An Analysis of Change of Major Behavior of Cal Poly Students

A Senior Project

Presented to

The Faculty of the Statistics Department

California Polytechnic State University, San Luis Obispo

In Partial Fulfillment

Of the Requirements for the Degree

Bachelor of Science

By Logan Lossing

Under the Guidance of Dr. Jeff Sklar

Table of Contents

Introduction

There has been much work done in educational research that attempts to model and explain various student behaviors from a multitude of perspectives. Much of the research done on university students focuses on trying to understand factors associated with college degree attainment. This is of no surprise as it is a fairly common belief that a college education is important for the individual and is inextricably linked to the health of the national economy. Further, graduation rates are a common measurement of the success of a university.

Educational researchers in the past have investigated various background and personality characteristics as well as institutional factors that are significantly associated with college graduation rates, but little work has been done to assess the impact that changing majors has on degree attainment. In particular, there has been very little work estimating the relationship between switching majors and time until degree attainment, especially using survival analysis methods. The main goal of this study was to present a reasonable way with which to accomplish such an analysis. Though the end goal is to be able to answer questions about the relationship between switching major and college degree attainment on a national level, performing the relevant analyses on the data available from Cal Poly San Luis Obispo was used to assess the effectiveness of such an approach and provide a good starting point for future research.

Before looking into the Cal Poly data, it is advantageous to first take a look at what results other researchers have found regarding factors related to graduation from college as well as any analyses of major switching behaviors.

Literature Review

Chen and Weko (2009) focus their attention on developing a profile of undergraduate students who enter and subsequently graduate from STEM majors. They analyzed longitudinal national level data and conducted simple t-tests without adjusting for multiple comparisons but they did obtain some interesting results. They did not detect any gender differences in STEM degree attainment, though they noted that more men entered STEM majors. They found that Asian/Pacific Islander students were much more likely to enter a STEM major than White, African American, and Hispanic/Latino students, between which no measurable differences were discovered. However, both White and Asian/Pacific Islander students were more likely than African American or Hispanic/Latino students to graduate from a STEM major. Younger students and those considered dependent (usually under age 24) had a higher STEM major entrance rates and STEM graduation rates. Students having at least one parent with a four year degree and students with considerable academic preparation also were shown to have higher STEM graduation rates. Overall, students entering college in STEM majors were determined to be more likely to graduate from college. An interesting aspect of their analysis was that they broke all STEM majors into four groups: Mathematics, the natural sciences, engineering, and computer science. They conducted analyses that sought to determine if there were different graduating behaviors for the various STEM types and did indeed find several important differences.

Shaw and Barbuti (2010) looked at factors that are associated with a student remaining in their original major through the third year of college with an explicit focus on STEM majors. Using national level data collected from 39 four year institutions, they found several interesting relationships between various background and demographic factors and major persistence through the third year as well as academic performance measures while in college. Statistical significance was assessed using Cohen's d, a method that standardizes mean differences between groups and uses this as a measure of effect size. Though their study only looks at persistence in matriculating major through the third year, they do find some interesting results. They found small differences in matriculating major persistence through the third year in gender, parental income, and first generation status, as well as with ethnicity. They found that women were less likely to persist than men, underrepresented ethnicities were less likely to persist than White and Asian/Pacific Islander students, and first generation students were less likely than students whose parents had some college to persist. They also note that students that persisted in their matriculating major through the third year also tended to have higher high school and university grade point averages (GPA) but this varied by majors.

Allen and Robbins (2007) attempted to predict persistence to the third year in matriculating major using academic preparation measures, first year academic performance, and the student's vocational interests. Using logistic regression methods they determined that first year GPA was associated with persistence but that neither high school GPA nor ACT composite score was significant in the presence of first year GPA. They also note that their finding of the importance for first year college GPA is consistent with results obtained by other researchers.

Chizmar (2010) published his analysis of the role that gender plays in the persistence of economics majors in a paper published by *The Journal of Economic Education* in 2000. Though focused on specifically economics he failed to detect a gender difference but was able to conclude that students whose grades in economics classes were lower than their grades in other classes were more likely to switch as well as students with only a small amount of coursework completed in the major. More interesting than his results was his method of analysis. He made use of the discrete time hazard model as formulated by researcher Judith Singer and John Willett of Harvard. This is notable as it is a method that we use to analyze our data and will be discussed at length later.

Methods

Work on the project began in Summer 2011. The first big task undertaken was to clean the data, create and/or modify variables, and recode them with meaningful labels. All of the statistical analysis and manipulation for the project was done using the R statistical software package.

Research Questions

Four main questions were developed to study the relationship between switching majors, graduating from college, and time until graduation.

(1) Can we identify important demographic and academic characteristics associated with the chance that a student will change his/her major?

(2) Can we determine, after controlling for background information and academic performance (both during high school and while at Cal Poly), what effect does switching major have on the risk of graduation?

(3) Can we determine if students that switch majors have different graduation rates than those that do not change majors?

(4) Can we determine how the timing of the major switch affects time until graduation?

The Data, Variables, and Sample Used

The data were acquired from Institutional Planning and Analysis at Cal Poly San Luis Obispo and provides information about the incoming freshman class of 2005. This data set includes academic information about their studies at Cal Poly for 6 years beginning Fall Quarter 2005 through Spring Quarter 2011, as well as various background information concerning ethnicity, gender, parental education, California residency and high school academic performance. A comprehensive list of the variables used, their levels, and explanations are located in Table 1.

The type of information provided to us by the variables fell into three classes. The first class of variables capture demographic background information. These include *Gender*, *Ethinicity, First Generation Status*, *Geography*, and *Pell Grant*. The second class of variables contain academic background information that perhaps partly capture the preparedness of the student for the rigors of Cal Poly coursework. These include *High School GPA*, *SAT Score*, and *Remedial Work*. The final class of variables include information about each student's Cal Poly career: the units and term GPA they had for every quarter from Fall 2005 to Summer 2011, what major/college they belonged to, and major switching information.

Table 1: Variables Used in Analysis

Variable	Levels	Additional Information
Demographic Background Characteristics:		
Gender	Male Female	
Ethnicity	Hispanic/Latino African American Native American White Haw/ Pacific Islander Asian American International Unknown	International students and students of unknown ethnicity were not used in inferential models.
FirstGeneration	Student is First Gen Student is not First Gen	A Student is considered to be a First Generation Student if the maximum educational level of either parent was high school or below.
Geography	CA Resident non-CA Resident	A student is a CA Resident if they were a resident at the time they applied to Cal Poly.
Pell Grant	Received Pell Grant No Pell Grant	A Pell Grant is a form of Financial Aid provided by the government.
Academic Background Characteristics:		
High School GPA		The final high school GPA, on a 4.0 scale.
SAT Score		The original data contained information about each student's SAT and ACT scores. Not all students took both entrance exams. For those students that only took the ACT, an SAT equivalence score was computed. The SAT score used was the sum of the SAT reading and SAT mathematics scores. computed using a concordance table.

Variable	Levels	Additional Information
Remed	Did some remedial work No remedial work required	When entering the university students take mathematics and english placement exams to see if they are ready for college level work.
Academic Performance at Cal Poly:		
Units		Number of units attempted for each quarter from 2005-2011.
GPA		Term GPA for each quarter from 2005-2011.
College		The college the student's major belonged to upon entering Cal Poly in 2005.
	CAFES	*College of Architecture and Engineering Science*
	CAED	*College of Education*
	CENG	*College of Engineering*
	CLA	*College of Liberal Arts*
	OCOB	*Orfalea College of Business*
	CSM	*College of Science and Mathematics*
STEM Matric	STEM non-STEM	All majors at Cal Poly were classified as being a STEM or a non-STEM major based on course work required. This variable refers to the classification of the student's matriculating major.
FirstSwitch.Vary		This is a variable we created to keep track of type and timing of a student's first major switch. This variable is coded as "Persist" for all quarters before the student's first major change, is either a 1 or a 2 for all quarters after the first major switch. A "1" indicates that the major change was to a STEM major, "2" indicates the change was to a non-STEM major.
First Switch	Persist Switch to STEM Switch to non-STEM	This variable tracks if a major change occurred. The coding only applies to the first major change.
Number of Major Changes		This variable represents the number of major changes made by the student from 2005-2011.

Variable	Levels	Additional Information
Year Switch	Persist Switch Year 1 Switch Year 2 Switch Year 3 Switch Year 4 Switch Year 5	This variable indicates the academic year in which the first major switch was made.

Descriptive Statistics

We are specifically interested in looking at the association between switching majors and overall college graduation rates. To begin, we take a look at some descriptive statistics that give us a feel for what the overall switching and graduation behaviors are.

In Table 2, we see that of the original 3425 students, 2596 managed to graduate within six years and 829 had yet to complete their degrees. We see that of those that managed to graduate in the six years, about 24.4% switched majors compared to 12.1% for those that have not yet graduated. Looking more specifically at the type of major switch, we see that of the 21.4% of students that switched majors at some point, 9.9% switched to a STEM major and 11.5% switched to a non-STEM major. In addition, we see that the relative percentage contribution of those that switched to STEM majors and those that non-STEM majors are higher to the Graduation group and lower for the Drop Out/Still Enrolled group. This is encouraging as it seems to point to higher graduation rates for those that switch majors. Also note, we see that the overwhelming majority of students that did change majors did so only once, and that it is fairly rare that a student change majors two or more times, only about 1.3% of all students did so.

Table 2: Major Change Information Broken Down by Graduation Status for 2005 Cal Poly Freshman Class

Variable	Levels	$n_{DropOut/StillEnrolled}$	$\%_{DropOut/StillEnrolled}$	n_{Grad}	$\%_{Grad}$	n_{all}	$\%_{all}$
EverChange	Persist	729	87.9	1964	75.7	2693	78.6
	Switched	100	12.1	632	24.4	732	21.4
	all	829	100.0	2596	100.0	3425	100.0
First Major Switch	Persist	729	87.9	1964	75.7	2693	78.6
	Sw STEM	61	7.4	278	10.7	339	9.9
	Sw Non-STEM	39	4.7	354	13.6	393	11.5
	all	829	100.0	2596	100.0	3425	100.0
NumberMajorChanges	0	729	87.9	1964	75.7	2693	78.6
	1	90	10.9	597	23.0	687	20.1
	2	8	1.0	34	1.3	42	1.2
	3	1	0.1	1	0.0	2	0.1
	4	1	0.1	0	0.0	1	0.0
	all	829	100.0	2596	100.0	3425	100.0

Now, let's look at the graduation and major switching rates tabulated by specific categorical variables, this information is available in Table 3.

Table 3: Background Characteristics Broken Down by Graduation Status for 2005 Cal Poly Freshman Class

Variable	Levels	$n_{Graduated}$	$\%_{Graduated}$	$n_{DropOut/StillEnrolled}$	$\%_{DropOut/StillEnrolled}$	n_{all}	$\%_{all}$
College	CAED	234	9.0	72	8.7	306	8.9
	CAFES	483	18.6	171	20.6	654	19.1
	CENG	520	20.0	268	32.3	788	23.0
	CLA	449	17.3	108	13.0	557	16.3
	CSM	294	11.3	100	12.1	394	11.5
	OCOB	616	23.7	110	13.3	726	21.2
	all	2596	100.0	829	100.0	3425	100.0
Ethnicity	AfAm	22	0.8	11	1.3	33	1.0
	AsAm	292	11.2	87	10.5	379	11.1
	Haw	8	0.3	7	0.8	15	0.4
	Hisp	214	8.2	105	12.7	319	9.3
	Inter	12	0.5	5	0.6	17	0.5
	NaAm	24	0.9	7	0.8	31	0.9
	Unknown	228	8.8	85	10.2	313	9.1
	White	1796	69.2	522	63.0	2318	67.7
	all	2596	100.0	829	100.0	3425	100.0
Gender	Male	1261	48.6	524	63.2	1785	52.1
	Female	1335	51.4	305	36.8	1640	47.9
	all	2596	100.0	829	100.0	3425	100.0
FirstGeneration	FirstGen	2397	92.3	721	87.0	3118	91.0
	NotFirstGen	199	7.7	108	13.0	307	9.0
	all	2596	100.0	829	100.0	3425	100.0
Remed	AtLeastSomeREMED	282	10.9	156	18.8	438	12.8
	NoREMED	2314	89.1	673	81.2	2987	87.2
	all	2596	100.0	829	100.0	3425	100.0

For the *College* variable, we note that most colleges have roughly equal contributions (in terms of percentage) to both the DropOut/Still Enrolled and Graduation groups. However, we see that 32.3% of all Cal Poly students that did not graduate in the six years from 2005-2011 belong to the CENG, whereas only 20% of all students that graduated were from that college. There is a similar but opposite behavior for the students of the OCOB, 23.7% of all graduating students were a member of the business college, but of all those that had not graduated by Spring 2011, only 13.3% belonged to OCOB. We also see a less pronounced difference in the *Gender* variable. Males have a higher contribution to the total number of DropOut/Still Enrolled group (63.2% Male, 36.8% Female) and conversely a higher percentage of females graduated in the six year period (51.4% to 48.6%).

Turning our attention to the *Ethnicity*, *FirstGeneration*, *Geography*, and *Remed* variables, we see that the the relative percentage contribution of each level to each category for both the Graduation and DropOut/Still Enrolled groups are approximately equal. This seems to indicate that a strong relationship between these background characteristics and graduation rates may not exist.

Now, let's compare the switching major behaviors for each of these groups. This information is provided in Table 4. Major switching rates and switching behaviors vary tremendously between the colleges. We see that although only about 21.2% of all students belong to the OCOB, they are responsible for 24.9% of all students that persist in their original major. Similarly, 23% of Cal Poly students entered in the CENG but they are responsible for 52.8% of all major switches that are to STEM majors. Finally, though only 16.3% and 19.1% of all students belonged to CLA and CAFES respectively, they both are responsible for about 30% of all major switches to non-STEM majors. Turning our attention to the *Gender* variable, we see that though almost equal percentages of males and females persist in their majors, the switching behaviors are very different, about 60% of all major switching by males is to a STEM major whereas about 60% of all major switching by females is to a non-STEM major.

With respect to the other variables, we see very comparable major switching rates and behaviors through each level.

Variable	Levels	$n_{Persist}$	$\%_{Persist}$	n_{SwSTEM}	$\%_{SwSTEM}$	$n_{SwNon-STEM}$	$\%_{SwNon-STEM}$	n_{all}	$\%_{all}$
College	CAED	254	9.4	19	5.6	33	8.4	306	8.9
	CAFES	491	18.2	49	14.4	114	29.0	654	19.1
	CENG	562	20.9	179	52.8	47	12.0	788	23.0
	CLA	422	15.7	26	7.7	109	27.7	557	16.3
	CSM	294	10.9	47	13.9	53	13.5	394	11.5
	OCOB	670	24.9	19	5.6	37	9.4	726	21.2
	all	2693	100.0	339	100.0	393	100.0	3425	100.0
Ethnicity	AfAm	24	0.9	3	0.9	6	1.5	33	1.0
	AsAm	295	10.9	50	14.8	34	8.6	379	11.1
	Haw	12	0.4	2	0.6	1	0.2	15	0.4
	Hisp	259	9.6	30	8.8	30	7.6	319	9.3
	Inter	17	0.6	0	0.0	0	0.0	17	0.5
	NaAm	25	0.9	2	0.6	4	1.0	31	0.9
	Unknown	246	9.1	33	9.7	34	8.6	313	9.1
	White	1815	67.4	219	64.6	284	72.3	2318	67.7
	all	2693	100.0	339	100.0	393	100.0	3425	100.0
Gender	Male	1423	52.8	212	62.5	150	38.2	1785	52.1
	Female	1270	47.2	127	37.5	243	61.8	1640	47.9
	all	2693	100.0	339	100.0	393	100.0	3425	100.0
FirstGeneration	FirstGen	2448	90.9	313	92.3	357	90.8	3118	91.0
	NotFirstGen	245	9.1	26	7.7	36	9.2	307	9.0
	all	2693	100.0	339	100.0	393	100.0	3425	100.0
Remed	AtLeastSomeREMED	344	12.8	42	12.4	52	13.2	438	12.8
	NoREMED	2349	87.2	297	87.6	341	86.8	2987	87.2
	all	2693	100.0	339	100.0	393	100.0	3425	100.0

Table 4: Background Characteristics Broken Down by Major Switching Behavior for 2005 Cal Poly Freshman Class

Now, let's quickly look at the breakdown in the quantitative variables *High School GPA* and *SAT Score* by graduation and major switching behaviors. This information is displayed in Tables 5 & 6.

Variable	Levels	$\bar{x}$	$\tilde{x}$	s
HSGPA	Graduated	3.8	3.8	0.4
	DropOut/Still Enrolled	3.6	3.6	0.4
	all	3.7	3.8	0.4
SAT	Graduated	1204.6	1210.0	119.6
	DropOut/Still Enrolled	1202.4	1210.0	137.3
	all	1204.0	1210.0	124.1

Table 5: Academic Background by Graduation Status

Variable	Levels	$\bar{x}$	$\tilde{x}$	s
HSGPA	Persist	3.7	3.8	0.4
	Sw STEM	3.8	3.8	0.4
	Sw Non-STEM	3.7	3.7	0.4
	all	3.7	3.8	0.4
SAT	Persist	1204.3	1210.0	124.8
	Sw STEM	1230.4	1240.0	123.5
	Sw Non-STEM	1179.6	1180.0	115.1
	all	1204.0	1210.0	124.1

Table 6: Academic Background by Major Switch

By examining the tables we see that there are is not any substantial differences in High school GPA or SAT scores when comparing switching behaviors, however, when looking at graduation there does seem to be a noticeable difference between those that graduated and those that had not after 6 years when looking at the *High School GPA* variable. Students that ultimately graduated from Cal Poly within the six year period have a slightly higher mean and a trimmed mean High School GPAs.

Analysis/Results

Multinomial Logistic Regression Model

Our first research goal was to identify background characteristics and academic performance measures that are associated with various switching behaviors. To aid us in this analysis we began by creating a variable that we named *FirstSwitch* that has three nominal levels. The levels corresponded to different possible switching behaviors that we were interested in studying. The first level represent those that persisted (never switched majors), the second level represent those whose first major switch was into a STEM major, and the final level represent those whose first major switch was to a non-STEM major. To successfully accomplish the goal of identifying characteristics that are associated with various major switching behaviors we need to find a way to relate this nominal categorical variable *FirstSwitch* to the categorical and continuous predictors. An effective statistical tool for analyzing relationships between a categorical nominal dependent variable and categorical or continuous independent variables is multinomial logistic regression, and we made use of it to accomplish the first part of our analysis.

Multinomial regression allows us to assess which independent variables are useful in determining the odds that a student will switch majors rather than persist but it also allows us to see how the relationships between various predictors change when estimating the odds the major switch would be to a STEM versus a non-STEM major. Multinomial regression output provides us two sets of coefficients, one set for each comparison that needs to be made. One set to compare those that switch to STEM majors to those that persist and another set to compare those that switch to a non-STEM major to those that persist. For this analysis, it makes the most sense to make those that persist in their original major the baseline group because it is the most natural to interpret. The multinomial regression model assumes that the natural logarithm of the odds of a student being in a major switching group to persisting is a linear function of the independent variables. In this case we have:

$$\ln\left(\frac{\Pr(Switch\ to\ STEM)}{\Pr(Persist)}\right) = \boldsymbol{B}_{\mathbf{1}}^{\boldsymbol{T}} * \mathbf{X} \ \text{ and } \ \ln\left(\frac{\Pr(Switch\ to\ non{-}STEM)}{\Pr(Persist)}\right) = \boldsymbol{B}_{\mathbf{2}}^{\boldsymbol{T}} * \mathbf{X},$$

where **X** is the vector of predictors and the $\mathbf{B_i}$ is the vector of parameters coefficients

Multinomial Regression is sensitive to having combinations of predictor variables with few observations. For this reason we had to make a few adjustments, both to our variables and our dataset. First, we removed *College* and used *STEM Matric* to indicate whether the student's initial major was STEM or non-STEM. The method of determining which majors were considered to be a STEM major was guided by the classifications made by Chen (2009) as well as a close inspection of the coursework required for each major using the Cal Poly Course Catalog. A table listing all Cal Poly majors and our corresponding classifications is located in Table 13 in the appendix on page 30. Secondly, we had to cut out groups of students that had small populations. This included those of Hawaiian/Pacific Islander, African American, and Native American ethnicity. Students with missing values in any predictor were also removed from the analysis.

The mlogit package was used to obtain maximum likelihood estimates of the model parameters. Results are available in Tables 7 below. To assess model fit, a global likelihood ratio test is carried out.

For this dataset, the resulting test statistic had a very small p-value and thus we can conclude that the model is useful in differentiating major switching behavior. Further, drop in deviance tests can be conducted for each predictor in the model to test whether or not the independent variables have an overall relationship to the response variable, *FirstSwitch*. The drop in deviance test statistic is computed by fitting two models: one model with all the relevant parameters and a second model that excludes the predictors being tested, thus the first model has g more parameters being estimated. The null hypothesis is that the parameter estimates corresponding to predictors included in the full model but not the reduced model are zero and the alternative hypothesis is that at least one of the parameter estimates is not zero. The test statistic is:

$$D = 2(\text{LogLikelihood}_{\text{Full Model}} - \text{LogLikelihood}_{\text{Reduced Model}})$$

Under the null hypothesis, D is approximately chi-square distributed with g degrees of freedom. Once it has been determined that a predictor has an overall relationship to the response then we can look at the significance of each predictor in being able to distinguish between those that switch majors to STEM and those that persist, as well as distinguishing those that switch to non-STEM majors and those that persist.

Multinomial Regression models carry with them the assumption of independence of irrelevant alternatives. Essentially this assumption asserts that a person's preference between two alternatives is not affected by the presence of another alternative. In the context of this project for example, the amount a student prefers to switch to a STEM major rather than to persist in their matriculating major is not dependent on the option to switch to a non-STEM major. The Hausman-McFadden test was carried out to test this assumption. The resulting p-value of the test was very near one, we have no reason to believe that this assumption has been violated in this case.

The Discrete Time Hazard Model

Of the four main research questions previously laid out, there are still three that we have yet to answer. Instead of trying to identify who might be likely to change majors, the remaining three questions were created to guide us in exploring the relationship between switching majors and the risk of graduation from Cal Poly. To be able to answer research questions (2) – (4) we need to be able to take the following circumstances into consideration.

Since the data covers the incoming 2005 Cal Poly freshman class only over a six year period of time we have no way of knowing if students that have not graduated by summer quarter 2011 will eventually do so. This introduces the problem of censoring. Ignoring censored observations has very dire analytical consequences and a branch of statistics has been developed to deal with censoring, survival analysis. Most survival analysis techniques, however, are designed to be used on continuous time to event random variables. In this study the event of interest, graduating from Cal Poly, can only happen at the end of a quarter. This makes the time to event random variable for this analysis discrete.

In addition, the data we have been provided contains both time varying and time invariant predictors. A time invariant predictor is one that remains constant over time, such as *Gender*. A time varying predictor is one that changes value over time. The variable *Units* is an example of a time varying predictor as the unit load a student takes can vary from quarter to quarter, though note, the values of all predictors, even time varying predictors, are frozen during each quarter. We want to be able to make use of both types of predictors in our analysis.

Finally, we note that graduation is a non-repeatable event. Once a student graduates, they are ineligible to graduate at a later date. This makes time to graduation inherently conditional in the sense that you are only eligible to graduate in quarter j given that you have not already graduated in quarters 1,2,…,j-1.

A model introduced by the famous statistician Dr. David Cox in the 1970's and elucidated by many researchers over the years, most notably Dr. Judith Singer and Dr. John Willett of Harvard, can handle all of these problems simultaneously. This model is known as a Discrete-Time Hazard Model (DTHM). Hazard, a common quantity in survival analysis, is the "backbone" of the DTHM model. The benefit to using the hazard function is that it allows us to assess the risk of graduation in every quarter. In this case, we define hazard to be the conditional probability that a student graduates in some quarter j given that the student did not graduate in the previous j-1 quarters. (Note, in traditional survival analysis with a continuous time to event random variable hazard is a rate but in a DTHM hazard is a probability.) If we define random variable T to indicate the time period j when a randomly selected student graduates, then we can express hazard as:

$$h_j = \Pr(T = j | T \geq j)$$

We are interested in identifying predictors that help to determine whether different types of students have different hazard functions. To do this we must include the various predictors into our definition of hazard. The predictor adjusted hazard can be defined as the conditional probability that student i, i=1,2,…,N (N = number of student in sample), graduates in quarter j, j=1,2,…J (J = maximum number of quarters a student can be in the dataset, 24 quarters), given values for each of P predictor values $z_{ij} = (z_{1ij},\ldots, z_{Pij})$

$$h_{ij} = \Pr(T_i = j | T_i \geq j, Z_{1ij} = z_{1ij}, \ldots, Z_{Pij} = z_{Pij})$$

The most general form of the DTHM assumes that logistic transformation of hazard is a linear function of the P predictors and an intercept (represented by the α's with use of time dummies D) where α_j is the intercept for quarter j. We can write this formally as:

$$\ln\left(\frac{h_{ij}}{1-h_{ij}}\right) = (\alpha_1 D_{1ij} + \cdots + \alpha_{24} D_{24ij}) + (\beta_1 Z_{1ij} + \cdots + \beta_p Z_{Pij})$$

Where $D_{hij} = 1$, if h = j and zero otherwise.

By solving for h_{ij} we obtain:

$$h_{ij} = \frac{1}{1 + e^{-[(\alpha_1 D_{1ij} + \cdots + \alpha_{24} D_{24ij}) + (\beta_1 Z_{1ij} + \cdots + \beta_p Z_{Pij})]}}$$

To obtain parameter estimates we turn to the method maximum likelihood. As usual in survival analysis, the likelihood function is the product of the probabilities of observing the sample data. If we introduce a vector called the event indicator, y_{ij} where for student i, $y_{ij} = 1$ if the student graduates in time period j and is zero otherwise, and let j_i represent the total number of quarters that have passed since Fall 2005 in which student I had yet to graduate Cal Poly, then it can be shown that the likelihood function is:

$$L = \prod_{i=1}^{n} \prod_{j=1}^{j_i} h_{ij}^{y_{ij}} (1 - h_{ij})^{(1-y_{ij})}$$

Here we see that for each time period, a student either graduates or does not. If they do graduate, then at time j_i then they contribute h_{ij} to the likelihood function, otherwise they contribute $(1\text{-}h_{ij})$. This likelihood function is identical to a sequence of independent Bernoulli trials in parameters h_{ij}, where the number of trials is $K = j_1+j_2+\ldots+j_n$, the sum across all students, of the quarters each student is at Cal Poly without graduating. This equivalence allows for easy maximum likelihood estimation of the model parameters using standard logistic regression routines. The only caveat is that the data set needs to be transformed into what is known as "person period" form. A more detailed explanation of this and an example is located in Figure 3 of the appendix on page 29.

After parameter estimates have been obtained we need to assess model fit. Traditionally, the first model fit in a DTHM is the baseline hazard, the model that results from only using time as a predictor. If additional predictors are then added to the model we can assess if their inclusion into the model helps to fit the data better. In logistic regression we can accomplish this using a drop in deviance test, as described earlier.

In survival analysis, it is usually of interest to obtain an estimate of the survival function. To "survive" in this context means to "not graduate." Once hazard estimates are obtained, the corresponding estimated survival probabilities at time j can be easily calculated as:

$$\hat{S}_j = \prod_{k=1}^{j} (1 - \hat{h}_k)$$

The mean time to graduation in can be easily estimated by: $\widehat{E(T)} = \sum_{i=1}^{24} \widehat{S_i}$

For this part of the analysis we worked with an initial sample size of 46,412 person period records for 3012 students. Of the original 3425 students, we excluded students that declared themselves to be international or unknown in the *Ethnicity* variable, and students that had missing values in any of the predictors.

In order to have a more parsimonious model, a quartic polynomial model in time was fit. As is traditional, the first model fit was a model that only included time, baseline hazard. After estimating the baseline hazard we introduced the demographic background characteristics into the model. A drop in deviance test showed that the demographic and academic background characteristics have helped provide a better fit to the data. Next, information about Cal Poly academic performance and behavior, were added to the model and again a drop in deviance test identified these predictors as helping to improve model fit. Finally, interactions between *FirstSwitch.Vary* and *Ethnicity*, *Year Switch*, *Gender*, and *College* were added as well as interactions between time and GPA and time and Units. Again, these were found to improve model fit. Results of the test are located in Table 9. Now that we have a useful model we can take a look at the research questions and see what the model says.

Table 9: Discrete Time Hazard Model Results

	Model A	**Model B**	**Model C**	**Model D**
Model df	5	16	22	67
AIC	11341	11079	10438	9937.2
LL	-5665.4	-5523.6	-5196.8	-4901.6
-2(Δ LL)	-	283.55	653.73	590.29
df	-	11	6	45
P-Value Comparing	-			
A to B		<.001		
B to C			<.001	
C to D				<.001

Model A: Only time as a predictor
Model B: Time and Background (Academic and Demographic)
Model C: Time, Background (Academic and Demographic), and Academic Performance at Cal Poly
Model D: Time, Background (Academic and Demographic), Academic Performance at Cal Poly, and relevant interactions.

Discussion

Interpretation of the Multinomial Results

The first research question was designed to assess background characteristics associated with various major switching behaviors. The predictors that are useful both overall and to distinguish those switching to STEM majors and those that persist are *STEM Matric* and *Fall05 GPA*. Students that enter Cal Poly in a STEM major are more likely to switch to another stem major than to persist in their original majors. Further, the higher a student's *Fall 2005 GPA* the more likely that student is to switch to a STEM major rather than persist. The *Geography* variable is just barely not significant at the $\alpha = .05$ significance level but it suggests that California residents are slightly more likely to switch to a STEM major rather than persist.

The predictors that are useful both overall and to distinguish those switching to non-STEM majors and those that persist are *STEM Matric, Gender, High School GPA, SAT score, Fall05 GPA*, and the interaction term *STEM Matric*Fall05 GPA*. Those that enter as a STEM major are more likely to switch to a non-STEM major than to persist. Males are less likely to switch to a non-STEM major rather than persist. The higher a student's High School GPA the less likely they are to switch to a non-STEM major rather than persist, however, the higher a student's Fall 2005 GPA the more likely that student is to switch to a non-STEM major rather than persist, however, this effect is stronger for those entering Cal Poly in STEM majors when compared to those that entered in non-STEM majors.

Table 7: Drop in Deviance Tests of Significance for Predictors Included in the Multinomial Logistic Regression

Predictor	Log Likelihood	Drop in Deviance χ^2	DF	P-Value
STEM Matric	-1900.8	138.17	6	<.001
Gender	-1842.1	20.90	2	<.001
Ethnicity	-1835.9	8.40	4	0.08
First Generation	-1831.9	0.56	2	0.76
High School GPA	-1838.7	14.02	4	0.007
SAT Score	-1836.2	9.030	2	0.011
California Residency	-1834.6	5.96	2	0.051
Remedial Work	-1832.7	2.14	2	0.34
Fall 05' Units	-1831.9	0.47	2	0.79
Fall 05' GPA	-1852.6	20.90	4	<.001
Pell Grant Recipient	-1831.9	0.43	2	0.81
STEM:High School GPA	-1832	0.58	2	0.75
STEM:Fall 05 GPA	-1838.4	13.42	2	<.001
Full Model	-1831.7	238.42		<.001

Table 8: Multinomial Model Results

Predictor	Estimate	Std. Error	Odds Ratio	P-Value
Compare Persist to Switch STEM				
1:(intercept)	-5.62	1.63	0.00	<.001
1:STEM Matric	3.82	1.62	45.72	0.018
1:Male	0.18	0.14	1.20	0.19
1:Hispanic/Latino	-0.32	0.26	0.73	0.22
1:White	-0.33	0.18	0.72	0.06
1:First Generation	-0.12	0.28	0.88	0.66
1:High School GPA	0.087	0.43	1.09	0.84
1:SAT Score	0.000033	0.00066	1.00	0.95
1:Out Of State	0.40	0.19	1.49	0.04
1:No Remedial Work	-0.077	0.23	0.93	0.74
1:Fall05 Units	0.027	0.04	1.03	0.50
1:Fall05 GPA	0.70	0.25	2.02	0.0055
1:Received Pell Grant	0.1447	0.22	1.16	0.51
1:STEM Matric*High School GPA	-0.26	0.47	0.77	0.58
1:STEM Matric*Fall05 GPA	-0.44	0.27	0.64	0.11
Compare Persist to Switch non-STEM				
2:(intercept)	-0.038	1.11	0.96	0.97
2:STEM Matric	2.93	1.18	18.80	0.013
2:Male	-0.53	0.13	0.59	<.001
2:Hispanic/Latino	-0.57	0.27	0.94	0.83
2:White	0.29	0.19	1.34	0.14
2:First Generation	-0.183	0.28	0.83	0.52
2:High School GPA	-0.56	0.26	0.57	0.03
2:SAT Score	-0.0019	0.00063	1.00	0.003
2:Out Of State	-0.27	0.24	0.76	0.27
2:No Remedial Work	0.28	0.21	1.33	0.17
2:Fall05 Units	-0.0022	0.036	0.99	0.95
2:Fall05 GPA	0.73	0.15	2.08	<.001
2:Received Pell Grant	0.018	0.22	1.02	0.93
2:STEM Matric*High School GPA	-0.20	0.35	0.82	0.56
2:STEM Matric*Fall05 GPA	-0.66	0.20	0.52	<.001

Interpretation of DTHM Results

The second and third main research questions sought to determine if switching major is associated with the risk of graduating. Predictors in the model associated with major switching behavior are statistically significant and these include the interactions between *FirstSwitch.vary* and *YearSwitch*, *Gender*, and *College*. Thus we have evidence that the association between changing major and the risk of graduation is complicated and is dependent on the student's matriculating college, gender, the year in which the switch was made, and whether the first switch was made to a STEM major or a non-STEM major. Though this is a partially satisfying answer, it turns out that we can make some more insightful conclusions. If we were to look through all the majors in each college at Cal Poly, we could roughly separate the colleges into "Science" and "non-Science" colleges. The "science" colleges would be CAFES, CSM, CAED, and CENG leaving the "non-science" colleges CLA and OCOB. For the "science" colleges it turns out that if the student is going to change their major in the first three years[1], the hazard of graduation generally increases with a major change, the increase in hazard is always highest when switching to non-STEM majors. For the "non-STEM" colleges, persisting in the original major has the highest hazard of graduation for these colleges, followed by major switches to non-STEM majors.

To obtain a visual representation of this we will introduce a "typical" student profile and allow other predictors to vary. For this purpose we will fix the various background predictors at the following levels unless otherwise stated:

- *Gender*: male
- *Geography*: California resident
- *Pell Grant Status*: does not receive Pell grants
- *Ethnicity*: White
- *FirstGeneration*: not a first generation student
- *High School GPA*: 3.75(median value)
- *SAT Score*:1200 (median value)
- *Units*: 12 per quarter
- *GPA*: 3.0
- *Switch Yr:* 3(Median Switch Year)

Allowing both *FirstSwich.vary* and *College* to vary and holding all the other predictors fixed to the values described above, we can obtain a nice picture of the hazard estimates plotted by quarter.

Taking a look at Figure 1a, it is clear to see that switching majors greatly increases the estimated hazard of graduation for students that matriculated to "science" colleges. For CENG, this jump in hazard of graduation is dramatic when comparing those that persist to those that switched to non-STEM in the third year. We see that for OCOB and CLA the reverse behavior holds, switching to a STEM major in the third year approximately halves the hazard of graduation. This fits well with observations made in the descriptive statistics section. These patterns hold for switches made in years one and two, as well.

[1] Most majors at Cal Poly are designed to be completed in four years. It seems logical that switching majors in the fourth or fifth years would not result in higher hazard graduation.

Table 10: DTHM Results

Variable	Coef (Logit Scale)	Exp(Coef)	Individual Significance at α = .05
Intercept	-22.523	<.001	Yes
Time	-4.841	0.008	
Time Squared	0.867	2.38	
Time Cubed	-0.046	0.955	
Time Fourth	0.001	1.001	
First Generation Student	-0.207	0.813	No
Not a California Resident	-0.147	0.864	No
Receive Pell Grant	-0.235	0.79	Yes
Male	-0.418	0.658	Yes
Asian-American	-0.08	0.923	No
Hawaiian/Pacific Islander	-0.44	0.644	
Hispanic/Latino	0.048	1.049	
Native American	0.41	1.506	
White	-0.029	0.972	
HSGPA	0.568	1.765	Yes
SAT Combined Score	<.001	>.999	No
No Remedial Work Required	0.107	1.113	No
Number of Major Changes	-0.668	0.512	Yes
First Switch to STEM Major	7.558	1915.767	Yes
First Switch to non-STEM Major	9.296	10897.181	
Units Attempted	0.07	1.072	Yes
GPA	1.396	4.04	Yes
CAFES	0.897	2.453	
CENG	0.311	1.365	
CLA	1.547	4.697	Yes
CSM	0.912	2.49	
OCOB	1.853	6.377	
T:GpaVector	-0.042	0.959	Yes
T:Units	-0.009	0.991	Yes

CAFES: Switch to STEM	-0.033	0.968	Yes
CENG:Switch to STEM	-0.54	0.583	
CLA:Switch to STEM	-1.13	0.323	
CSM:Switch to STEM	0.048	1.049	
OCOB: Switch to STEM	-1.515	0.22	
CAFES: Switch to non-STEM	0.108	1.114	
CENG:Switch to non-STEM	0.387	1.473	
CLA:Switch to non-STEM	-0.607	0.545	
CSM:Switch to non-STEM	-0.106	0.899	
OCOB: Switch to non-STEM	-0.928	0.395	
Asian-American:Switch to STEM	1.639	5.149	No
Asian-American:Switch to non-STEM	-0.791	0.453	
Hawaiian/Pac Islander: Switch to STEM	2.132	8.432	
Hawaiian/Pac Islander: Switch to non-STEM	1.184	3.267	
Hispanic/Latino: Switch to STEM	0.726	2.066	
Hispanic/Latino: Switch to non-STEM	-0.638	0.528	
Native American: Switch to STEM	0.604	1.83	
Native American: Switch to non-STEM	-1.14	0.32	
White: Switch to STEM	1.298	3.663	
White: Switch to non-STEM	-0.758	0.468	
Male: Switch to STEM	0.241	1.272	Yes
Male: Switch to non-STEM	0.101	1.107	
Persist:Switch Yr 2	0.769	2.158	Yes
Switch to STEM: Switch Yr 2	-0.291	0.748	
Switch to non-STEM: Switch Year 2	0.411	1.509	
Persist:Switch Yr 3	0.563	1.756	
Switch to STEM: Switch Yr 3	-0.623	0.536	
Switch to non-STEM: Switch Year 3	0.045	1.046	
Persist:Switch Yr 4	-4.469	0.011	
Switch to STEM: Switch Yr 4	-0.917	0.4	
Switch to non-STEM: Switch Year 4	-0.594	0.552	
Persist:Switch Yr 5	-7.276	0.001	
Switch to STEM: Switch Yr 5	-1.261	0.283	
Switch to non-STEM: Switch Year 5	-0.559	0.572	
Persist:Switch Yr 6	-8.694	<0.001	
Switch to STEM: Switch Yr 6	-0.361	0.697	
Switch to non-STEM: Switch Year 6	-20.46	<0.001	
Persist:No Switch	7.711	2232.502	

Also, for CENG we notice that if the major switch was made to a STEM major the hazard of graduation actually drops substantially. This actually makes sense for a few reasons. First, engineering majors tend to take students longer to complete, five years is a very typical. In addition, engineering students have a tendency to switch to other majors within the engineering college. So, since all engineering majors fall under the STEM umbrella then it seems logical that we would see switches to STEM majors resulting in lower hazard graduation.

Taking a look a Figure 1b, we see this same phenomenon displayed a little differently. We see that for the "non-science" colleges, persisting in ones matriculating major has between two and three times higher hazard graduation. Another interesting thing that we can see in Figure 1b is that there are three grouping of colleges that have similar hazard estimates in all three cases: OCOB and CLA, CENG and CAED, and CSM and CAFES.

Another perspective that was to be addressed by the last research question was to investigate how the timing of the first major switch affects the risk of graduation. A visual representation of this is available in Figure 2a, as well as a table of estimated mean survival times in Table 11.

Again, there are two distinct behaviors that are noticeable, one for the "science" and one for the "non-science" colleges. Switching in the first three years is associated with higher risk of graduation for those matriculating into "science" colleges. Further, switching to a non-STEM major has the highest risk of graduation if done in the second year where as for those switching to STEM majors the highest risk of graduation is obtained if done in the first year. It is true that having a higher risk of graduating is equivalent to having smaller estimated mean time to graduation and this is demonstrated in Table 11.

The behavior for the "non-science" colleges previously mentioned holds as well. Students who persist tend to graduate sooner than those that switch, regardless of STEM classification. An interesting thing to note however is that the optimal year for switching by type of major change is the same for the "science" and "non-science" colleges, hazard of graduation is highest when switching to STEM majors in the first year, and highest when switching to non-STEM in the second year.

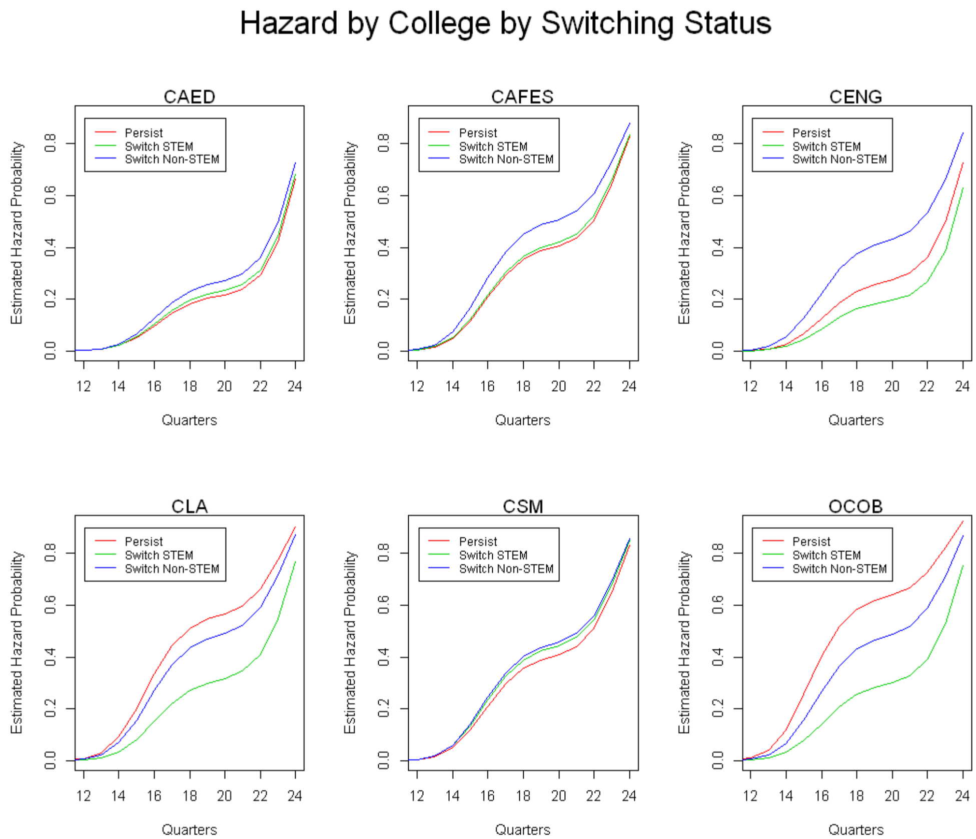

Figure 1a: Comparison of major switching behaviors by matriculating college.

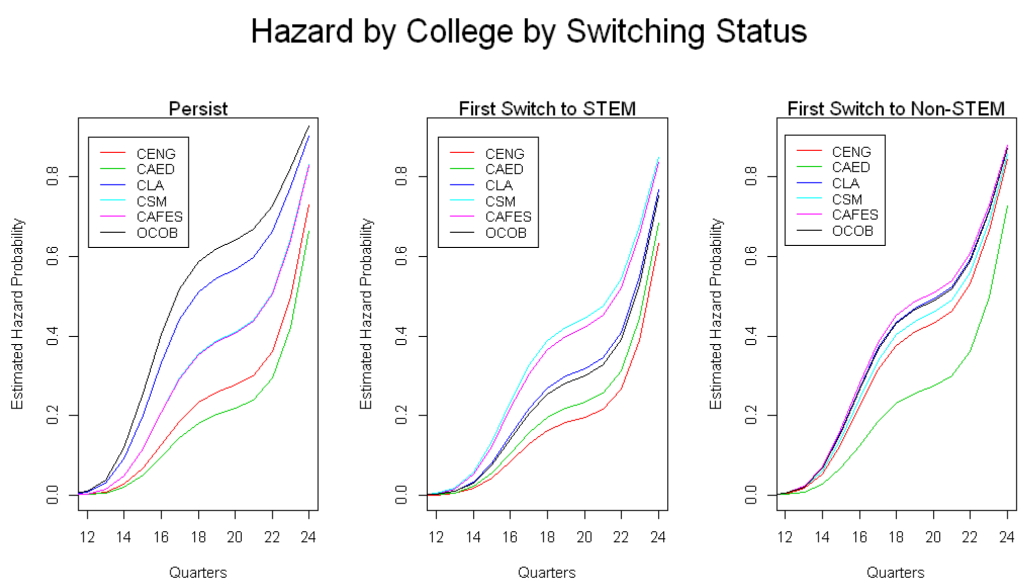

Figure 1b: Comparison of matriculating college by major switching behavior

(SW Yr 1)	Persist	Switch STEM	Switch non-STEM
CENG	17.75	17.34	16.42
CAED	18.45	16.88	17.86
CLA	15.46	16.11	16.01
CSM	16.51	15.28	16.23
CAFES	16.53	15.41	15.90
OCOB	15.06	16.24	16.03
(SW Yr 2)			
CENG	17.75	18.00	15.74
CAED	18.45	17.48	16.97
CLA	15.46	16.63	15.38
CSM	16.51	15.71	15.57
CAFES	16.53	15.86	15.29
OCOB	15.06	16.78	15.39
(SW Yr 3)			
CENG	17.75	18.78	16.34
CAED	18.45	18.23	17.76
CLA	15.46	17.30	15.93
CSM	16.51	16.25	16.15
CAFES	16.53	16.42	15.83
OCOB	15.06	17.46	15.96
(SW Yr 4)			
CENG	17.75	19.48	17.63
CAED	18.45	18.94	19.26
CLA	15.46	17.95	17.10
CSM	16.51	16.79	17.38
CAFES	16.53	16.98	16.97
OCOB	15.06	18.13	17.13

Table 11: Estimated Mean Time to Graduation (In Quarters)

Though the main research questions have been more or less answered, it is time to take a look at a few other interesting findings that our model helps to reveal. The first thing of interest is that women have a higher hazard of graduation than men and thus lower estimated time until graduation, regardless of major switch behavior. This can be seen below in Table 12. Since differing behavior between "science" and "non-science" colleges has been so consistent, a representative college of each type was selected for comparison. The significant interaction between *Gender* and *FirstSwitch.vary* is also interesting in that shows that though women tend to graduate faster than men, switching majors regardless of STEM classification actually raises their expected mean time to graduation. In contrast, a major switch of either type is associated with a decrease in estimated mean time to graduation for males, but switching to a STEM major is associated with a larger decrease than switching to a non-STEM major.

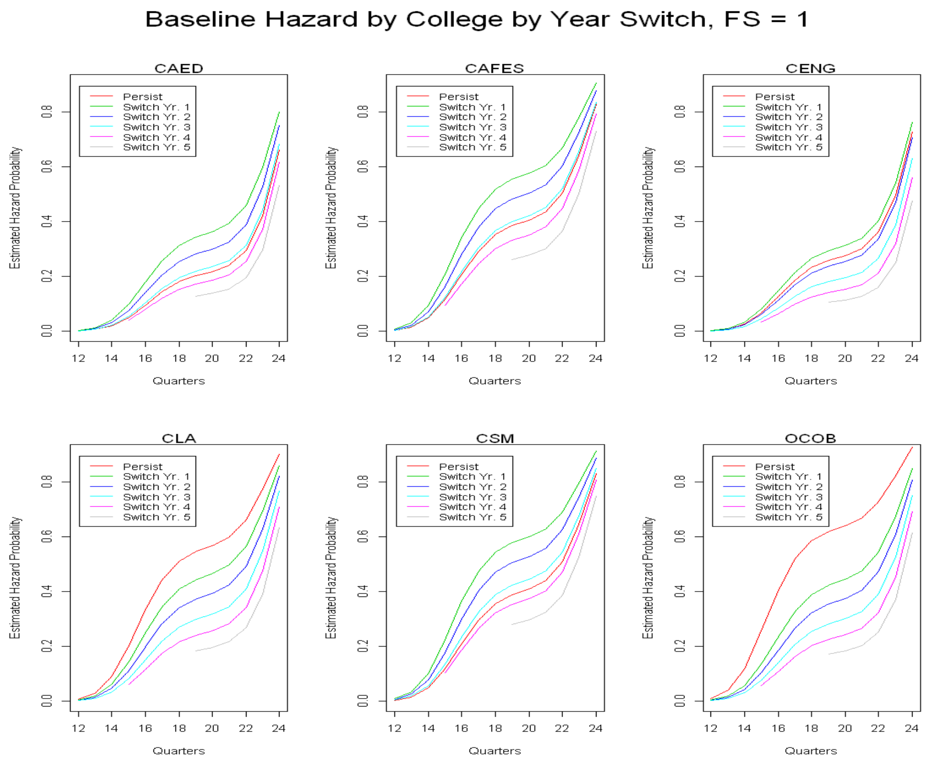

Figure 2a: Comparison of how timing of switching to STEM majors relates to hazard estimates

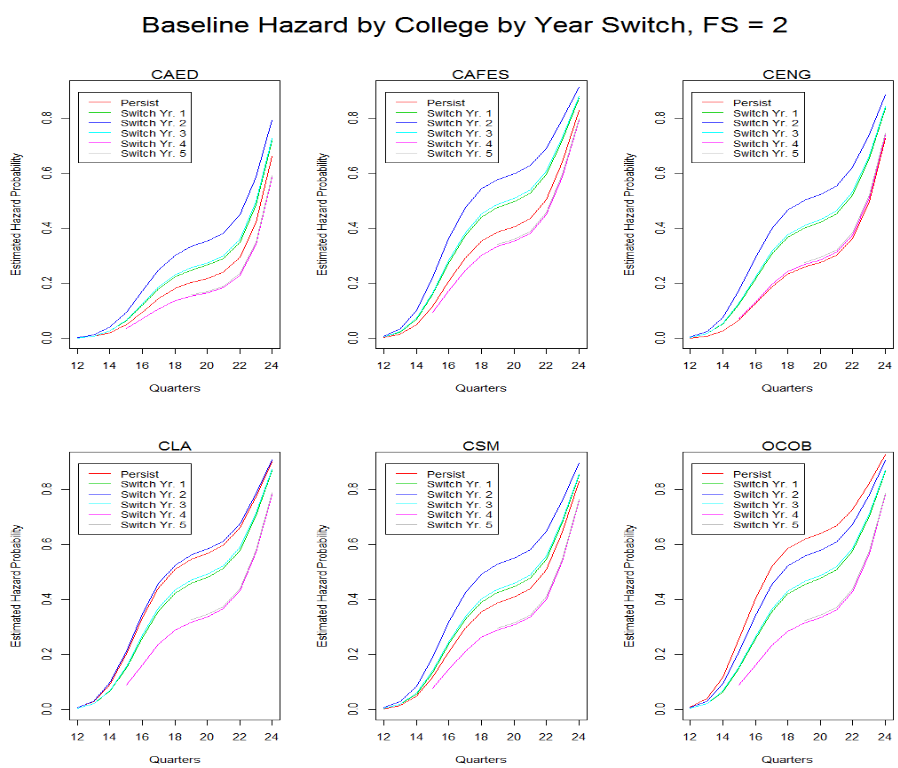

Figure 2b : Comparison of how timing of switching to non-STEM majors relates to hazard estimates

Gender/College Type	Persist	Switch STEM	Switch non-STEM
Male/SCI	17.75	19.48	17.63
Female/SCI	16.86	19.07	16.95
Male/Non-SCI	15.46	17.95	17.10
Female/Non-SCI	14.91	17.55	16.49

Table 12: Estimated mean time to graduation, in quarters. SCI = CENG, Non-SCI = CLA

Another interesting thing to look at is what specific predictors were significantly associated with hazard of graduation individually, in the presence of the other predictors. Though they were included as control variables, it is still important to compare to the findings of other researchers. Interactions between Time and both *GPA* and *Units* were statistically significant. Higher GPA and higher unit loads are associated with higher hazard graduation initially but the interaction tells us that the association decreases with time. Also, students receiving Pell Grants were estimated at having a decrease in hazard of graduation. *High School GPA* was significantly associated with hazard of graduation but neither *SAT Score* nor *Remed* were. Higher High School GPAs are associated with much higher hazard of graduation. *Ethinicity*, *FirstGeneration*, and *Geography* were all found to be not significantly associated with hazard of graduation in the presence of the other predictors.

Limitations of Analyses

This study was conducted using data from a moderately selective, technically focused, public university in a relatively ethnically homogenous area. The results found may not generalize to other types of universities, however, the methodology used is flexible, allowing for similar analyses to be carried out with different datasets easily. The variables used are such that they should be readily available at most universities, allowing for investigation of how the association between major switching behaviors and the hazard of graduation might differ by characteristics of the university and student population.

There are some complications with the data that would have been nice to be able to circumvent. One big issue results from data not available to us. The dataset provided to us only had records of a student's primary major. Information about double majors and minors were not available. It seems reasonable that the inclusion of this information would allow a clearer picture of the relationship between hazard of graduation and the predictors used in the model. Unfortunately, there is no way of estimating the number of students that double major or minor making estimation of the size of the problem difficult. Another possible bit of information that could have been useful in our analysis regards student participation in study abroad programs or in Cal Poly athletics programs. Other types of analyses could have taken into consideration the effect that summer school attendance has on the risk of graduation.

Conclusion

The explicit purpose of this analysis was to illuminate the relationship between switching majors and degree attainment, with a particular emphasis on STEM majors. Data from the 2005 Cal Poly freshman cohort was used to assess this relationship while controlling for various background characteristics. Four research questions were developed in order to guide the analysis. These questions try to help identify what characteristics are associated with various major switching behaviors as well as assess how switching majors is related to college degree attainment while controlling for background characteristics, and how this relationship changes with the timing of the major change.

Extensive literature searches and review as well as a careful examination of descriptive statistics were performed to guide our expectations. The descriptive statistics showed what seemed to be different graduation rates for those switching majors and those that persisted in their original major. There were several other background characteristics that seemed to be related to degree attainment, most notably *College* and *Gender*.

In order to identify which background characteristics are associated with various major switching behaviors, a multinomial logistic regression model was created. The model is useful in determining how various background characteristics help to differentiate those that switch to STEM majors rather than persist as well as those that switch to non-STEM majors rather than persist. It was determined that students matriculating into a STEM major are more likely to switch majors than students matriculating to non-STEM majors, however, they are more likely to switch to another STEM major than to a non-STEM major. Further, it was shown that higher *Fall05 GPA* is associated with a higher chance of switching majors regardless of the STEM classification of the matriculating major. It was found that males are less likely than women to switch to a non-STEM major rather than persist. Higher *High School GPA* was found to lower the odds that a student would switch to a non-STEM major.

To assess the relationship between various major switching behaviors and degree attainment, a discrete time hazard model was fit. The DTHM allows for assessment of the hazard of graduation in every quarter while controlling for various background and academic performance measures. A useful model was constructed and results obtained. The relationship between hazard of graduation and major switching behavior was found to be complicated and depended on the student's matriculating college, gender, and the year that the switch occurred. This relationship between hazard of graduation and major switching behavior can be broken down into two basic varieties that depend on whether or not the student's matriculating college is a "science" or a "non-science" college. If a switch is made in the first three years, students matriculating to "science" colleges seem to benefit from switching majors in that it generally increases their hazard of graduation whereas persisting in one's original major is best for students matriculating to non-science majors. Another interesting finding was that if a student switches major, regardless of matriculating to a "science" college or not, hazard of graduation is increased the most if the switch takes place in the first year for switches to STEM majors and the second year for switches to non-STEM majors. Additionally, it was determined that women have higher hazard graduation regardless of major switching behavior than men but switching majors usually increases a women's time to graduation while decreasing it for men.

Appendix

Figure 3: Example Showing Conversion of Data from Person Form to Person Period Form

In Person Form

Student	Grad?	Censor?	Total Quarters	M/F	GPA Q1	GPA Q2	GPA Q3	GPA Q4	GPA Q5	GPA Q6	GPA Q7	GPA Q8
1	Yes	No	6	M	3.45	3.6	3.5	3.6	2.9	3.6	-	-
2	No	Yes	2	F	3.5	2.9	-	-	-	-	-	-
3	No	Yes	8	M	3.3	2.75	3.56	.3.61	3.15	2.85	3.2	3.5

In Person Period Form

Quarter	Student	Y	M/F	D1	D2	D3	D4	D5	D6	D7	D8	GPA Vector
1	1	0	M	1	0	0	0	0	0	0	0	3.45
2	1	0	M	0	1	0	0	0	0	0	0	3.6
3	1	0	M	0	0	1	0	0	0	0	0	3.5
4	1	0	M	0	0	0	1	0	0	0	0	3.6
5	1	0	M	0	0	0	0	1	0	0	0	2.9
6	1	1	M	0	0	0	0	0	1	0	0	3.6
1	2	0	F	1	0	0	0	0	0	0	0	3.5
2	2	0	F	0	1	0	0	0	0	0	0	2.9
1	3	0	M	1	0	0	0	0	0	0	0	3.3
2	3	0	M	0	1	0	0	0	0	0	0	2.75
3	3	0	M	0	0	1	0	0	0	0	0	3.56
4	3	0	M	0	0	0	1	0	0	0	0	3.61
5	3	0	M	0	0	0	0	1	0	0	0	3.15
6	3	0	M	0	0	0	0	0	1	0	0	2.85
7	3	0	M	0	0	0	0	0	0	1	0	3.2
8	3	0	M	0	0	0	0	0	0	0	1	3.5

- Gender is an example of a time invariant predictor. Since its value is always constant, then its value is repeated in the person period dataset for each quarter attended by the student.

- Term GPA is an example of a time varying predictor. The student's GPA for quarter j is listed in row j of the person period dataset.

- In the person form, all information about each student is located in a single row. In particular, there is information indicating whether or not a student ever graduates and their total number of quarters. This information can be turned into the Y vector that indicates for each quarter whether the student graduated or not.

Suppose that graduation at some university typically takes place in six quarters but information about eight quarters is known. We see that student 1 graduates in six quarters. This results in Student 1's record to be transformed into six rows of the person period dataset. Student 1's gender, a time invariant predictor, is repeated in each quarter. GPA, a time varying predictor, has the quarter j's GPA located in row j of the Student 1's person period record. The censor indicator and information about the total number of quarters attended by the student allow for creation of the Y vector. Students 2 and 3 are both censored; they have not graduated by the end of 8 quarters, the length of the data collection process. Student 2 only attended for two quarters (perhaps they dropped out) and student 3 was still taking courses at the end of the data collection process.

Table 13: Cal Poly Majors STEM Classification

Major Abbreviation	Major Name	STEM Classification
AERO	Aerospace Engineering	STEM
AGB	Agribusiness	non-STEM
AGSC	Agricultural Science	STEM
ARCE	Architectural Engineering	STEM
ARCH	Architecture	non-STEM
ART	Art and Design	non-STEM
ASCI	Animal Science	STEM
ASM	Agricultural Systems Management	non-STEM
BCHM	Biochemistry	STEM
BIO	Biology	STEM
BMED	Biomedical Engineering	STEM
BRAE	BioResource and Agricultural Engineering	STEM
BUS	Business	non-STEM
CD	Child Development	non-STEM
CE	Civil Engineering	STEM
CHEM	Chemistry	STEM
CM	Construction Management	non-STEM
COMS	Communication Studies	non-STEM
CPE	Computer Engineering	STEM
CRP	City and Regional Planning	non-STEM
CRSC	Crop Science	STEM
CSC	Computer Science	STEM
DSCI	Dairy Science	STEM
ECON	Economics	non-STEM
EE	Electrical Engineering	STEM
EHS	Environmental Horticulture Science	STEM
ENGL	English	non-STEM
ENVE	Environmental Engineering	STEM
ENVM	Environmental Management and Protection	STEM
ERSC	Earth Sciences	STEM
ES	Ethnic Studies	non-STEM
FDSC	Food Science	STEM
FNR	Forestry and Natural Resources	STEM
FRSC	Fruit Science	STEM
GENE	General Engineering	STEM
GRC	Graphic Communication	non-STEM
HIST	History	non-STEM
IE	Industrial Engineering	STEM
IT	Industrial Technology	STEM

JOUR	Journalism	non-STEM
KINE	Kinesiology	STEM
LAES	Liberal Arts and Engineering Studies	STEM
LARC	Landscape Architecture	non-STEM
LS	Liberal Studies	non-STEM
MATE	Materials Engineering	STEM
MATH	Mathematics	STEM
MCRO	Microbiology	STEM
ME	Mechanical Engineering	STEM
MFGE	Manufacturing Engineering	STEM
MLL	Modern Languages and Literature	non-STEM
MU	Music	non-STEM
NUTR	Nutrition	STEM
PHIL	Philosophy	non-STEM
PHYS	Physics	STEM
POLS	Political Science	non-STEM
PSC	Physical Science	STEM
PSY	Psychology	non-STEM
REC	Recreation	non-STEM
SOCS	Social Sciences	non-STEM
SS	Soil Science	STEM
STAT	Statistics	STEM
TH	Theatre	non-STEM
WVIT	Wine and Viticulture	non-STEM
SE	Software Engineering	STEM

Bibliography

Allen, J., & Robbins, S. B. (2007). Prediction of college major persistence based on vocational interests, academic preparation, and first-year academic performance. *Research In Higher Education*,*49*, 62-79.

Bayaga, A. (2010). Multinomial logistic regression: Usage and application in risk analysis. *Journal of Applied Quantitative Methods* , *5*(2), 288-297.

Chen, X. U.S. Department of Education, Stats in Brief. (2009). *Students who study science, technology, engineering, and mathematics (stem) in postsecondary education*

Chizmar, J. (2000). A discrete-time hazard analysis of the role of gender in persistence of the economics major. *The Journal of Economic Education*,*31*(2), 107-118.

R Data Analysis Examples Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. from http://www.ats.ucla.edu/stat/r/dae/mlogit.htm/ (accessed August 24, 2011)

Shaw, E. J., & Barbuti, S. (2010). Patterns of persistence in intended college major with a focus on stem majors.*NACADA Journal*, *30*(2), 19-34.

Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Stasistics*, *18*(2), 155-195.