observe when UH is present may be caused by simple variations in the risks of graduating across individuals that is related to their different, but unobserved characteristics. This may happen when students with greater chances of graduating do so early in their academic careers and are therefore eliminated from the risk set. Educational theory suggests that students with high "hazards" of graduating early (i.e., the 4th year) may be more motivated or have higher aspirations than other students do.

The bias that results from model misspecification due to UH is typically a difficult problem to remedy in event history models. Advances in statistical modeling have, however, made it possible to control for unobserved sources of heterogeneity. Thus, controlling for UH is important if we are to avoid observing spurious declining hazard results. If, on the other hand, we observe that a hazard function increases we can safely imply that the underlying hazard really does increase over time (Allison, 1984).

Conceptualizing the Empirical Model

The empirical model used is a *discrete-time* event history model since the duration data is collected year-by-year until the occurrence of the event (graduation). Graduation is defined as the awarding of a bachelor's degree any time within the (approximately) twelve-year observation window.

This study is a continuation and enhancement of event history models developed over the past several years (DesJardins, 1993; DesJardins, et al., 1994, DesJardins, 1996; DesJardins, et al., 1998; DesJardins, et al., 1999b). The samples used in the previous studies, however, were from a single Research I institution. Although the results of these studies have influenced policy making within this institution, the results are not generalizable to higher education as a whole. Therefore, applying event history modeling to a *national* sample is a logical extension of the single institution studies.

Most studies of student departure are typically deficient in that they do not adequately account for the longitudinal nature of the data, are suboptimum in their handling of censored data, and often do not adequately control for unmeasured or unobserved factors that could help to explain student departure. Our single institution research has clearly shown that it is important to adequately include time as a factor, and that controlling for unobservables is important when trying to explain student departure from college (DesJardins, et al., 1999a; 1999b).

The Empirical Model

The HS&B data file is longitudinal in nature, therefore using a methodology designed specifically to study events that take place over time seems appropriate. The empirical model that was used to study this data is a discrete-time event history (hazard) model since the transcript file data is collected at discrete points in time. We examined the time to graduation and dropout, though only the graduation results are reported in this paper. Graduation is defined as the awarding of a bachelor's degree anytime within the observation period of the sample (1982 to 1993).

Formally, the discrete-time event history model is described by the following set of conditions. Let K be a discrete random variable measuring the number of periods until an event occurs. It is assumed that the event of interest is influenced by a vector of time-varying explanatory variables (z_k). Let $P(K=k \mid K > k-1, z_1, \ldots, z_k)$ represent the conditional probability that the event occurs in period k given that it has not occurred by k-1, where $z_1, \ldots, z_k$ represents the set of values of the explanatory variables until time k. The standard model for this conditional probability is the (discrete-time equivalent of the) proportional hazards model (see Cox, 1972; Prentice and Gloeckler, 1978; Meyer, 1986, 1990; and Han and Hausman, 1990):

$$P(K=k \mid K>k-1, z_1, \ldots z_k) = 1-\exp(-\exp(\alpha_k + \beta z_k)) \qquad \textbf{(1)}$$

where β is a vector of coefficients that measure the effects of the explanatory variables (z_k) and α_k is a time-varying constant term, k=1,2,3,.... An alternative model, typically used in earlier education-related studies using event history methods (Singer and Willett, 1991; DesJardins, 1993; Ronco, 1996) is the "logit" model:

$$P(K=k \mid K>k-1, z_1, \ldots z_k) = \exp(\alpha_k + \beta z_k) / 1+ \exp(\alpha_k + \beta z_k)) \qquad \textbf{(2)}$$

Like the structural and path analytic models used in earlier student departure research, one drawback of model (1) is that it assumes that *all* of the determinants of the event being studied are accounted for by the explanatory variables (z_k). Model (1) also assumes that the *effects* of the explanatory variables are constant over time. Violations of either of these assumptions, which are common when doing social science research, may cause biased estimates.

The model outlined below (3) generalizes model (1) by allowing for time-varying effects *and* includes a variable that controls for unmeasured or unobserved factors. The model below is therefore a substantial improvement over the proportional hazards model presented in (1) (see McCall, 1994 for details). Let $P(K=k \mid K > k-1, z_1, \ldots, z_k, \theta)$ represent the conditional probability that the event occurs in period k given that it has not occurred in the first k-1 periods of enrollment. The values of the time-varying regressors in periods 1 through k ($z_1, \ldots, z_k$) are observed and the unobserved variable is specified as θ. Thus, the model we will use is

$$P(K=k \mid K>k-1, z_1,\ldots z_k, \theta) = 1-\exp(-\exp(\alpha_k + \beta_k z_k')\theta) \qquad \textbf{(3)}$$

where β_k measures the (possibly time-varying) effect of z_k in period k and α_k is again a time-varying constant term, k=1,2,3,.... Model (3) is estimated by maximum likelihood and non-parametric maximum likelihood techniques (see Heckman & Singer, 1984). The latter method was demonstrated to be feasible by McCall (1994).

The explanatory variables $(z_1,\ldots,z_k)$ used in (3) includes pre-college, demographic, college achievement factors, and financial variables that were available in the HS&B data file. The independent variables included in the models are chosen based on theoretical considerations, previous research on student departure, and the event history research we have done to date.

Constructing the Event History Data File

Although it is possible to apply event history methods to the data elements provided in the various HS&B CD's discussed above, we decided to construct a customized event history file. We did this for two reasons: 1) the HS&B data and elements limit one's ability to define time in more discrete terms than is possible if one constructs their own data set; and 2) construction of the event history file required the authors to learn more about the national data and this is one of the objectives of the AIR grant.

Construction of the event history file began using CD #98-135, however, as mentioned above, not all variables used in Adelman's Toolbox study are contained on this CD. NCES graciously provided us with a pre-production release of an update to CD 135 (henceforth "preliminary"). However, the preliminary CD does not include the financial aid detail that we require so we obtained a copy of CD #95-361. This disk contains all the HS&B data, a very valuable source of longitudinal information.

Using the preliminary CD, an initial cohort of students for whom transcripts and complete records are available are selected (henceforth the "student" file). Then a file of college course data is constructed (henceforth the "course" file) for students with enrollments between fall of 1982 and the end of 1993. For each year of enrollment, year-specific college credits and grade point averages, and cumulative college credits are calculated. This course file is then matched against a file containing institution-specific information like the type and selectivity of the institution attended, and the institution (FICE) code.

The "student" and "course" files are then matched to create indicators of enrollment status (enrolled or not, by year), yearly performance data (GPA and credits), and descriptions of the institution(s) attended in any given year. The results are then matched against a file of degree recipients to obtain the year in which a student received their *first bachelor's degree*. The enrollment status variable created above is then updated to include the year of graduation in addition to whether a student is enrolled, or not. This file is used to match against other data files on the preliminary CD in order to attach variables that are included as regressors in the statistical model. Also, the cohort file is matched against a file from CD #95-361 that contains financial aid information. This match allows us to construct information about students' financial aid receipt (loans, grants, work/study) by year (but only through 1985). Variables specifically used in event history models (time to degree, censoring, GPA lags) are added to the above and the final result is a longitudinal file that contains a large number of variables describing each student's background and academic history.

Table 2 provides additional detail about the event history file construction. As you can see, the effective sample size is 4,681. Since one of our objectives is to replicate Adelman's Toolbox results, we restrict the sample to students who have ever attended a four-year institution. Most of these students (78%) enrolled some time in the academic year following high school graduation (fall of 1982). Of the 1571 students who delayed college entry, 42% of them delayed by only one year and 57% of them enrolled within two years.

[TABLE 2 ABOUT HERE]

EMPIRICAL RESULTS

Descriptive Results

A description of the effective sample is presented in Table 3 below. The Academic Resources variable is an index measuring the academic resources students bring forward from secondary school (see Adelman, 1999, for details on the construction of this variable). One would expect that the better prepared a student is for college, the more likely they are to graduate. In our analysis we substitute the three components of the Academic Resources variable (Academic Intensity/Quality, High School Rank, and Senior Year Test score; all measured in quintiles) to examine how these variables operate independently. It may be that aggregation of these variables into an index is masking, especially over time, variation in the way these factors influence student graduation chances.

We also include a variable to indicate whether a student is a parent by 1986. As Adelman notes we must "account for change in family status as students move into their early and mid-20's" (1999, p. 37). Our expectation is that parenthood will negatively impact a student's chances of graduating from college. A quintile version of SES is included with the expectation being that higher SES individuals will tend to graduate in greater numbers than their lower SES colleagues. A variable indicating student "Anticipations" (see Adelman, 1999, p. 32-35 for a discussion) is included to account for whether a student believes they will earn a bachelor's degree. Gender is included as an indicator variable where males=1 and females=0. Race is included as a dummy variable indicating whether a student is African American, Hispanic, or Native American (Whites and Asian Americans are the reference group).

[TABLE 3 ABOUT HERE]

Financial aid and student work-related variables are also included. However, "the HS&B/So limits our utilization of financial aid variables" (Adelman, 1999, p. 63) to dichotomous representations since the dollar amounts that students report may not be reliable. We expect, however, that using multiple sources to obtain aid-related information and then dichotomizing to a yes/no variable will tend to mitigate this potential source of bias. Also, since aid information is not available in HS&B/So beyond 1985, we include a dummy variable (After 1985) indicating whether a student is enrolled beyond 1985.

Actual college performance (GPA) has been shown to be an important determinant of college graduation. Adelman uses both freshman GPA and the ratio of final college GPA to first-year college GPA (in an attempt to test for a "trend" in college performance; 1999, Table 37, p.75). We allow GPA to vary by year and have demonstrated (DesJardins et al., 1999a) that one's performance in college has effects on graduation that vary over time. The objective herein is to examine whether grades have a time varying effect on bachelor's degree completion.

Modeling Time to Graduation

When event history techniques are applied to education-related topics, researchers typically use a "single risk" specification. This type of model allows the researcher to estimate single events, like time to graduation, or duration to stopout or dropout (for instance, see Ahlburg, et al. 1997; Chizmar and Cummins, 1994; Guerin, 1997; Murtaugh, et al., 1999; Ronco, 1994; Xiao, 1997 for single-risk departure studies). In this study we use a single risk model to examine how a number of factors influence time to graduation. The objective is to test if the event history results are similar to the Toolbox study results obtained using logistic regression.

The model building process begins by estimating "naïve" models and making these models more complex as we progress. The naïve results are used as "benchmarks" against which we can compare the results of Adelman's research. One objective of this model building approach is to determine whether the addition of time-varying coefficients and controls for unobserved student differences improves the model fit. If so, then we have statistical evidence that an event history approach is an improvement over cross-sectional or time-constant methods.

Time-Constant Coefficient Models

The "Baseline" Model

The first models estimated (displayed in Table 4), are designed to mimic Adelman's "Background Model" (1999, p. 62, Table 30). The models displayed are time-constant coefficient models (henceforth TCC). These models do not incorporate time and are therefore similar to the logistic regression approach used by Adelman. There are three different models estimated. The first (No UH) does not contain a control for unobserved factors (i.e., there is no "error" term). The second model (Gamma) is a parametric approach in which unobservable or unmeasured factors are assumed to be gamma distributed. The third approach allows the distribution of the unobserved heterogeneity control to be determined by the sample. Elsewhere it has been demonstrated that this "flexible" approach is typically preferred (based on model fit) over parametric specifications (DesJardins, et al., 1999a). The results of all of the TCC models are very similar to Adelman's findings. It appears, however, that the models containing UH controls fit the data better (based on the likelihood values) than the model that does not control for unobservables. This result demonstrates the utility of controlling for UH when using event history models.

[TABLE 4 ABOUT HERE]

The "Financial Aid" Model

The next model (see Table 5) estimated adds financial aid variables (loans, grants, and work study). This approach again parallels that of Adelman. The same model testing process noted above is conducted whereby we estimate the new set of variables with no UH, gamma distributed UH, and a flexible heterogeneity control (this strategy remains the same throughout the study). Again, the addition of UH controls improves the model fit but the results begin to diverge slightly from what Adelman found (1999; see Table 32, p. 65).

[TABLE 5 ABOUT HERE]

The "GPA" Model

This is where the parallel to Adelman's research diverges. In our next model, we add college grade point average to the model noted above in an attempt to see how actual performance changes our estimates. In the Toolbox study, after estimating the financial aid model, Adelman includes variables relating to the number, order and characteristics of the institutions a student attends. We decided to diverge from Adelman's strategy because we feel the process of institutional choice, enrollment, and reenrollment is endogenous and the inclusion of a large number of endogenous variables may bias the estimates.

When actual performance, as measured by college GPA, is included (see Table 6) the effect of the Academic Resources variable is cut by about 41% (from .4241 to .25). This should not be a surprise as the Academic Resources variable is probably an indicator of student potential, whereas one's GPA in college provides evidence of *realized* potential. The results indicate that college GPA is a very powerful predictor of bachelor's degree attainment. For instance, every one-grade increase in GPA more than doubles a student's chances of graduating. Also, when we control for GPA, the negative relationship between graduation and being a member of a minority group is less pronounced (the coefficient value changes from -.3049 to -.1596, a change of 58%).

[TABLE 6 ABOUT HERE]

The "Components" Model

The final time constant coefficient model estimated is designed to test whether the three components of the Academic Resources variable (Academic Intensity/Quality, High School Rank, and Senior Year Test score) have differential effects. It may be that these variables operate differently and that using them in an index masks their independent effects.

[TABLE 7 ABOUT HERE]

Our results (see Table 7) indicate that the Academic Intensity/Quality and High School Rank variables are significantly and positively related to bachelor's degree completion. In this model specification, we find no evidence that the Senior Year Test is significantly related to graduation. We will again use this model specification when the time-varying coefficient models are estimated to test the stability of these results.

Summarizing the time-constant coefficient results, we find that models that incorporate UH are (generally) preferred to models that do not control for unobservables. The results of the "background" model and the model that includes financial aid are similar to those found by Adelman. When actual performance in college is controlled for (by GPA) other effects wane (i.e., Academic Resources, the Minority effect). Finally, it appears, though more testing needs to be done, that using the Academic Resources index masks the differential effects of its components.

Time-Varying Coefficient Models

The same model building strategy demonstrated above is used to examine how the same regressors are related to graduation when we include a more appropriate incorporation of time. Given that graduation is a longitudinal process, theory, and prior event history research (Ahlburg, et al., 1997; DesJardins, et al., 1999a) suggests that allowing the values of variables, and their effects, to change over time is an improvement over time-constant approaches.

Generally, our results indicate that time-varying coefficient (TVC) models are an improvement over the time-constant models (based on model fit). Again we find that including controls for UH is preferred to models that do not account for unobserved factors. Also, we again find that flexible UH controls fit the data better than the parametric (gamma) specifications tested. The flexible approach is less restrictive because using a parametric distribution may impose unrealistic constraints on the sample. To conserve space, we will briefly discuss the differences in the models estimated and include the tables of the time-varying results in the appendix.

In general, the time-varying models allow us to observe the relationship between the regressors and graduation *and* the direction that these relationships take over time. For instance, in even the most basic model ("Baseline" with No UH in Table 8), we observe that the negative relationship found in the TCC models between graduation and being a parent changes over time. The hazard of graduation is initially negative (-1.728 from Table 8 column 1, "Baseline Effect"), but the hazard rate trend is positive as time passes (Table 8, column 2, "Trend"). For a graphic example look at Figure 2. Using the coefficients in Table 8, we calculated year-specific hazard rates and plotted them for three variables. As you can see, in year four students who are parents have very low hazard rates, but as time passes this function moves toward zero (a positive trend). Conversely, students with high academic resources are more likely to graduate (Baseline Effect of .5590) but the hazards for these students decline over time (-.0504 "Trend" value and a declining hazard function in Figure 2). Also note in Figure 2 that male students are initially less likely to graduate than females, but between years 6 and 7 this effect reverses.

[TABLE 8 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

Like we did for the TCC models, we display four tables of results for the TVC models estimated (Baseline, Financial Aid, GPA, and Components). For each of these models we again display the results assuming different types of UH controls (No UH, Gamma, and Flexible). What is different about the TVC models is that we now have two sets of coefficients (and standard errors) for each model estimated. The inclusion of two sets of estimates reflects the longitudinal nature of the TVC approach.

The addition of this time dimension permits us to observe that variables thought to have constant effects over time have effects that change over the course of a student's academic career. For instance, the TCC Baseline results discussed earlier in the paper indicate a negative relationship between graduation and being male. The TVC models demonstrate, however, that this relationship varies over time, and eventually *reverses*. These empirical result differences are indicative of how a proper inclusion of time can reveal important temporal effects.

Another difference between the TCC and TVC models is that we begin to see marked differences in the estimates depending on how UH is controlled. For instance, and as mentioned above, students who are parents by 1986 are initially less likely to graduate, but this negative effect seems to "wane" over time (see Figure 2). But when UH is controlled for the trend effect for this variable becomes insignificant. There are many reasons why we observe different results when UH is controlled. For example, as the sample of students changes over time so does their observable characteristics (GPA's tend to inflate over time, the racial makeup of the sample changes). Given that this is the case, undoubtedly students' *unobservable* characteristics also change over the observation period. Just as it is very important to be able to control for changes in observed differences among students, it is also important to control for the changes we do not observe. Theory and prior experience tell us that models using flexible controls for UH typically provide a better fit to the data, so from here on we will focus our discussion on the Flexible model results only.

Adding Financial Aid to the Baseline Model

When financial aid is added to the TVC baseline model discussed above we again observe time-varying results. A student's commitment toward a bachelor's degree is initially positively-related to graduation, but this effect becomes less pronounced over time. Loans and grants are positively related to graduation but the hazards decline as time passes (see Figure 3 or Table 9). Work/study assistance is initially negatively-related to graduation but this effect actually reverses signs in about year 6 (see Figure 3).

[FIGURE 3 ABOUT HERE]
[TABLE 9 ABOUT HERE]

Because the NCES data does not contain financial aid information after 1985 we included a dummy variable to control for students who matriculate after 1985. The inclusion of this dummy tests whether students who delay entry have hazards of graduating that are different from students who initially enrolled in the 1982-1985 period. The formula displayed below is how we convert the coefficients produced by the hazard model to arrive at the year-specific hazards graphed in the Figures above. The calculation of year-specific hazards may help to better understand how the After

1985 variable operates. In general:

$$b(t) = b_o + b_1\,(t\text{-}1) \qquad \textbf{(4)}$$

Where b(t) is the year-specific hazard rate, b_o is the Baseline Effect, b_1 is the Trend coefficient, and t is the year in question. Using (4) and the After 1985 results from Table 9 (Flexible model), we estimate the hazard of graduating in the fourth-year for students who start college after 1985.

$$b(4) = 5.4473 - 1.5981\,(4\text{-}1)$$

$$b(4) = 5.4473 - 4.7943$$

$$b(4) = .653$$

This result suggests that students who delay matriculation until after 1985 are about exp(.653) = 1.92 times *more* likely to graduate in four years than students who matriculate in 1982-1985. This result seems counterintuitive, but if we examine the year-specific hazard rates we find that this effect reverses the very next year (to -.94). The -.94 hazard indicates that students who start school in 1982-1985 are 2.6 times more likely to get a bachelor's degree in year five than students who start after 1985. As time passes, the negative graduation effect for students who delay matriculation gets very large (in years 6-12).

Controlling for College Performance

As we did for the TCC models, we add college GPA to the TVC financial aid model discussed above. We focus on the results of the Flexible model and find that when GPA is added, the Academic Resources Baseline Effect remains positive, but the effect of this variable drops by about 36% compared to the financial aid model. A similar result is found for minority students. When GPA is added the Minority coefficient loses about 43% of its negative effect (-.4564 to -.261; see Table 10). Another interesting difference between the aid model and the model that includes GPA is that SES is not significant in the latter. We find that GPA is initially positively-related to graduation but the trend is negative. GPA is a very important predictor of graduation but as time passes the effect is less pronounced.

[TABLE 10 ABOUT HERE]

In the model that includes GPA we find that the Grant variable becomes insignificant compared to the financial aid model results. This result deserves more in-depth analysis but it may be that students who remain in school beyond year four are students for whom the Grants variable is comprised mostly of merit-based aid. If this is the case, then when GPA is not included in the model we would expect this source of variation to be picked up by a correlated measure, like Grants.

[FIGURE 4 ABOUT HERE]

The Time-Varying "Components" Model

The final TVC graduation model includes the components of Academic Resources rather than the actual index used by Adelman. The results indicate that Academic Intensity/Quality is not signficantly related to graduation initially, but that it helps students graduate later in their academic careers (it has an insignificant

Baseline Effect and a significant and positive Trend; see Table 11 or Figure 5 for a graphical display). High School Rank is positively related to graduation but displays a weak and statistically insignificant trend. The Senior Year Test is initially positively-related to graduation, but this effect wanes and then reverses around year five. Thus, we again see that the components of the Academic Resource index have differential effects. These effects vary depending on the 1) the type of model estimated (TCC vs. TVC), 2) the controls used for UH, and 3) the time period observed.

[FIGURE 5 ABOUT HERE]

[TABLE 11 ABOUT HERE]

In summary, methodologically we find that models that incorporate time are an improvement over time-constant models, controlling for unobserved and unmeasured factors is important, and that factors often thought to have effects that are time-constant actually have time-varying effects. Substantively, we find that the Academic Resources index used in the Toolbox study positively affects graduation hazards but that when GPA is included the effect of this variable becomes less powerful. Also, it appears that more research needs to be done into how the component parts of the Academic Resource index affect graduation. We also find that the negative effect associated with membership in a minority group becomes less so when controls for aid and GPA are included. The Anticipations variable displays remarkably consistent results over all of the models estimated. The effect of being a male is negatively related to graduation no matter which model is estimated. Financial aid variables exhibit different patterns depending on whether time is controlled for or not. Given our preferred TVC - "GPA" model, we believe that loans enhance and work/study inhibits timely-graduation, but the work/study effect reverses around year 6. Finally, as one would expect, GPA is positively related to graduation, the interesting point is that GPA hazards remain quite constant over time.

LIMITATIONS

Sample weights are not used in our models given that our sample was a nonrandom subset of the entire HS&B sophomore cohort. As mentioned earlier, our subset includes only members of the HS&B sophomore cohort who had at least one postsecondary transcript (transcript record must be complete), had information in the course table, and were enrolled in a four-year institution at some time during 1982-1993. Thus, applying sample weights would not have created a representative sample of students who were enrolled in a four-year postsecondary institution during this time period. This is consistent with how Murnane, Willett and Tyler (2000) handled their nonrandom subset (individuals who took the tenth-grade math test and who reported earnings for 1990 or 1991) of the total HS&B sophomore cohort when investigating who benefits from obtaining a GED.

We did, however, test the sensitivity of the significance of our results by adjusting the standard errors in the models using a "design effect" (see Adelman, 1999, pp. 109-112) of 1.55. This value was chosen since Adelman uses it in a number of his models and we believe that increasing standard errors by 55% is a conservative approach. We find that the significance of the variables remains virtually identical to our unadjusted results.

Another limitation for this particular study is that we restricted our sample to only those individuals who were enrolled in a four-year institution at some point during 1982-1993. This is appropriate for examining bachelor's degree completion rates. However, this is a restriction that we may wish to relax in future studies in order to examine enrollment patterns for individuals who attended community colleges or other institutions.

In a future study we will construct a competing risks model of graduation and dropout. Estimating this type of model will allow us to test whether the single or independent risk outcome (examining graduation only) estimates are sensitive to the single risk assumption. It may be, for instance, that the single risk graduation model is a misspecification due to functional form because we are inadequately controlling for the interdependence of graduation and dropout.

IMPLICATIONS

The findings of this research are significant for research, practice, and policy making in postsecondary education. First, this study demonstrates to educational researchers the potential of using the HS&B postsecondary transcript file for studying student outcomes. Second, detailing how event history modeling can be applied to longitudinal data informs researchers about a new approach to study events that take place over time. Since many educational events are longitudinal in nature (tenure decisions, student transfer, faculty and staff turnover), providing information about this methodology should improve researchers' understanding of an important analytic tool. Third, our results provide researchers and policy makers with a better understanding of the factors that affect student departure from college and how these effects change over time.

FUTURE RESEARCH

Currently, we are estimating models of student dropout using the strategy and methods explained above. This analysis will provide more insight into the factors that are related to why students decide to leave college. Given that we now possess a national dataset, we will be able to model both institutional dropout and system-wide dropout. The former is important for institutional policy makers while the latter is an important social policy issue.

There are a number of additional analyses that we intend to do in the near future. In addition, we would also like to model competing or correlated risks in future studies. As mentioned above, a competing or correlated risk model will be estimated to examine the interrelationship between graduation and dropout. (see DesJardins, et al., 1999 for a demonstration of this modeling approach).

In the future we intend to model the selection process into higher education. To do this we need to expand the cohort to all high school graduates (and GED's). This type of analysis will allow us to not only model the timing of bachelor's degree attainment, but also to expand the analysis to individuals enrolling in and graduating from different types of institutions of higher education (i.e. community colleges, two-year trade schools). This type of analysis will enable us to model the transitions between institutions, how delayed entry and stopout affects graduation timing, and the path that students take toward two- and four-year degrees.

Finally, we intend to more closely examine the most effective use of academic resources-related variables. This construct appears to be very important in explaining bachelor's degree completion. We need to understand better, however, how the components of this construct can be used, especially in models that allow effects to vary over time. The ultimate objective of this line of analysis is to see if the components of these variables have effects that differ over time.

REFERENCES

Adelman, C. (1998). "More than 13 ways of looking at degree attainment." National Cross-Talk, vol 6. no. 4. Washington DC: National Center for Public Policy and Higher Education.

Adelman, C. (1999). Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment. Jessup, MD: U.S. Department of Education.

Ahlburg, D.A., McCall, B.P., and Na, I. (1997). "Time to Dropout from College: A Hazard Model of College Dropout with Endogenous Waiting Time". Working paper, Industrial Relations Center, University of Minnesota.

Allison, P. D. (1984). Event history analysis: Regression for longitudinal event data. Sage University Papers: Quantitative Applications in the Social Sciences, 07-046. Newbury Park, CA: Sage Publications.

Bean, J.P. (1978). "Dropouts and turnover: The synthesis of a causal model of student attrition". Unpublished Doctoral Dissertation, The University of Iowa.

________ (1980). "Dropouts and turnover: The synthesis and test of a causal model of student attrition". Research in Higher Education, Vol.12: 155-187.

________ (1981). "The synthesis of a theoretical model of student attrition". Paper presented at the American Education Research Association's annual meeting, Los Angeles.

________ (1982). "Student attrition, intentions, and confidence: Interaction effects in a path model". Research in Higher Education, Vol.17: 291-319.

________ (1983). "The application of a model of turnover in work organizations to the student attrition process". Review of Higher Education, Vol.6: 129-148.

________ (1992). "Student dependency theory: An explanation of student retention in college". Paper presented at the Annual Meeting of Association for the Study of Higher Education, Minneapolis.

Blossfeld, H., Hamerle, A., and Mayer, K. U. (1989). Event history analysis: Statistical theory and application in the social sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Blossfeld, H., and Rohwer, G. (1995). Techniques of event history modeling: New approaches to causal analysis. Mahwah, NJ: Lawrence Erlbaum Associates.

Cabrera, A.F., Nora, A., and Castaneda, M.B. (1992a). "The role of finances in the persistence process: A structural model". Research in Higher Education, Vol. 33, No. 5.

________ (1993). "College persistence: Structural equations modeling test of an integrated model of student retention". Journal of Higher Education, Vol. 64, No. 2.

Chizmar, J., and Cummins, W. (1994). "Discrete-time survival analysis of the role of gender in persistence in the economics major". Paper Presented at the 34th Annual Forum of the Association for Institutional Researchers, New Orleans, LA.

Cox, D. R. (1972). "Regression models and life-tables (with discussion)". Journal of the Royal Statistical Society, B, 34.

Crowley, J., and Hu, M. (1977). "Covariance analysis of heart transplant survival data". Journal of the American Statistical Association, 72, 557-565.

DeBrock, L., Hendricks, W., and Koenker, R. (1996). "The economics of persistence: Graduation rates of athletes as labor market choice," Journal of Human Resources. 31: 513-539.

DesJardins, S.L. (1993). "Using hazard models to study student careers". Paper Presented at the 33rd Annual Forum of the Association for Institutional Researchers, Chicago, IL.

________ (1996). "Using event history methods to study the temporal dimensions of student departure from college". UMI Dissertation Document No. 9706199.

DesJardins, S.L., Ahlburg, D.A., and McCall, B. P. (1994). " Studying the determinants of student stopout: Identifying "true" from spurious time-varying effects". Paper Presented at the 34th Annual Forum of the Association for Institutional Researchers, New Orleans, LA.

DesJardins, S.L., Ahlburg, D.A., and McCall, B. P. (1998). "Simulating the longitudinal effects of changes in financial aid on student departure." Paper Presented at the 38th Annual Forum of the Association for Institutional Researchers, Minneapolis, MN.

DesJardins, S.L., Ahlburg, D.A., and McCall, B. P. (1999a). An event history model of student departure. Economics of Education Review, 18: 375-390.

DesJardins, S.L., Ahlburg, D.A., and McCall, B. P. (1999b). Investigating factors related to timely degree completion. Paper Presented at the 40th Annual Forum of the Association for Institutional Researchers, Seattle, WA.

Diaz-Jiminez, J., V. Quadrini, and J-V. Rios-Rull. (1997). "Dimensions of inequality: Facts on the U.S. distribution of earnings, income, and wealth," Federal Reserve Bank of Minneapolis Quarterly Review. Spring: 3-21.

Guerin, K.J. (1997). Tracking student stop-out, transfer and graduation: An event history analysis of competing risks. Paper Presented at the 37th Annual Forum of the Association for Institutional Researchers, Lake Buena Vista, FL.

Han A., and Hausman, J. (1990). "Flexible parametric estimation of duration and competing risk models". Journal of Applied Econometrics, 5, 1-28.

Heckman, J. J., and Singer, B. (1982). "The identification problem in econometric models for duration data." In W. Hildenbrand (Ed.) Advances in Econometrics , New York: Cambridge University Press.

________ (1984). "A method for minimizing the impact of distributional assumptions in econometric models for duration data". Econometrica, 52, 271-320.

Jencks, C. , and Edlin, K. (1995). "Do poor women have a right to bear children?," The American Prospect. 20:43-52.

Kiefer, N. M. (1988). "Economic duration data and hazard functions." Journal of Economic Literature, Vol. XXVI: 646:679.

Lancaster, T. (1990). The Econometric Analysis of Transition Data. New York: Cambridge University Press.

Lederman, D. (1991). "College athletes graduate at higher rate than other students but mens' basketball players lag far behind, a survey finds," Chronicle of Higher Education. 37:1.

McCall, B. (1994). "Testing the proportional hazards assumption in the presence of unmeasured heterogeneity". Journal of Applied Econometrics, 9, 321-334.

Meyer, B. (1986). "Semi-parametric estimation of hazard models". M.I.T. working paper.

________ (1990). "Unemployment insurance and unemployment spells". Econometrica, 58,757-782.

Michael, R. T., and Tuma, N. B. (1985). "Entry into parenthood by young men and women: The influence of family background". Demography, 22, 515-544.

Murnane, R. J., Willett, J. B., and Tyler, J. H. (2000). "Who benefits from obtaining a GED? Evidence from High School and Beyond". The Review of Economics and Statistics, 82 (1): 23-37.

Murphy, K. , and Welch, F. (1993). "Inequality and relative wages," American Economic Review. 83: 104-109.

Murtaugh, P.A., Burns, L.D., and Schuster, J. (1999). "Predicting the Retention of University Students," Research in Higher Education 40(3): 355:371.

Pascarella, E.T. (1986). "A program for research and policy development on student persistence at the institutional level". Journal of College Student Personnel, 27: 100-107.

Pascarella, E.T., and Terenzini, P.T. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco, CA: Jossey-Bass.

Prentice, R. and Gloeckler, J. (1978). "Regression analysis of grouped survival data with application to breast cancer data". Biometrics, 34, 57-67.

Price, J.L. (1977). The study of turnover. Ames, Iowa: Iowa State University Press.

Ronco, S. L. (1994). Meandering ways: Studying student stopout with survival analysis. Paper Presented at the Annual Forum of the Association for Institutional Researchers, New Orleans, LA.

________ (1996). How enrollment ends: Analyzing the correlates of student graduation, transfer and dropout with a competing risks model. AIR Professional File, Number 61.

Rossi, P. H., Berk, R. A., and Lenihan, K. J. (1980). Money, work, and crime: Some experimental results. New York: Academic.

Singer, J.D. and Willett, J.B. (1991). "Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events". Psychological Bulletin, Vol. 110, Number 2: 268:290.

Sorensen, A. B. (1977). "Estimating rates from retrospective questions." In D. R. Heise (Ed.), Sociological Methodology. San Francisco: Jossey-Bass.

Spady, W. (1970). "Dropouts from higher education: An interdisciplinary review and synthesis". Interchange, Vol. 1: 64:85.

________ (1971). "Dropouts from higher education: Toward an empirical model", Interchange. Vol. 2: 38-62.

Tinto, V. (1975). "Dropout from higher education: A theoretical synthesis of recent research". Review of Educational Research, Vol. 45: 89-125.

________ (1982). "Limits of theory and practice in student attrition". Journal of Higher Education, Vol.53: 687-700.

________ (1987). Leaving college: Rethinking the causes and cures of student attrition. Chicago, IL: The University of Chicago Press.

________(1988). "Stages of student departure: Reflections on the longitudinal character of student leaving". Journal of Higher Education, Vol.59: 438-455.

________(1993). Leaving college: Rethinking the causes and cures of student attrition. Chicago, IL: The University of Chicago Press.

Trusell, J. and Richards, T. (1985). "Correcting unmeasured heterogeneity in hazard models using the Heckman-Singer approach." In N. B. Tuma (Ed.) Sociological Methodology 1985. San Francisco: Jossey-Bass.

Xiao, Beiling. (1997). "Using discrete-time survival analysis to examine time to master of science degree attainment." Paper Presented at the 37th Annual Forum of the Association for Institutional Researchers, Lake Buena Vista, FL.

Yamaguchi, K. (1991). Event history analysis. Applied Social Research Methods Series, Volume 28., Newbury Park, CA: Sage Publications.

Figure 1: "Baseline" Hazard Function
No UH Control

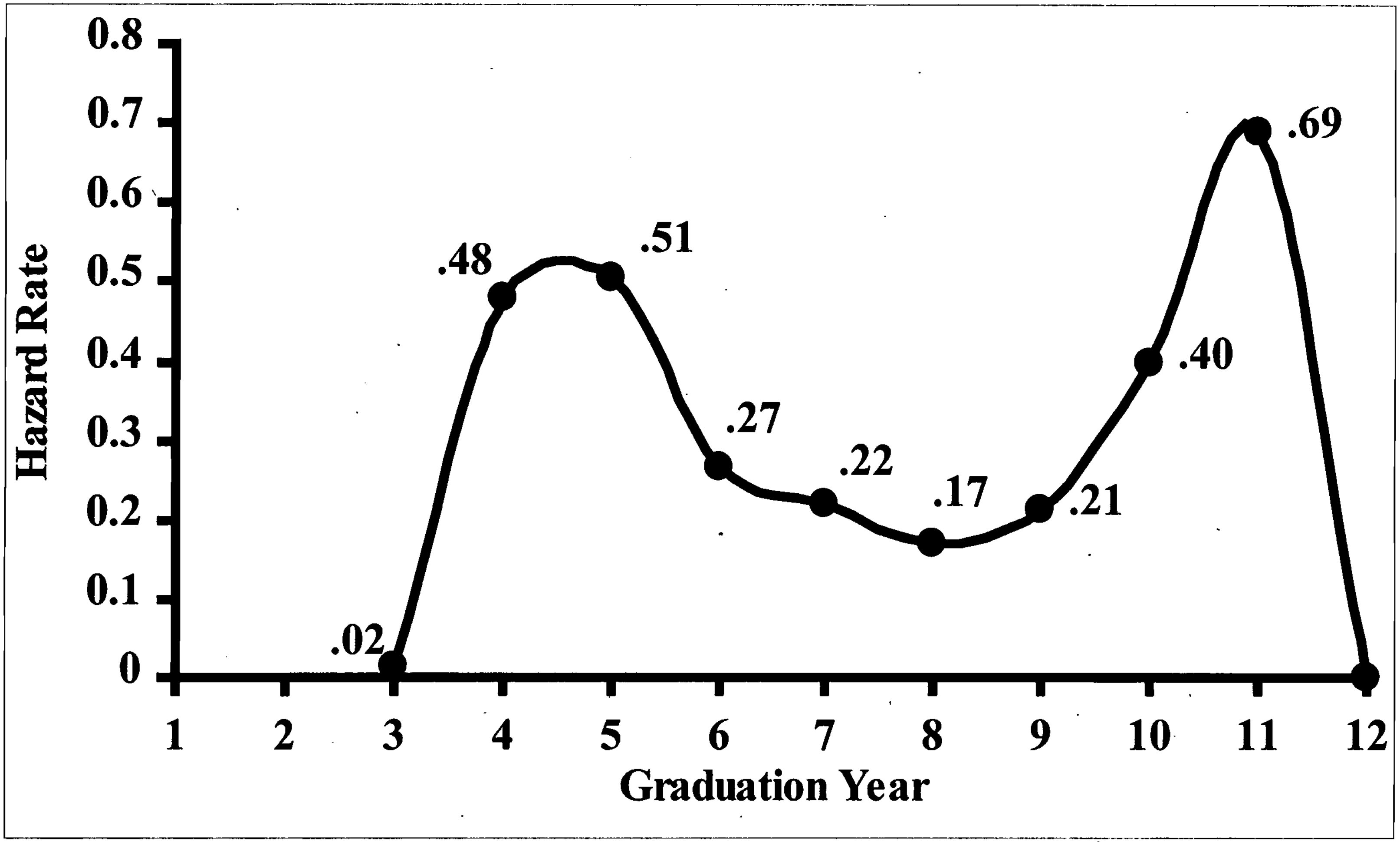


Table 2: Construction of Effective Sample

Sample Description	N	%	Description
All HS&B Sophomores	14,799	100%	-
Postsecondary Transcript / History Information Available	7,166	48%	At least one postsecondary education transcript received from the student and completeness of their postsecondary transcript record
Students With Postsecondary Coursework	7,081	48%	Students with no course records were deleted (N=85)
Of Students With Coursework…			
No Delay in Matriculation	5510	78%	Students who began postsecondary education in 1982
Delayed Matriculation	1571	22%	Students who began postsecondary education after 1982
Ever Attended a Four-Year Institution	4681	32%	Had enrolled in a four-year institution at some point durin the 12 year observation period

30

31

Table 3: Descriptive Statistics of the Sample
(N=4681)

Variable	Min	Max	Mean	Variance	Variable Description (Type of Variable)	NCES Variable Name
Academic Resources	1	5	4.05	1.13	Index of HS Academic Resources (Quintiles)	ACRES
Parent by 1986	0	1	0.04	0.04	Whether student was parent by 1986 (Dummy)	CHILD86M
SES Quintile	1	5	3.01	1.13	Socio-economic Status (Quintiles)	SESQ
Anticipations	1	5	4.38	0.81	Postsecondary Attendance Anticipations (1)	EDUCEXP
Minority Student	0	1	0.25	0.19	If Black, Hispanic, Native American (Dummy)	RACE
Male	0	1	0.47	0.25	If Male (Dummy)	PSEX
Loan	0	1	0.44	0.25	If Awarded Loan (Dummy)	*
Grant	0	1	0.56	0.25	If Awarded Grant (Dummy)	*
Work/Study	0	1	0.70	0.21	If Awarded Work/Study (Dummy)	*
After 1985	0	1	0.02	0.02	If Enrolled After 1985 (Dummy)	*
College GPA	0	4	2.54	0.79	GPA in College (Continuous)	*
Acad. Resources Components						
Academic Intensity/Quality	1	5	4.00	1.18	Intensity/Quality of HS Curriculum (Quintiles)	ACCURHSQ
High School Rank	1	5	3.86	1.41	High School Rank (Quintiles)	CLSRANKQ
Senior Year Test	1	5	3.96	1.25	Test from Senior Year in HS (Quintiles)	SRTESTQ2

[1] 5=Consistently Stated Expected a Bachelor's Degree
4=Expectations Raised to Bachelor's Over Time
3=Expectations Lowered from Bachelor's Over Time
2=Expectations Lowered to No Degree
1=No Degree Expectations

Variable names in italics are permitted to vary over time. The statistics associated with these variables are for a student's first-year of enrollment.

* These variables were created by the authors from a number of NCES variables and sources. The programming code (in SAS) is available on request.

Table 4: "Baseline" Time-Constant Coefficient Models with Various UH Controls

Variable	No U.H.		Gamma		Flexible	
Coefficient / (se)						
Academic Resources	0.3440	**	0.5125	**	0.4476	**
	(.025)		(.044)		(.032)	
Parent by 1986	-0.9764	**	-1.4006	**	-1.2747	**
	(.166)		(.196)		(.172)	
SES Quintile	0.0750	**	0.1027	**	0.0840	**
	(.016)		(.024)		(.021)	
Anticipations	0.1838	**	0.2544	**	0.2327	**
	(.03)		(.041)		(.035)	
Minority Student	-0.2440	**	-0.3632	**	-0.3186	**
	(.051)		(.075)		(.062)	
Male	-0.2257	**	-0.4090	**	-0.3549	**
	(.039)		(.064)		(.051)	
Likelihood (L)	-4831		-4809		-4807	

* p < .05
** p < .01

34

Table 5: "Financial Aid" Time-Constant Coefficient Models With Various UH Controls

Variable	No U.H.		Gamma		Flexible	
Coefficient / (se)						
Academic Resources	0.3337	**	0.4953	**	0.4241	**
	(.025)		(.044)		(.032)	
Parent by 1986	-1.0082	**	-1.4500	**	-1.3014	**
	(.167)		(.196)		(.172)	
SES Quintile	0.1017	**	0.1461	**	0.1199	**
	(.017)		(.026)		(.021)	
Anticipations	0.1682	**	0.2387	**	0.2161	**
	(.031)		(.042)		(.036)	
Minority Student	-0.2978	**	-0.4446	**	-0.3833	**
	(.052)		(.077)		(.064)	
Male	-0.2216	**	-0.4013	**	-0.3417	**
	(.039)		(.064)		(.051)	
Loan	0.1401	**	0.2328	**	0.1911	**
	(.042)		(.067)		(.055)	
Grant	0.2516	**	0.3565	**	0.3049	**
	(.044)		(.069)		(.057)	
Work/Study	-0.0991	**	-0.1735	**	-0.1294	**
	(.045)		(.068)		(.057)	
After 1985	0.2090		0.5952	**	0.4072	*
	(.213)		(.257)		(.224)	
Likelihood (L)	-4800		-4777		-4776	

* p < .05
** p < .01

Table 6: "GPA" Time-Constant Coefficient Models
With Various UH Controls

Variable	No U.H.		Gamma		Flexible	
Coefficient / (se)						
Academic Resources	0.2064	**	0.2967	**	0.2500	**
	(.025)		(.037)		(.031)	
Parent by 1986	-1.0262	**	-1.4426	**	-1.3183	**
	(.175)		(.197)		(.177)	
SES Quintile	0.0949	**	0.1265	**	0.1103	**
	(.016)		(.024)		(.021)	
Anticipations	0.2216	**	0.2928	**	0.2656	**
	(.03)		(.041)		(.036)	
Minority Student	-0.1037	*	-0.1820	**	-0.1596	**
	(.054)		(.074)		(.064)	
Male	-0.1335	**	-0.2599	**	-0.2273	**
	(.038)		(.059)		(.05)	
Loan	0.1937	**	0.2986	**	0.2730	**
	(.043)		(.065)		(.056)	
Grant	0.1443	**	0.1827	**	0.1565	**
	(.044)		(.065)		(.057)	
Work/Study	-0.0780	*	-0.1360	**	-0.1180	**
	(.045)		(.066)		(.058)	
After 1985	0.0677		0.2371		0.1370	
	(.221)		(.255)		(.23)	
College GPA	0.6556	**	0.9337	**	0.8410	**
	(.029)		(.057)		(.043)	
Likelihood (L)	-4582		-4550		-4544	

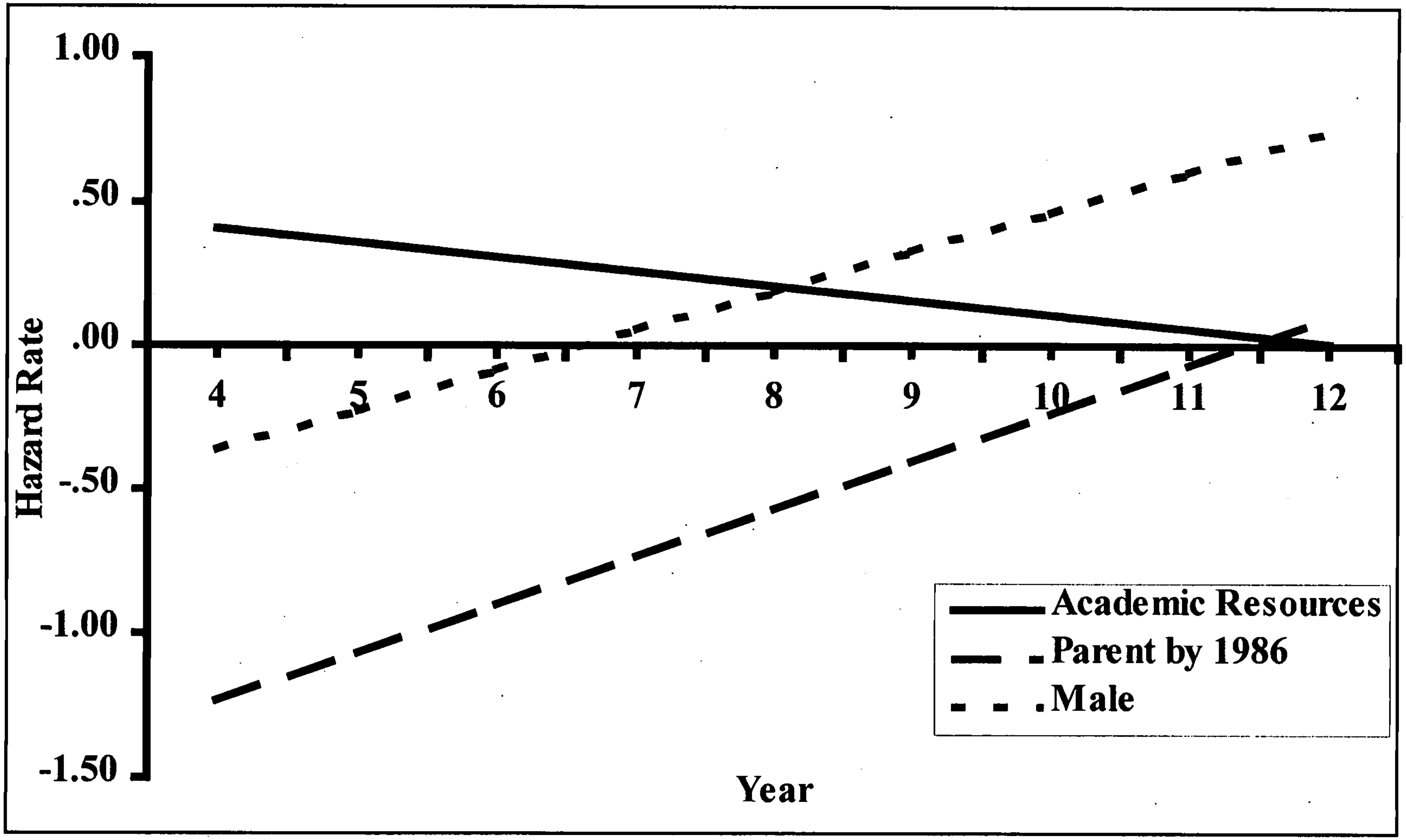
* p < .05
** p < .01

Table 7: "Components" Time-Constant Coefficient Models
With Various UH Controls

Variable	No U.H.		Gamma		Flexible	
Coefficient / (se)						
Academic Intensity/Quality	0.1849	**	0.2421	**	0.2096	**
	(.024)		(.035)		(.029)	
High School Rank	0.0701	**	0.1064	**	0.0977	**
	(.021)		(.029)		(.026)	
Senior Year Test	0.0257		0.0423		0.0264	
	(.024)		(.034)		(.03)	
Parent by 1986	-1.0082	**	-1.3989	**	-1.2985	**
	(.175)		(.196)		(.174)	
SES Quintile	0.0935	**	0.1246	**	0.1101	**
	(.017)		(.025)		(.021)	
Anticipations	0.2054	**	0.2699	**	0.2486	**
	(.031)		(.041)		(.036)	
Minority Student	-0.1190	**	-0.1917	**	-0.1678	**
	(.054)		(.074)		(.064)	
Male	-0.1404	**	-0.2556	**	-0.2181	**
	(.039)		(.059)		(.051)	
Loan	0.1929	**	0.2904	**	0.2671	**
	(.043)		(.064)		(.055)	
Grant	0.1457	**	0.1786	**	0.1508	**
	(.044)		(.065)		(.057)	
Work/Study	-0.0819	*	-0.1343	**	-0.1204	**
	(.045)		(.066)		(.057)	
After 1985	0.1128		0.2688		0.1704	
	(.225)		(.256)		(.231)	
College GPA	0.6486	**	0.9094	**	0.8264	**
	(.03)		(.057)		(.043)	
Likelihood (L)	-4566		-4536		-4531	

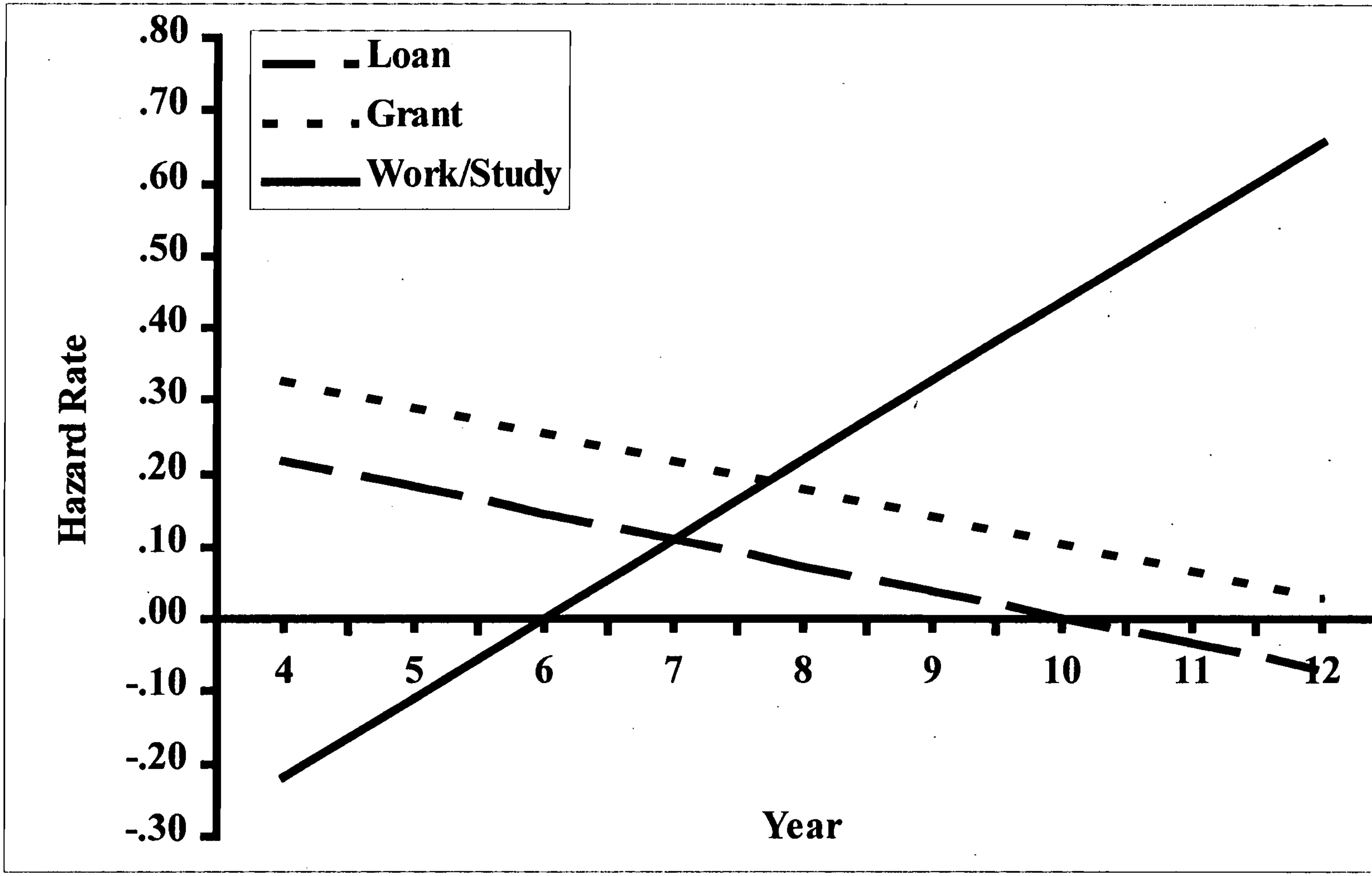
* p < .05
** p < .01

Figure 2: Hazard Rates for Selected Variables from the "Baseline" Time-Varying Coefficient Model



Fig

Figure 3: Hazard Rates for Selected Variables from the "Financial Aid " Time-Varying Coefficient Model



41 42

Figure 4: Comparing Academic Resources Effects

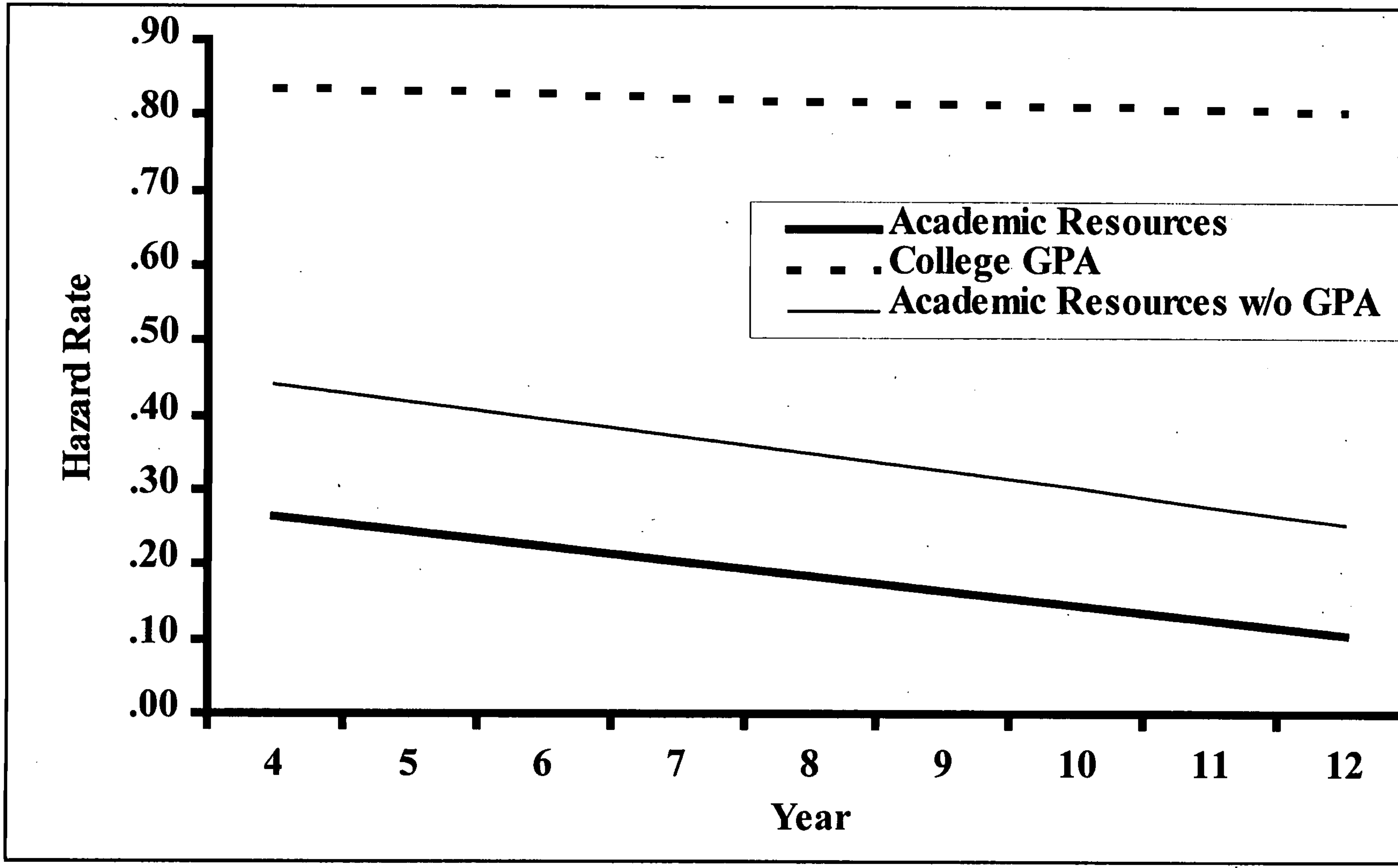
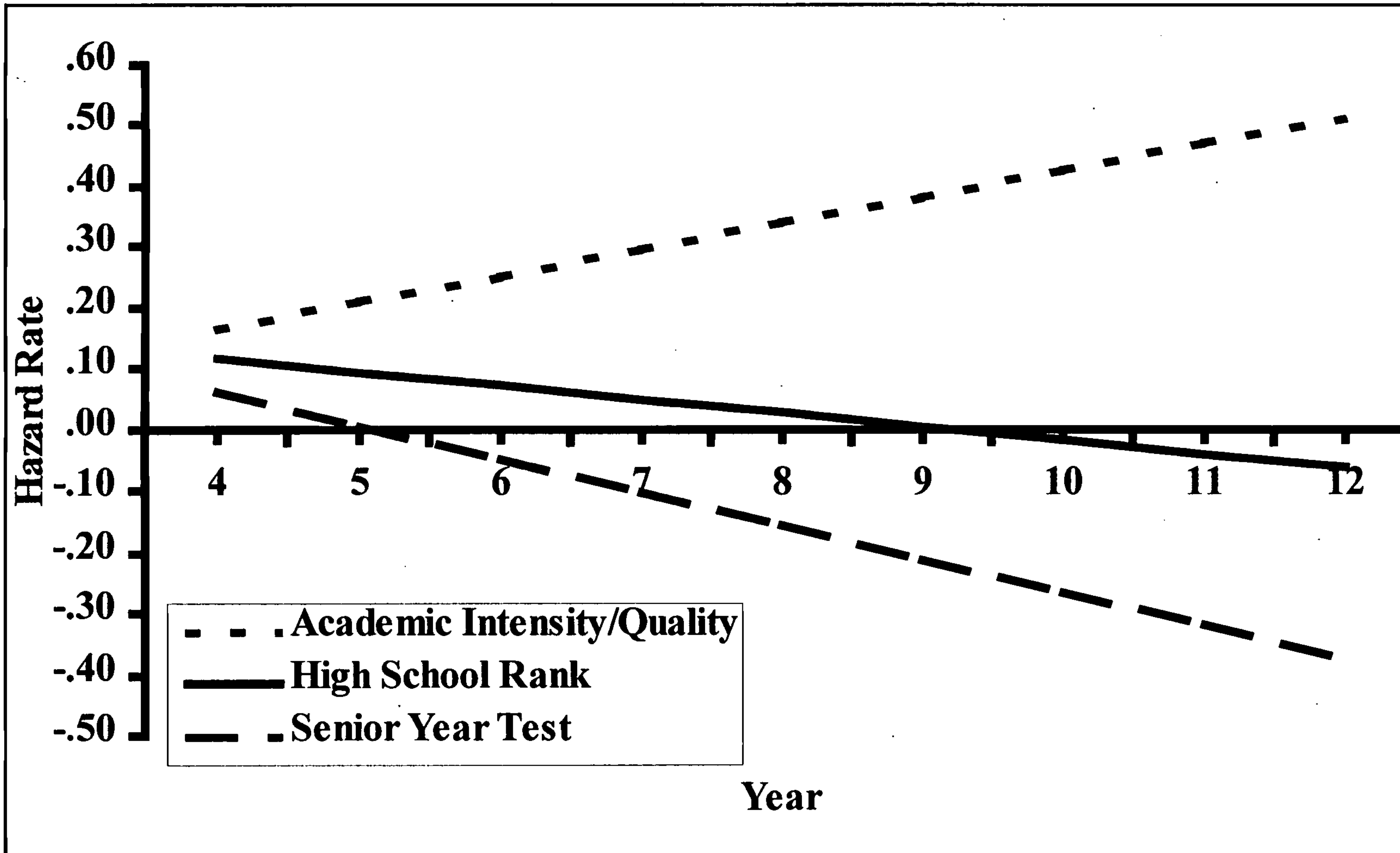


Figure 5: Hazard Rates for Selected Variables from the "Components" Time-Varying Coefficient Model



45

46

Table 8: "Baseline" Time-Varying Coefficient Model with Various UH Controls

	No UH		Gamma		Flexible	
Variable	Baseline Effect	Trend	Baseline Effect	Trend	Baseline Effect	Trend
Coefficient / (se)						
Academic Resources	0.5590 **	-0.0504 **	0.5556 **	-0.0325 *	0.4695 **	-0.0038
	(.064)	(.014)	(.069)	(.019)	(.076)	(.019)
Parent by 1986	-1.7280 **	0.1660 **	-1.6719 **	0.1156	-1.4800 **	0.0537
	(.369)	(.07)	(.383)	(.08)	(.389)	(.081)
SES Quintile	0.0603	0.0039	0.0551	0.0087	0.0447	0.0122
	(.046)	(.01)	(.05)	(.012)	(.052)	(.013)
Anticipations	0.3990 **	-0.0519 **	0.3800 **	-0.0398 **	0.3264 **	-0.0235
	(.076)	(.017)	(.082)	(.019)	(.08)	(.019)
Minority Student	-0.3986 **	0.0395	-0.3752 **	0.0203	-0.3094 **	-0.0032
	(.135)	(.03)	(.146)	(.035)	(.151)	(.037)
Male	-0.7761 **	0.1385 **	-0.8453 **	0.1404 **	-0.8717 **	0.1441 **
	(.111)	(.026)	(.123)	(.029)	(.124)	(.031)
Likelihood (L)	-4798		-4797		-4793	

* p < .05
** p < .01

Table 9: "Financial Aid" Time-Varying Coefficient Models with Various UH Controls

Variable	No UH Baseline Effect		No UH Trend		Gamma Baseline Effect		Gamma Trend		Flexible Baseline Effect		Flexible Trend	
Coefficient / (se)												
Academic Resources	0.5578	**	-0.0534	**	0.5557	**	-0.0385	**	0.5115	**	-0.0233	
	(.065)		(.014)		(.07)		(.018)		(.075)		(.018)	
Parent by 1986	-1.7947	**	0.1732	**	-1.7463	**	0.1271		-1.6422	**	0.0913	
	(.372)		(.07)		(.382)		(.078)		(.387)		(.079)	
SES Quintile	0.1176	**	-0.0043		0.1156	**	0.0008		0.1086	**	0.0036	
	(.047)		(.011)		(.051)		(.012)		(.053)		(.013)	
Anticipations	0.3809	**	-0.0504	**	0.3719	**	-0.0419	**	0.3492	**	-0.0346	*
	(.078)		(.017)		(.083)		(.019)		(.082)		(.019)	
Minority Student	-0.5254	**	0.0582	*	-0.5044	**	0.0384		-0.4564	**	0.0217	
	(.138)		(.031)		(.148)		(.035)		(.151)		(.036)	
Male	-0.7421	**	0.1302	**	-0.8044	**	0.1320	**	-0.8226	**	0.1336	**
	(.112)		(.027)		(.122)		(.029)		(.124)		(.03)	
Loan	0.3496	**	-0.0523	*	0.3496	**	-0.0435		0.3265	**	-0.0357	
	(.121)		(.029)		(.131)		(.031)		(.134)		(.033)	
Grant	0.5006	**	-0.0659	**	0.4833	**	-0.0505		0.4425	**	-0.0375	
	(.128)		(.03)		(.138)		(.033)		(.141)		(.035)	
Work/Study	-0.5123	**	0.1034	**	-0.5475	**	0.1078	**	-0.5466	**	0.1094	**
	(.135)		(.033)		(.146)		(.036)		(.147)		(.037)	
After 1985	5.4438	**	-1.6285	**	5.5332	**	-1.6279	**	5.4473	**	-1.5981	**
	(.896)		(.297)		(.908)		(.301)		(.754)		(.255)	
Likelihood (L)	-4749				-4748				-4746			

* p < .05
** p < .01

49 50

Table 10: "GPA" Time-Varying Coefficient Models with Various UH Controls

Variable	No UH Baseline Effect	No UH Trend	Gamma Baseline Effect	Gamma Trend	Flexible Baseline Effect	Flexible Trend
Coefficient / (se)						
Academic Resources	0.3573 **	-0.0350 **	0.3587 **	-0.0210	0.3235 **	-0.0198
	(.068)	(.015)	(.076)	(.017)	(.071)	(.016)
Parent by 1986	-1.8845 **	0.1866 **	-1.8589 **	0.1144	-1.7706 **	0.1073
	(.363)	(.071)	(.39)	(.083)	(.377)	(.079)
SES Quintile	0.0936 **	0.0000	0.0933 *	0.0062	0.0670	0.0109
	(.048)	(.011)	(.054)	(.013)	(.052)	(.013)
Anticipations	0.4414 **	-0.0530 **	0.4195 **	-0.0350 *	0.3780 **	-0.0289
	(.079)	(.017)	(.087)	(.02)	(.083)	(.019)
Minority Student	-0.2835 **	0.0449	-0.2835 *	0.0321	-0.2610 *	0.0289
	(.142)	(.032)	(.159)	(.037)	(.151)	(.037)
Male	-0.6447 **	0.1268 **	-0.7775 **	0.1442 **	-0.7353 **	0.1400 **
	(.113)	(.027)	(.131)	(.032)	(.124)	(.031)
Loan	0.4275 **	-0.0583 **	0.4481 **	-0.0446	0.3909 **	-0.0320
	(.124)	(.029)	(.142)	(.034)	(.136)	(.034)
Grant	0.2304 *	-0.0268	0.1655	0.0008	0.1308	0.0063
	(.13)	(.03)	(.148)	(.036)	(.143)	(.036)
Work/Study	-0.4752 **	0.0980 **	-0.5432 **	0.1088 **	-0.4986 **	0.1016 **
	(.139)	(.035)	(.157)	(.039)	(.149)	(.038)
After 1985	5.1675 **	-1.5771 **	5.3828 **	-1.6307 **	5.4920 **	-1.6855 **
	(.878)	(.296)	(.892)	(.304)	(.765)	(.264)
College GPA	0.9087 **	-0.0630 **	0.9412 **	-0.0157	0.8482 **	-0.0038
	(.095)	(.022)	(.11)	(.028)	(.107)	(.027)
Likelihood (L)	-4531		-4519		-4514	

* p < .05
** p < .01

51

52

Table 11: "Components" Time-Varying Coefficient Models with Various UH Controls

	No UH		Gamma		Flexible	
Variable	Baseline Effect	Trend	Baseline Effect	Trend	Baseline Effect	Trend
Coefficient / (se)						
Academic Intensity/Quality	0.1096	0.0198	0.0706	0.0402 **	0.0360	0.0436 **
	(.069)	(.015)	(.078)	(.018)	(.076)	(.018)
High School Rank	0.2233 **	-0.0367 **	0.2164 **	-0.0292 *	0.1830 **	-0.0225
	(.059)	(.013)	(.066)	(.015)	(.063)	(.015)
Senior Year Test	0.2000 **	-0.0447 **	0.2344 **	-0.0518 **	0.2254 **	-0.0543 **
	(.066)	(.014)	(.075)	(.017)	(.072)	(.017)
Parent by 1986	-1.8537 **	0.1849 **	-1.8004 **	0.1101	-1.6733 **	0.0924
	(.372)	(.074)	(.396)	(.085)	(.383)	(.081)
SES Quintile	0.1084 **	-0.0044	0.1052 *	0.0023	0.0736	0.0084
	(.049)	(.011)	(.055)	(.013)	(.054)	(.013)
Anticipations	0.4099 **	-0.0486 **	0.3952 **	-0.0338	0.3653 **	-0.0303
	(.08)	(.018)	(.089)	(.021)	(.086)	(.02)
Minority Student	-0.2242	0.0253	-0.2092	0.0089	-0.2236	0.0181
	(.144)	(.033)	(.16)	(.037)	(.154)	(.038)
Male	-0.6652 **	0.1318 **	-0.8017 **	0.1527 **	-0.7907 **	0.1608 **
	(.117)	(.028)	(.134)	(.033)	(.128)	(.033)
Loan	0.4376 **	-0.0609 **	0.4565 **	-0.0480	0.3734 **	-0.0290
	(.123)	(.029)	(.142)	(.034)	(.138)	(.035)
Grant	0.2155	-0.0235	0.1477	0.0038	0.1139	0.0086
	(.131)	(.031)	(.15)	(.036)	(.146)	(.037)
Work/Study	-0.4817 **	0.0984 **	-0.5383 **	0.1068 **	-0.5023 **	0.1024 **
	(.138)	(.034)	(.155)	(.039)	(.147)	(.038)
After 1985	5.2484 **	-1.5915 **	5.4466 **	-1.6395 **	5.2890 **	-1.6089 **
	(.886)	(.299)	(.903)	(.308)	(.782)	(.269)
College GPA	0.8455 **	-0.0497 **	0.8747 **	-0.0046	0.7289 **	0.0223
	(.097)	(.022)	(.111)	(.029)	(.105)	(.027)
Likelihood (L)	-4509		-4498		-4492	

* p < .05
** p < .01

53

54

NOTICE

Reproduction Basis

☑ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").