The Roads to Success:
Analyzing Dropout and Degree Completion at University

Elena Arias Ortiz
ECARES, Université Libre de Bruxelles

Catherine Dehon
ECARES, Université Libre de Bruxelles

ECARES working paper 2011-025

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

The Roads to Success: Analyzing Dropout and Degree Completion at University*

Elena ARIAS ORTIZ and Catherine DEHON†

Abstract

In this paper we study the factors that influence both dropout and degree completion (4 or 5 years to earn a degree) at university using survival analysis. In particular, we apply the set of discrete-time methods for competing risks event history analysis described in Scott and Kennedy (2005). Using the competing risks model, we show that foreign students are more likely to experience consecutive enrollments without actually getting a degree. Also, having a mother with a higher education degree reduces significantly the risk of dropping out and at the same time increases the chance of graduation. Finally, the impact of a variable can evolve throughout the academic path. For example, "having chosen a strong mathematical profile during high school " reduces significantly the risk of dropping out only in the early years of study.

1 Introduction

Holding a higher education degree results in important financial rewards in the labor market. In OECD countries, the average income of workers with a university degree is in general two times higher than the income of those who graduated from high school[1]. In addition, higher educational attainment reduces the risk of unemployment. Across OECD countries, the average unemployment rate among those who attain only lower secondary education is 5 percentage points higher than those whose highest attainment is upper secondary, and seven points higher than those

*Catherine Dehon gratefully acknowledges research support from FRFC (Fonds de Recherche Fondamentale Collective) and from the ARC contract of the Communauté Française de Belgique. Catherine Dehon is also member of ECORE, the association between CORE and ECARES.

†Université libre de Bruxelles, ECARES, SBS-EM. CP 114/04, Av. F.D. Roosevelt, 50. B-1050 Brussels, Belgium.

[1]Source: OECD (2007), *Education at Glance: OECD indicators*, OECD, Paris.

who attain tertiary level education. However, choosing to invest in education does not guarantee that the student will actually graduate. Enrolling in a higher education institution represents a considerable financial risk. If a student drops out before graduation, he will have to bear the tuition costs and the foregone income of each year of tuition, without any return (or a very low return) from his investment.

Thus, it is an important policy issue to identify the factors that influence student dropout at university, both for society and for the individuals themselves. However, research in the economics of education has focused on success either at the beginning of the university path (enrollment in a higher education institution) or at the end (probability of earning a degree or of dropping out from university), and has neglected the temporal dimension involved in the learning process. In reality, we observe highly complex educational histories that may involve delayed enrollments, stop-outs, or part-time enrollments, for example. Yet, in educational research, temporal investigations of dropout and timely graduation have been done only infrequently (DesJardins et al. (2002)). To understand the changing circumstances of students as they proceed through their academic careers, a methodology is required that allows us to study transitions from one state to the next (e.g., from being enrolled to not enrolled); we thus use longitudinal data and temporal analytic techniques to fully capture how factors evolve throughout students' academic paths.

We study the factors that influenced both dropout and degree completion of students enrolled at the Universit libre de Bruxelles (ULB) in the academic years 1997-1998 and 2001-2002 using the "survival time" of an individual. This methodology, called event history analysis or survival analysis, was initially developed to study survival times of patients after an intervention. Since the outcome of interest need not necessarily be death, event history analysis can be applied to other fields of study and in particular to educational data (Willett and Singer (1991), Scott and Kennedy (2005)).

The main contribution of this paper is to show that much more can be learned about dropout and graduation by analyzing *when* the event occurs or how the impact of some factors *evolve* through time. For example, thanks to the competing risks model, we show significant differences between the academic paths of foreign and Belgian students. Indeed, being a foreign student reduces the probability of completing the degree, but it does not influence significantly the probability of dropping out.

In agreement with the literature, having a mother that holds a higher education degree influences both outcomes: students are more likely to graduate and less likely to dropout. However, our results suggest that the effect of the mother having a degree on the probability of dropping out is probably due to financial constraints, because having an unemployed father is also a significant factor in student dropout.

Finally, we show that the estimated impact of a variable can evolve through the academic path, as it does for the variable "having a strong mathematical profile during high school". The impact of this variable is stronger in the early stages, as it reduces significantly the probability of dropping out only in the early years (the first four years) of study. Having a strong mathematical profile also increases the chance of timely degree completion.

In addition, our paper shows that studying the timing of dropout and degree completion at university requires some methodological changes that have been overlooked in previous international studies. Dropout and degree completion are two dependent competing risks that lead to "not surviving" at university (i.e. no longer being enrolled at university). If the event of interest is degree completion, dropout cannot be considered as censored observations as they are not independent of the outcome of interest and vice versa. Thus, modeling survival at university requires a competing risks model. Furthermore, in this type of setting, time is organized into teaching terms or academic years of enrollment and a discrete time modeling approach is required to provide appropriate estimates.

This article is organized as follows. In Section 2, we provide some descriptive statistics that explain why it is interesting to look at the entire academic path. In Section 3, we describe and justify our own methodological approach. Also, using the hazard function and the cumulative hazard function, we identify some of the key factors that influence the probability of dropping out and graduating on time. In Section 4, we develop and estimate the competing risks model that allows us to estimate the influence of multiple factors on the hazard probabilities. We conclude in Section 5.

2 Analyzing student's path at university

2.1 The data

In the educational system of the Belgian French community, almost 60 percent of the secondary student population that finishes the general high school system[2] enrolls at university. Furthermore, higher education is largely financed through public funds and as a result, all universities have very low, common registration fees and no entrance examinations (except for the faculty of applied sciences at the ULB). In such a framework, barriers to higher education are relatively low yielding a high en-

[2]The general high school system is the system that prepares students to enter a higher education institution. This excludes students in the technical or the professional system that prepares them to immediately start their professional career.

try rate. To our knowledge, our paper is the first serious attempt to analyze students' entire path at university in the Belgian French Community.

Our data base consists of 5822 newly enrolled students at the Université libre de Bruxelles (ULB) in the academic years 1997-1998 and 2001-2002. It was created by merging the university's administrative data base with the results of a non-compulsory survey that students filled in anonymously at enrollment. The administrative data was collected over a 12-year period and contains information about the entire career path at university of each student as long as the student is enrolled at the ULB (enrollment, results, reorientation, dropout, graduation, etc.). Unfortunately, if a student drops out of the ULB, we do not know whether he enrolls at another university or higher education institution or whether he enters the labor market.

At the time of the data collection, some degree programs required students to successfully complete a minimum of 4 years to earn a degree while others required a minimum of 5 years (for the same level of academic degree)[3]. Thus, we split the data into two sub-samples. One sample contains students who graduated or dropped out from a 4-year degree program ($N_1 = 2973$) such as philosophy, hard sciences, social sciences, political sciences or economics, the other contains students who graduated or dropped out of a 5-year program ($N_2 = 2073$) such as law, business, psychology or engineering[4]. Throughout the paper, we will focus mainly on the 4-year program sample, given that for the 2001 cohort, we have data for more years after the minimum number required to get a degree. However, we analyze in detail the cases in which students in the 5-year program sample present particular or interesting behavior with respect to those in the 4-year program sample.

The database used is very rich as it covers 3 different aspects of the student's life (see Table 1). First, the personal characteristics of the student, such as the year of the first enrollment, gender and nationality. Second, the socioeconomic background of the student, measured by the mother's educational attainment and the father's professional activity. Third, previous high school track characteristics, such as the intensity of the mathematical profile the student chose during high school (a weak mathematical profile if the student chose less than 5 hours of math per week), starting the university on time (first enrollment in the academic year that started 17 years after his date of birth)[5], or the type of high school the student attended (the

[3]The students enrolled in Medicine are not studied in this paper because their degree requires a minimum of 7 years of full-time enrollment.

[4]For a complete description of the samples and the degree programs contained in each see Table 7 in the Appendix.

[5]In Belgium, the law requires children to start their primary education in the academic year in which they turn 6 years old.

traditionnel -more rigid with specialization tracks or the *rénové* - more flexible)[6]. We can see from Table 1 that there are few significant differences in the composition of the two samples. The most important difference between the two samples is previous school path. Indeed, 5-year programs seem to attract a higher proportion of students who completed their secondary schooling on time (65% against 53% in the 4-year program sample) as well as students that chose a strong mathematical profile (56% against 33% in the 4-year program sample). Also, there is a higher proportion of students who obtained their high school diploma in a *traditionnel* school enrolled in 5-year degree programs.

Table 1: Descriptive statistics of the 4-year program and the 5-year program samples

	4-year Sample	5-year Sample
Generation 2001	49%	52%
Gender (Female=1)	55%	50%
Foreigner	20%	15%
Mother with Higher Education	55%	58%
Father occupational activity		
Low level employee	28%	24%
High level employee	62%	67%
No profession/unemployed	10%	9%
Traditionnel	9%	14%
Weak Mathematical profile	67%	44%
On time	53%	65%
N	**2973**	**2073**

2.2 Why should we study outcomes after the first year at university?

We basically use the same data base as Arias and Dehon (2008) and Arias and Dehon (2010) but they focus on the probability of successfully completing the first year at university as a measure of academic success. Is success during the first year is a good predictor of a student graduating? Are students who fail their first year condemned to leave the university without a degree? So far, no answer has been provided for the Belgian French community. The objective of this paper is to find

[6]The difference in the success rates observed in the first year at university between students from these two educational systems is analyzed in Arias and Dehon (2010).

out what the factors are along the path that lead students to graduate or to leave the university without a degree.

The descriptive statistics shown in Table 2 suggest that many different things can happen after the first year as there are many different career paths leading to either dropout or graduation. First, almost 30% of students drop out of university after only one year of enrollment and, in that group, a very large majority of students fail their first year. The second most common path after the first year is failure and then re-enrollment in the same field (27.72%). Surprisingly, only 38% of these students will graduate at some point on their academic path, so a large proportion of them (62%) will invest in two or more years at university and never graduate.

Students who succeed their first year and obtain their degree on time represent only 17.86% of first year students. Interestingly, the fourth most common path is characterized by students who successfully complete their first year but obtain their degree only after more than 4 years (10.83%) implying that some students encounter difficulties along the way after the first year. Finally, we notice that after a failed year, 8.75% of the students re-enroll at university, but in a different field. In this group of students, a higher proportion obtain a degree than among those that re-enroll in the same field (49% vs. 38%).

Table 2: Main career paths after their first year at the ULB for students in the 4-year program sample (N=2973)

	Main career paths	
1.	Drop out 1st year	29.87%
2.	Failed 1st year and start again 1st year (Degree 38%)	27.72%
3.	Degree on time	17.86%
4.	Succeeded 1st year and degree late	10.83%
5.	Failed 1st year and reorientation (Degree 49%)	8.75%
6.	Other	4.98%
	Total	100%

Thus, even if a successful first year is a key step along the academic path, Table 2 shows the complexity and the multiplicity of paths students can follow. The configuration of the data is so complex that it requires a special methodology designed to deal with the multiple outcomes and multiple periods. For modelling purposes, we focus on two main outcomes: dropout and graduation.

The descriptive statistics corresponding to these two outcomes are given in Table 3 for both samples. The first thing to notice is that around 55% of the students will quit the university without a degree, whether we consider students in the 4-year program sample or in the 5-year program sample. More importantly, we can see that almost half of student dropout occurs after the first year of enrollment. Also, we observe that on average 40% of the students obtain a degree in the two samples. In the case of the 4-year program sample, actually a large majority obtain their degree after experiencing 1 or more failed years in their path. Finally, a higher proportion of students are censored in the 5-year program sample (that is, there was a higher chance that students enrolled in longer degree programs did not finish their degree by the end of the data collection). This can explain why the graduation rate is slightly higher for students enrolled in a 4-year program, even if in general aggregate dropout and graduation rates are very similar in both samples.

Table 3: Descriptive statistics of the 4-year and the 5-year program samples

	4-year Sample		5- year Sample	
Drop out:	1637	55.06%	1186	57.21%
a) Drop out 1 year	888	54.21%	552	46.19%
b) Drop out after the first year	749	45.75%	643	53.81%
Degree:	1281	43.09%	825	39.80%
a) On time	531	41.45%	486	58.9%
b) 1 year late	415	32.24%	195	26.63%
c) 2 or more years late	335	26.15%	144	17.45%
Censored	55	1.85%	62	2.99%
Total	2973	100%	2073	100%

As mentioned earlier, around 50% of the cases of dropout occur after the first year of enrollment. Among students who did not dropout during the first year in the 4-year program sample, the mean number of enrollments until dropout is 3 years. Is the risk of dropping out constant over time or does it disappear after the first year of enrollment? Is the probability of graduating equal to zero if a student fails one or several years at university? To answer these questions, we need to analyze the timing of events throughout the whole academic path. In the following sections, survival analysis is applied to analyze how the factors that influence dropout and graduation evolve throughout the path at university.

3 Analyzing educational data using survival analysis

3.1 Methodological approach and previous research

The statistical methodology known as survival analysis, event history analysis or hazards modeling, is appropriate for modeling response variables defined as the ‘time to the occurrence of an event”. As the name indicates, these methods were initially developed to study survival times for patients after an intervention and they have been widely applied in medicine and biology for example. Because the outcome of interest need not necessarily be death, event history analysis has been extended to other areas such as sociology (time to first marriage), economics (time to return to work) or education (time to withdrawal from university or time to graduation). In the previous section we showed how complex can be the paths followed by students during enrollment at university. Thus, in this paper, we apply survival analysis methods to analyze the behavior of students along the entire path at university.

In recent years, several papers have been entirely devoted to explaining the reasons why survival analysis is appropriate to studying the risk factors influencing the timing of educational events. For example, Willett and Singer (1991) explain how much more can be learned about student retention or timely graduation by modeling *when* the events occur instead of *whether* the events occur. In a nontechnical introduction to survival methods, they describe four main advantages offered by their use. First, describing cumulative differences in behavior until a determined time point, as in previous studies, does not take into account the fact that the risk of an event can vary over time. For example, the fact that two groups of students have identical attrition rates at one point in time does not imply they followed the same trajectories. Variation in risk is one of the main analytic focuses of such models. Second, methodologies that analyze separately individuals who experienced the outcome and those who did not, such as two-sample comparisons or even dichotomization, can obscure knowledge about educational transitions. They can also lead to contradictory results when using different databases and different cutoffs to split individuals (e.g. graduation after 3, 4 or 5 years). Third, there are always individuals for whom we do not observe the outcome of interest and survival analysis allows this partial information to be included in the model. For example, what value should be attributed to students who do not dropout and do not get a degree during the period of data collection? This partial data is called *censored* data, an aspect of the data that can only be incorporated using event history analysis. Finally, this method offers the possibility to model variables that evolve over time or to analyze predictors that are constant in value but whose effect on the behavior of students can vary across their paths at university.

Murtaugh et al. (1999) use survival analysis to study the time until student withdrawal at Oregon State University. Their findings suggest that studying the timing of events matters, as they discovered that withdrawals tend to concentrate at the end of each year. Furthermore, by applying multi-variable analysis, they show that observed differences in dropout rates between white and black students disappear and are even reversed when they account for differences in ability and family background. Another surprising result is that retention rates decrease as student age at enrollment increases. An obvious limitation of this paper is that individuals who graduated are considered as being censored, whereas in reality students can exit school in different ways (graduate, dropout, or stop out). Thus, individuals who graduated from university can no longer dropout and thus cannot be considered as "individuals who experience no event during the time period". When more than one event needs to be considered, we have to apply competing risks analysis, crucial in the setting of university because of the interdependencies that exist between competing outcomes like dropout and graduation. Still, these models have been used infrequently in the educational literature.

DesJardins et al. (2002) are among the first to model dropout and graduation as two jointly determined events in a competing risks framework. Analyzing a sample of students from the University of Minnesota, the authors find evidence suggesting that single risk models that assume event independence may be inappropriate and lead to spurious conclusions. For example, one of their main findings is that financial aid does not appear to increase graduation rates directly, but it does so indirectly by reducing student dropout. Jakobsen and Rosholm (2003) are a second exception as they apply a duration model taking into account that students can leave the educational system either by dropping out or after getting a degree. They are particularly interested in the educational behavior of first generation immigrants enrolled in qualifying education in Denmark from 1984 to 1999. They also find a difference in the factors that influence dropout and graduation rates, such as being married. Female students that get married have lower dropout rates, but there is no significant effect on completion rates.

By using a competing risks approach, the authors of these two articles improved on other event history models; however, their models are estimated using continuous time, indirectly assuming that the precise time of occurrence is known. In the case of educational data, this assumption is fairly unrealistic and does not apply to the school context as time periods are often discrete (quarters, semesters or years). This is why Singer and Willett (1993, 2003) derive maximum likelihood estimators for discrete-time hazard models and show that in a single outcome framework, the model can be fitted using standard logit regression. In addition, they consider that the comprehensibility of discrete-time survival analysis is an important feature of this model relative to a continuous time proportional hazard model (Cox regression)[7], known for being less intuitive than standard statistical tools.

Thus, in the following sections, we apply a model for the timing of dropout and degree completion at university that integrates all the methodological requirements that have been overlooked in previous international studies. We apply the set of discrete-time methods for competing risks event history analysis described in Scott and Kennedy (2005). The authors show that combining a model of discrete-time with a competing risks setting results in a model that can be estimated using a multinomial logistic regression, making the survival analysis approach more accessible.

3.2 Hazard probabilities

The descriptive statistics presented previously yield important information about the aggregate number of students who dropout and graduate. In this section, we analyze in greater detail *when* these outcomes occur along the academic path. Indeed, by applying survival analysis methodology to our data, we will be able to identify in the university path the specific years in which students are more likely to either dropout or graduate.

The very first step is to determine *when* students are more likely to experience one of the two events of interest, and to compute the sample hazard function. Generally speaking, suppose that there are K outcomes of interest $1, .., K$ (in our case $K = 2$) and the nonoutcome 0. For each point in time, we can compute the hazard that the outcome k occurs at time t, $h(k, t)$. Given that we have discrete time points (number of years), we call this a discrete time hazard; it can be defined as the conditional probability that a randomly selected individual will experience the event k, given that the nonoutcome 0 has happened in every period before t. Thus, for each year, we have to define the set of students who are at risk and calculate the proportion of this group that either leaves university or graduates, given that they survived all the previous years.

We start by defining exactly *who* is at risk in order to compute the hazard probabilities. In our competing risks setting, students can experience only one of the two outcomes of interest (dropout or graduation). If one of the outcomes occurs, the student is can no longer be at risk of experiencing any outcome (he is no longer enrolled at university). In Table 4 we see that out of the 2973 students that were at risk during the first year (at risk because they were enrolled in their first year), 888 students dropped out at the end of the year, yielding a hazard probability of

[7] For more details about these methods see Kleinbaum and Klein (2005).

approximately 30%.[8] In the second year, there are only 2085 students who are still at risk, given that they did not experience either of the two outcomes of interest in the first year, that is they "survived" at university. In this new point in time at university, the risk of dropping out decreases: around 20% of all the students enrolled in a second year (the ones at risk) dropout at the end of the year. The same logic applies in all subsequent years, until the 8th year of enrollment, which contains all the information about the 8th, 9th, 10th and 11th years of enrollment. The difference between the number of outcomes in period $8+$ (44) and the number of individuals at risk remaining at the beginning of the 8th year (99) gives the number of censored observations of the sample (55).

Table 4: Discrete time hazard probabilities for students enrolled in a 4-year program

		Frequencies		Hazard probabilities	
Year	Population	Drop-out	Degree	Drop-out	Degree
1	2973	888	0	29.87	0
2	2085	422	0	20.24	0
3	1663	143	0	8.59	0
4	1520	74	531	4.87	34.93
5	915	35	415	3.83	45.36
6	465	42	217	9.03	46.67
7	206	23	84	11.73	42.86
8+	99	10	34	7.46	25.37
Total	2973	1637	1281	55.06	43.62

In order to interpret the evolution of the risk through time, it is usual to plot hazard probabilities against time. This is called the hazard function and it is shown in Figure 1. The risk of dropping out is the highest during the first two years of enrollment at university: almost 30% of the students at risk dropout at the end of the first year and 20% at the end of their second year, conditional upon still being enrolled. The estimated risk of dropping out continues to decrease in subsequent years of enrollment, which is intuitive considering that the cost of dropping out without a degree is higher the more time you spend at university. Surprisingly, we observe that the risk of dropping out rises again after the 5th year of enrollment among the number of students at risk, that is those who are still enrolled at university after 5 years. Finally, as shown in the descriptive statistics, around 35% of the students obtain their degree on time. However, thanks to the analysis of the hazard function,

[8] As explained earlier a student cannot finish a 4-year career in less than 4 years, so the hazard probabilities of the outcome *degree* are equal to zero for the 3 first years of enrollment.

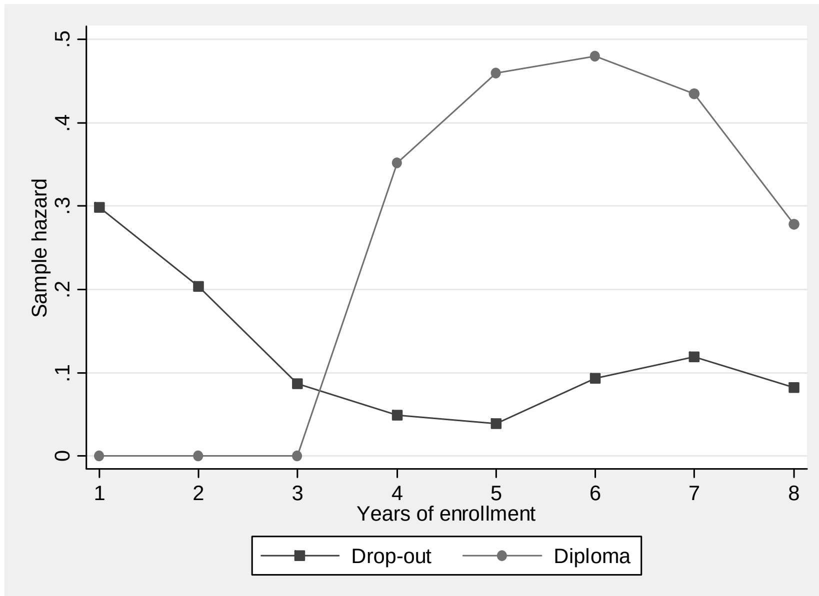

Figure 1: Drop-Out and Degree hazard functions (4-year program sample)

we can see that conditional upon surviving the first 4 years, the probability of graduating one year late, that is after five years, is actually higher than that of graduating on time (45% against almost 35% among those at risk, respectively). Overall, the hazard function of getting a degree shows that a large number of students obtained their degree after 5 or 6 years of enrollment, but the probability of graduating declines after that. The same pattern is observed for students in the 5-year program sample.

One useful feature of the hazard function is that we can compute it for different types of students in order to determine their specific path at university in terms of dropping out or graduating. For example, in Figure 2 we computed the hazard function of dropping out and obtaining a degree separately for Belgian and foreign students, revealing the specific behavior of these two subgroups on the two outcomes. Belgian students have a higher chance of graduating at each period than foreign students and they also have a lower probability of dropping out, particularly during the first two years of enrollment. It is important to highlight, however, that this descriptive analysis through hazard functions is a bivariate analysis and does not take into account differences in other variables. Again, the results we obtain for the 5-year program sample are very similar.

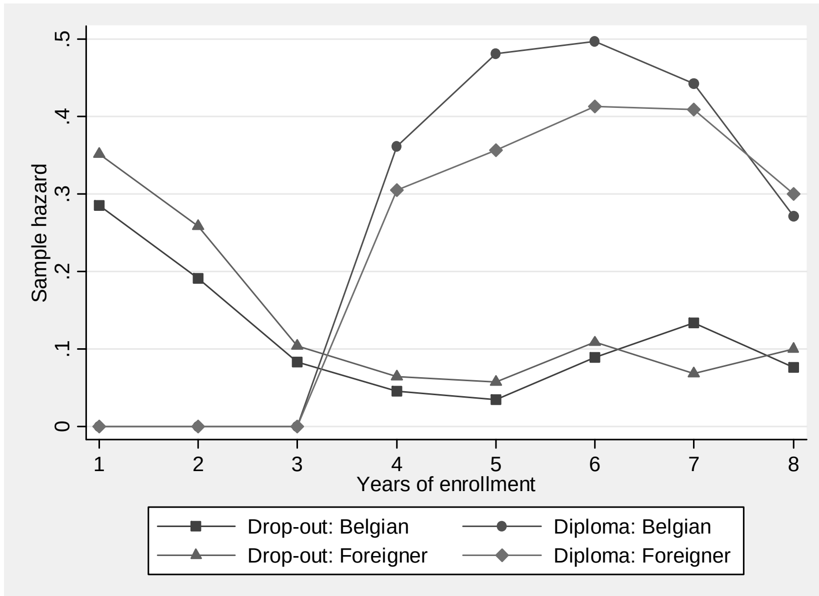

Figure 2: Drop Out and Degree hazard functions by nationality (4-year program sample)

The hazard functions also reveal that some sub-populations do not have the same behavior whether we consider 4-year programs or 5-year programs, as shown by the example in Figure 3 concerning gender differences. In the 4-year program sample, women have a higher probability of obtaining a degree during the 4th, 5th and 6th year of enrollment, which is not the case in the 5-year program sample. The difference in the effect of gender between the two samples is even more striking if we consider the risk of dropping out: while male students have a higher probability of quitting in the first 3 years of their university path in a 4-year program, males tend to dropout less if they are enrolled in a 5-year program. The descriptive statistics show that the percentage of women is approximately equivalent in both samples, so the difference in behavior cannot simply be explained by a sampling effect. A deeper analysis on gender issues should be carried out with other data in order to explain this phenomenon.

Finally, we analyze how the result of the first year is an important predictor of the rest of the career path at university. Figure 4 displays drop-out and graduation hazard rates for the three possible outcomes at the end of the first year: i) the student passed directly (i.e.*success in first session*), ii) the student first failed, retook the

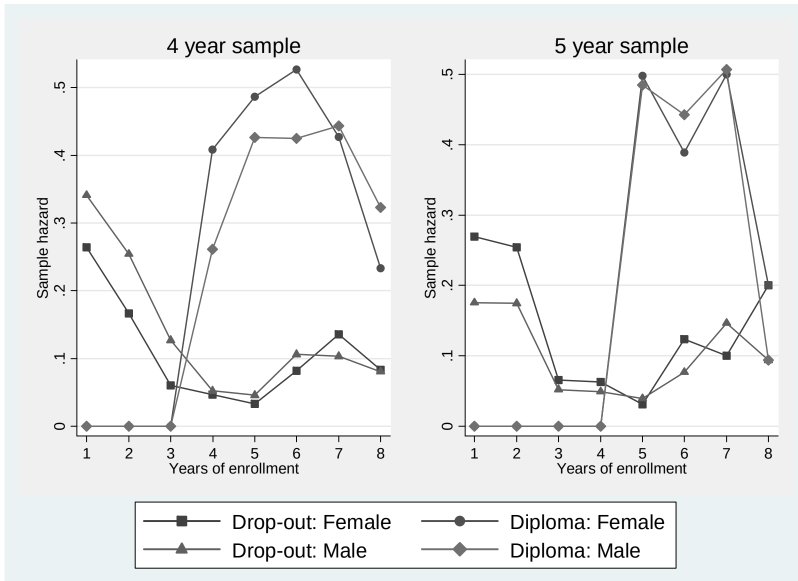

Figure 3: Drop Out and Degree hazard functions by gender

exams and passed (i.e. *success in second session*), iii) the student failed in the second session, meaning he failed the first year (i.e. *failed second session*). In other words, students can progress to the second year using two different paths or they can fail. Looking at the plot, we observe that students who passed directly have higher chances of getting their degree on time than students who passed using a second round of exams. For the latter, the probability of graduating in the fourth year of enrollment is barely over 40%. For the former, the probability of getting a degree on time is near 70%. The gap in the probability of getting a degree is not only observed for timely graduation, it is also observed for the probability of graduating after 5 years of enrollment. For those who failed the first year, the probability of getting a degree on time is obviously equal to zero and it remains significantly lower in the 5th year of enrollment compared to students who passed their first year. We also observe from the plot that students who fail their first year have a much higher probability of dropping out during almost all their career path at university. However, the probability of dropping out is practically the same whether we consider students who succeed in the first or in the second session of exams. Thus, even if some differences exist between passing in first or second session, whether the student fails or passes the first year is the outcome that will really influence the rest of his career path at university. In Arias and Dehon (2008), the authors analyze the factors that influence success at university by focusing on the outcome of the first year. Our results show that this is indeed a good predictor of dropout, especially if no distinction is made between passing in first or second session in the measure of success.

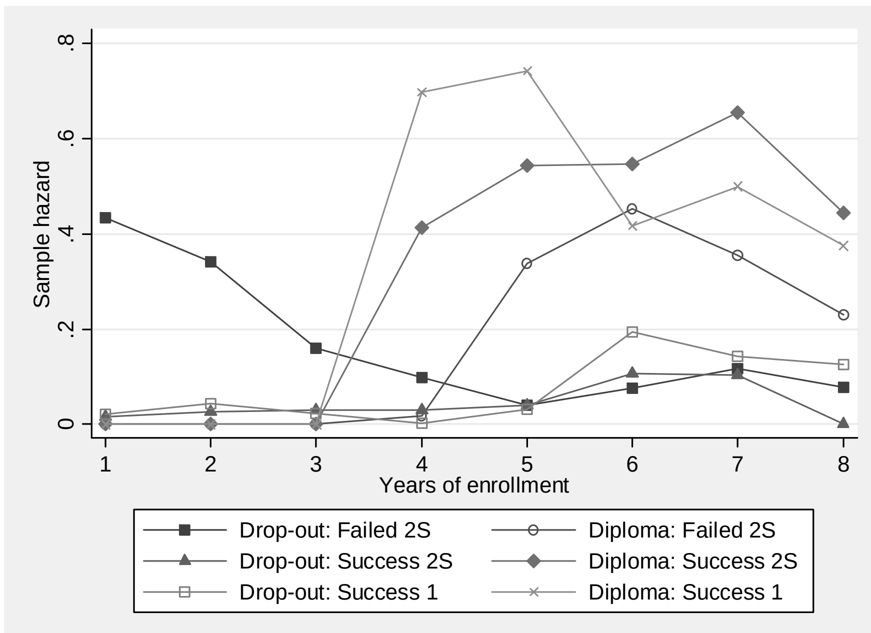

Figure 4: Drop Out and Degree hazard functions by first year outcome (4-year program sample)

3.3 Cumulative function

The hazard functions described in the previous sub-section yield instantaneous pictures of the different stages of students' path at university. They are very informative to understand which outcome we can expect to observe, at each period, for students who have survived until that period. However, in order to analyze in greater depth how many students have experienced a certain outcome by a certain point in time, we need to introduce cumulative functions. Let $M(k,t)$ be the cumulative probability of $k \in \{1,\dots,K\}$ at time t, that is the probability that outcome k occurs in the first t periods[9]. In addition, let

$$M(t) = \sum_k M(k,t),$$

be the probability that *any* of the K outcomes of interest occurs in the first t periods. In our case, this implies either dropping out or graduating. We could also compute the *survivor function* $1 - M(t)$, the probability that no outcome occurs by time t, interpreted as the probability of surviving until time t. We can see that $M(k,t)$ can be computed recursively using:

$$M(k,1) = h(k,1)$$

$$M(k,t) = h(k,t)[1 - M(t-1)] + M(k,t-1) \text{ for } t > 1 \tag{1}$$

where $h(k,t)[1-M(t-1)]$ is the probability that k occurs in t, given that the hazard in period t acts only on the population that survived on period $t-1$. In the same way, $M(t)$ is given by:

$$M(1) = \sum_k h(k,1)$$

$$M(t) = \left[\sum_k h(k,t)\right][1 - M(t-1)] + M(t-1) \text{ for } t > 1. \tag{2}$$

The estimates of $\widehat{M}(k,t)$ and $\widehat{M}(t)$ based on the sample are displayed in Figure 5 respectively for the 4-year program sample (the 2 plots in the top panel) and the 5-year program sample (the 2 plots in the lower panel). The figures reveal that the cumulative probability of getting a degree never outstrips the cumulative probability of dropping out in either of the two samples. In addition, the gap is more important for students enrolled in a 4-year program. Analyzing the total cumulative function gives a good summary about the *timing* of events at university: after 6 periods, 93% of the students in the 4-year program sample and 88% of the students in the 5-year program sample have experienced one of the two outcomes. For both samples, these percentages are obtained because around 50% of the students have dropped out and 40% have obtained a degree after 6 years.

The total cumulative plots display another useful observation: the median survival time of an individual. We find that the median survival time is approximately 3 years of enrollment for the 4-year program sample meaning that 50% of the students

[9]The statistical concepts defined in this article rely heavily on Scott and Kennedy (2005) and Willett and Singer (1991).

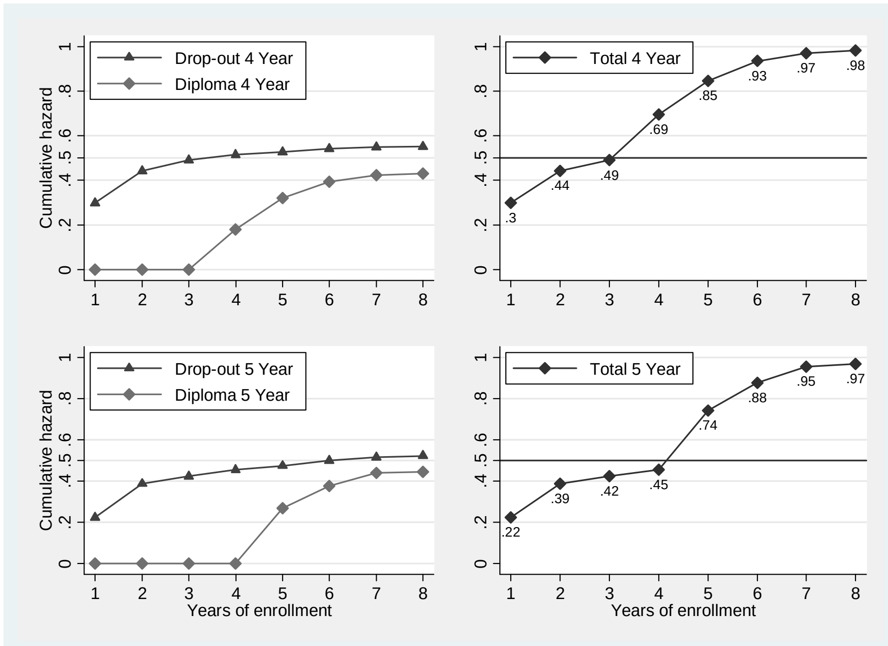

Figure 5: Cumulative hazard functions

experience an outcome before that time and the other 50% will experience an outcome after the 3 years. In our case, we have to bear in mind that surviving means experiencing no outcome, so only still being enrolled. Given that in this sample, students can graduate only after 4-years of enrollment, this result basically implies that 50% of the students drop out of university without a degree after a maximum 3 years of enrollment. For students enrolled in a 5-year program, the mean survival time is also inferior to the minimum amount of time required to get a degree, but in this case the 50% of the risk set is exhausted after approximately 4 years of enrollment.

The cumulative functions discussed so far did not show important differences between students enrolled in 4-year programs and those in 5-year programs. However, when we examine each subsample for specific student characteristics (socio-economic background, personal characteristics, previous high-school track), we do observe some differential effect on the career path at university between the two samples.

For example, Figure 6 shows that, unsurprisingly, the level of education of the mother has a strong impact on the cumulative hazard functions of both samples.

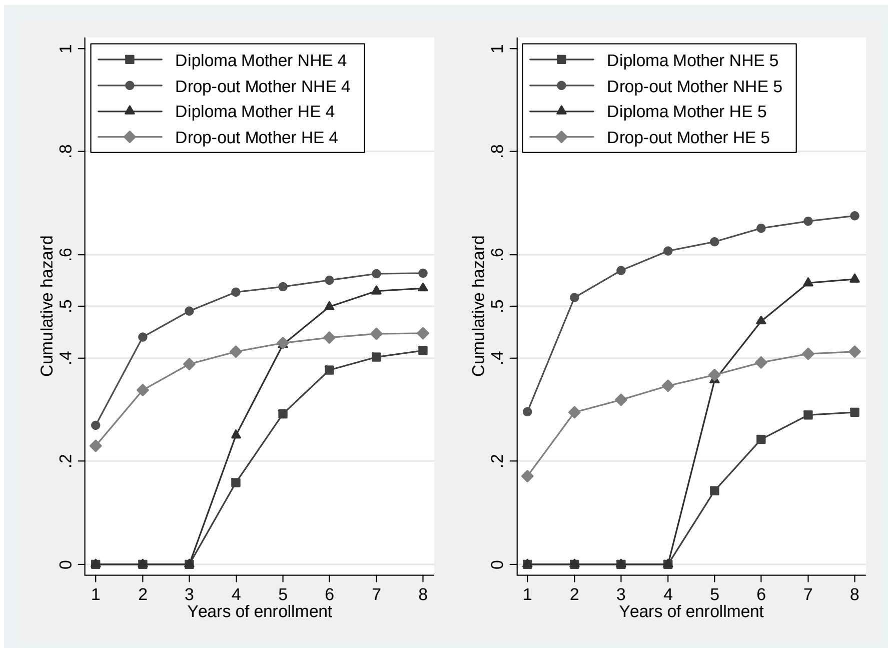

Figure 6: Cumulative hazard functions by educational level of the mother

Having a mother with a higher education degree decreases the cumulative probability of dropping out and increases that of getting a degree. What is surprising, though, is that the effect is much greater for students enrolled in the 5-year program sample. In the 5-year program sample, the cumulative probability of graduating if the student's mother has no higher education degree remains particularly low, barely exceeding 30%, whereas in the 4-year program sample, the cumulative probability of graduating reaches more than 40%.

The mathematical profile chosen during high school is also a crucial factor influencing a student's path at university as shown in Figure 7. Indeed, for the 5-year program sample, which includes more science-oriented studies, the cumulative probability of dropping out increases very sharply when the student chose a weak mathematical profile during high school and reaches 70% by the 7th year. In comparison, this probability never exceeds 40% for students with a strong mathematical profile. The latter group of students also displays a higher cumulative probability of getting a degree, especially for getting a degree on time. It is intuitive that the difference between these two types of students is much lower for the 4-year program sample since these tracks are more oriented to human and social sciences.

Finally, we also investigate the effects of attending a high school system with

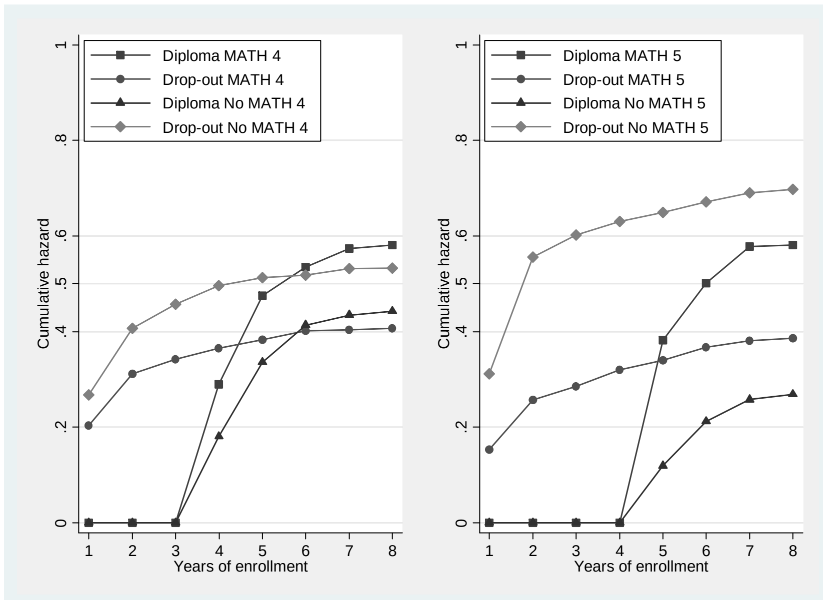

Figure 7: Cumulative hazard functions by mathematical profile

a more rigid curriculum in terms of optional courses (*traditionnel* system) instead of a more flexible secondary school curriculum (*rénové* system). This is interesting given that Arias and Dehon (2010) find that attending a *traditionnel* school has a positive and causal effect on the probability of passing the first year at university. Figure 8 shows that attending a *traditionnel* school as opposed to a *rénové* school has a stronger impact on the cumulative function of obtaining a degree in the 5-year program sample. The cumulative probability of getting a degree outstrips the probability of dropping out in this sample. The gap is smaller if we consider the cumulative probability of dropping out. In any case, we notice that students from the *rénové* have a higher cumulative probability of dropping out and at the same time a lower cumulative probability of graduating than students from a *traditionnel* school. The difference in performance is greater when we consider students from 5-year programs and it lasts for the whole academic path. This implies that the effect of this system does not disappear after the first year: it affects the probability of dropping out and graduating during the entire university path. However, important socioeconomic differences exist between these two groups of students, so before drawing any conclusions we need to validate the results in a multi-variable model.

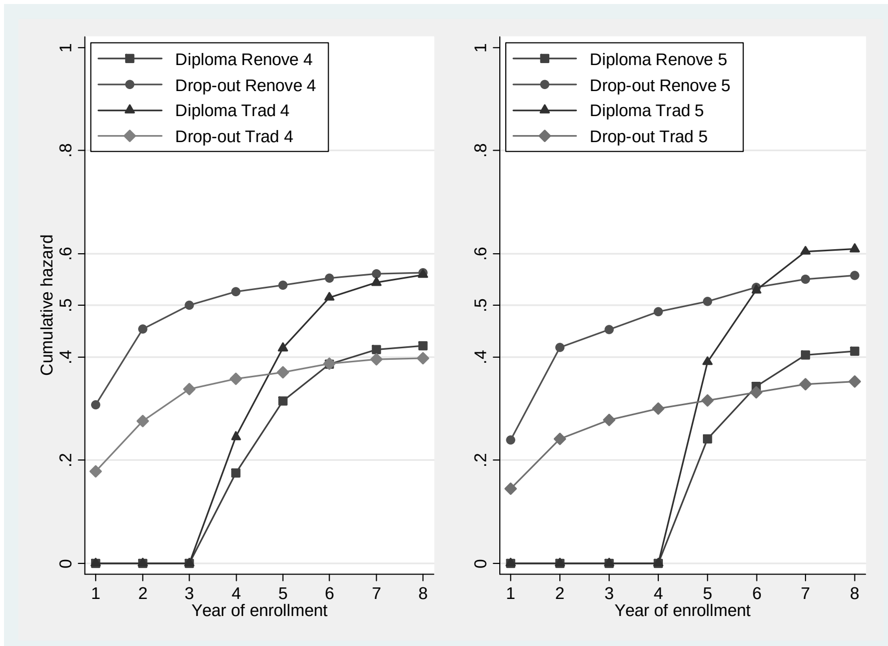

Figure 8: Cumulative hazard functions by type of high school

4 Competing risks: a multivariate analysis

4.1 Deriving a multiple-outcome discrete time model

The descriptive statistics do not allow us to control simultaneously for multiple factors that might affect the probability of a certain outcome. How to measure the effect of several covariates on the probability of either dropping out or obtaining a degree? As already mentioned, to combine the discrete-time base and the competing risks context, we follow the methodology introduced by Scott and Kennedy (2005). They showed that the single outcome case can very easily be extended to the competing risks setting using a multinomial logistic regression.

The statistical model has two essential attributes: (a) a baseline profile of risk and (b) a shift parameter that captures the effect of the predictor on the baseline profile in order to establish a relation between hazard probabilities and predictors. Thus, the proposed hazard model for subject $i \in \{1, \ldots, n\}$ of outcome

$k \in \{1, \ldots, K\}$ at time $t \in \{1, \ldots, T\}$ is:

$$h_i(k,t) = \frac{\exp[(\alpha_{k1}D_{i1} + ... + \alpha_{kT}D_{iT}) + (\beta_{k1}X_{1it} + ... + \beta_{kp}X_{pit})]}{1 + \sum_{l=1}^{K} \exp[(\alpha_{l1}D_{i1} + ... + \alpha_{lT}D_{iT}) + (\beta_{l1}X_{1it} + ... + \beta_{lp}X_{pit})]}, \quad (3)$$

where $(X_1, ..., X_p)$ are the p covariates included in the model, β_{lp} is the parameter associated with the covariate p and the outcome l (l ranges from $1, ..., K$ and the nonevent 0 is not considered as an outcome), $[D_{i1}, ..., D_{iT}]$ are dummy variables indexing time points and T refers to the last time period observed for anyone in the sample. The time period dummies are all identically defined as the example for the first dummy: $D_{i1} = 1$ when the observation for individual i came from the first year of enrollment, and $D_{i1} = 0$ when the observation came from any other year (2 through T). The intercept parameters $[\alpha_{k1}, ..., \alpha_{kT}]$, capture the baseline level of hazard in each time period. As we will see later, the slope parameters $[\beta_{k1}, ..., \beta_{kp}]$ describe the effects of the predictors on the baseline hazard function (on a logit scale). It is important to notice that these models allow for explanatory variables that vary over time (which explains why the X's are time indexed t), but in our paper we only have variables that are constant over time.

Taking logistic transformations of both sides of Equation (3) we obtain the following expression:

$$\log[\frac{h_i(k,t)}{h_i(0,t)}] = (\alpha_{k1}D_{i1} + ... + \alpha_{kT}D_{iT}) + (\beta_{k1}X_{1it} + ... + \beta_{kp}X_{pit}), \quad (4)$$

where $h_i(0,t)$ is the hazard of the nonevent defined as $1 - \sum_{l=1}^{K} h_i(l,t)$. This expression makes clear that we are assuming there is a linear relation between the logistic transformation of the hazard ratio and the predictors and not with the hazard probabilities themselves. The quantity $\frac{h(k,t)}{h(0,t)}$ is called the outcome-specific hazard ratio because it measures the relative risk of experiencing event k with respect to the risk of experiencing the nonevent. Thus, the basic element to take into account is that in the multinomial setting the outcome-specific hazard ratio compares one outcome to the nonevent, rather than to its complement (as in the single outcome setting).

For simplicity, the multivariate analysis is divided into two steps. First, we estimate an initial model containing only the effect of time (i.e. 8 period-indicator variables) in order to understand the interpretation of the time intercepts on hazard probabilities. In the second step, we add the effect of the covariates and estimate the full model.

4.2 Understanding the effect of time and the covariates

When the value of all the covariates is set to zero, the hazards probabilities depend only on the intercepts and the time dummies. They represent "the time period by time period conditional log-odds that the baseline group will experience the outcome k in each time period, given that they have not already done so" (Singer and Willett (1993)). To understand further the effect of time on the hazard probabilities, we estimate first an initial model containing only the effect of time $t \in \{1, \ldots, T\}$ (i.e. 8 period-indicator variables):

$$\log[\frac{h_i(k,t)}{h_i(0,t)}] = \alpha_{k1} D_{i1} + ... + \alpha_{kT} D_{iT} \quad i \in \{1, \ldots, n\}, k \in \{1, \ldots, K\}. \quad (5)$$

This is the simplest model that we can estimate. Not only it is useful to compare and understand more complex models, we can also show that we can recover the whole-sample hazard function estimated in Section 3.

The estimation results are displayed in Table 5. The top two rows of the table give the estimation of the coefficients $[\widehat{\alpha}_{k1}, ..., \widehat{\alpha}_{kT}]$ for $k = 1, 2$. In order to see the effect of a particular time period on the outcome-specific hazard ratio we need to apply the exponential function to both sides of Equation 5 (rows 3-4). Overall, we see that the odds of dropping out relative to the nonevent are above 0.2 during the first two semesters, and then they take values around 0.10 from years 3 to 6. In the case of the graduation outcome, we cannot interpret the estimated time coefficients for periods 5, 6 and 7 given that they are not significant. Finally, the hazard probabilities given in rows 5 and 6 are obtained simply by applying the relation given in Equation 3. These probabilities are identical to the non parametric MLE of hazard given in Section 3,[10] that is, the hazard function of the 4-year program population. Thus, the estimation of this baseline model allows us to see how the multinomial logit model is used only as a statistical tool to establish a relation between the hazard and the covariates.

Concerning the interpretation of the slope parameters $[\beta_{k1}, \ldots, \beta_{kp}]$, the estimated coefficient associated with the pth variable and the kth outcome $\widehat{\beta}_{kp}$ measures the vertical shift in the baseline hazard function between two different values of a particular predictor X_p. Indeed, the logit of the hazard defined by Equation 4 is linear in the covariates, implying that the effect of a particular predictor X_p is to

[10]It is important to highlight that these hazards probabilities are not exactly equal to those seen in Table 4 because in the first section we analyzed the whole sample and not the smallest complete sample required for estimating the model. The results, however, hold independently of the sample we use.

Table 5: Results of fitting the baseline model

	Years of enrollment							
Event type	1	2	3	4	5	6	7	8
Coefficients								
Dropout	-1.16	-1.61	-2.65	-2.24	-2.11	-1.82	-1.16	-2.19
Degree	-	-	-	-0.38	0.14	0.17	-0.22	-0.73
Odds ratio								
Dropout	0.31	0.20	0.07	0.11	0.12	0.16	0.31	0.11
Degree	0.00	0.00	0.00	0.68	1.15	1.19	0.80	0.48
Outcome hazard (x100)								
Dropout	23.83	16.69	6.59	5.92	5.30	6.90	14.87	6.98
Degree	0.00	0.00	0.00	38.23	50.75	50.57	37.84	30.23

Note: all coefficient are significant at 0.01 level except for periods 5, 6 and 7 of the outcome degree

vertically shift the logit hazard function at every point in time. This requires the logit-hazard profiles to have the same shape and to be parallel for each value taken by the explanatory variable (for example men and women) given that they differ only by this vertical shift. This is called the proportionality assumption. In a majority of cases this assumption is relatively reasonable. However, valuable information about the behavior of our estimates can be learned by verifying the validity of this assumption for our covariates before we add the set of explanatory variables to our model.[11]

By using a graphical method, we can easily verify visually which variables are likely to violate this proportionality assumption. The idea is just to plot together the logit-profile for each outcome, for example dropping out ($k = 1$), computed as:

$$\log\left(\frac{h_X(1,t)}{h_X(0,t)}\right) \text{ for } t = 1, ..., 8$$

[11]Actually, there are 3 assumptions required for estimating a discrete-time hazard model: (i) the linearity assumption, (ii) no unobserved heterogeneity and (iii) the proportionality assumption. The first one is not relevant in our case given that the entire set of explanatory variables is composed of dummy variables. The second assumption is derived from the fact that discrete-time hazards models do not include an error term, implying that all the variation in the hazards profiles comes from the variation in the value of the covariates. The problem with this assumption is of course that omitting an important predictor from the model can then have severe consequences on the estimates. Some have suggested ignoring this unobserved heterogeneity and some have suggested as a solution including a random error term in the model (Allison, 1982), but all the existing solutions bring with them other problems to the estimation of the model. The bottom line is that so far there are no simple statistical methods that deal directly with this problem.

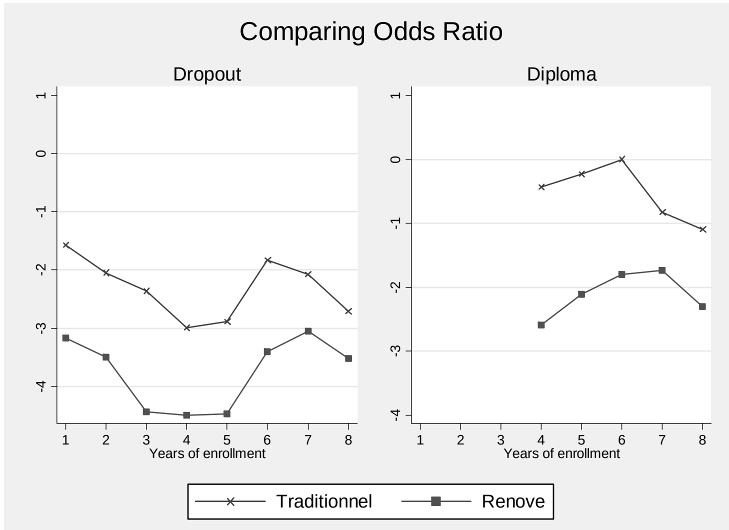

Figure 9: Comparing logit-hazard profiles between the types of school

for all values of X. We can apply the same calculations for the logit-profile of getting a degree ($k = 2$). In Figure 9, we can see the logit-hazards ratio computed separately for students from the *traditionnel* schools and the *rénové* schools. The logit-hazard profiles for each type of school are approximately parallel for both, dropout and degree. This implies that the effect of *traditionnel* on the probabilities of dropping out and getting a degree has the same shape at each point in time. Conversely, Figure 10 below shows that the variable "having a weak mathematical profile during high school" does not satisfy the proportionality assumption. Indeed, the difference in the effect between a strong mathematical profile and a weak mathematical profile on the probabilities of dropping out and getting a degree changes over time. We see that the logit-hazard profiles do not have the same shape at all and the profiles even cross each other at later stages of enrollment. After analyzing the behavior of all our explanatory variables, the only other variable that did not satisfy this assumption is finishing school on time with respect to peers of the same generation.

The results from the visual examination of the graphical logit-hazard profiles revealed that the effects of the explanatory variables $Math$ and $OnTime$ vary across time. In order to estimate correctly our discrete-time hazard model and satisfy the proportionality assumption, the variables that do not fulfill the proportionality condition after the visual analysis have to be interacted with period dummies. This

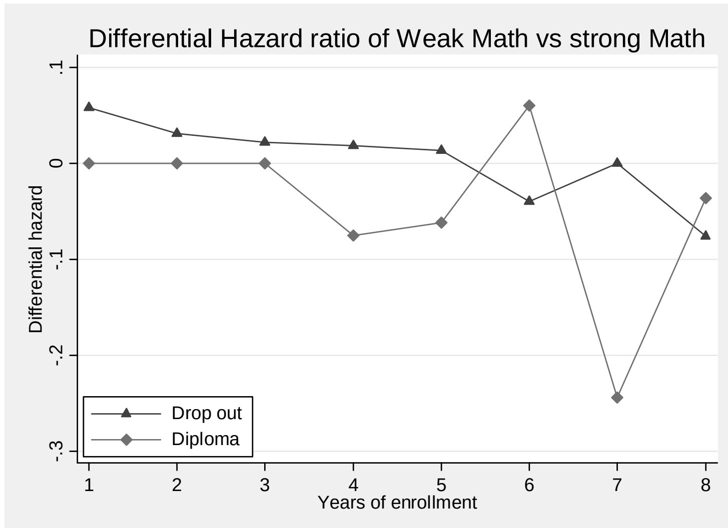

Figure 10: Comparing logit-hazard profiles between the two mathematical profiles

allows us to control for the change in their effect on hazard over time.

4.3 Estimating the full model

The results of the full model for the 4-year program population containing all the explanatory variables plus the interactions between the variables $Math$ and $OnTime$ with time are displayed in Table 6. We report the odds ratio and standard errors associated with the dropout outcome in the first two columns and those associated with degree in the two last columns. In this model, the way we interpret the coefficient for the pth variable and the kth outcome $\hat{\beta}_{kp}$ is the following: when a covariate X_q increases by one unit while holding everything else constant, the outcome-specific hazard ratio $\frac{h(k,t)}{h(0,t)}$ is multiplied by $exp(\hat{\beta}_{kp})$. The odds are reported in Table 6. From the previous section, we know that this effect is proportional during the complete academic path for all variables except for the variables $Math$ and $OnTime$ whose effects vary across time.

In the literature, there is a debate about the difference in success at university between foreign and Belgian students. Thanks to the competing risks model we partly clarify this debate given that there are significant differences that appear along the academic path between these two groups of students. Indeed, being a foreign student reduces the probability of getting all the way through the degree program, but it does not influence significantly the probability of dropping out. Thus, our competing risk setting shows that foreign students are more likely to experience consecutive enrollments without actually getting a degree. Indeed, Belgian students are enrolled, on average, for fewer periods than foreign students (3.53 years and 3.98 years respectively).

It is important to highlight that being a foreign student has a significant effect only on the probability of getting a degree. Other variables however, have a strong influence on both outcomes. Gender is a variable that influences both the probability of graduating and of dropping out. In Table 6 we see that female students have a significantly lower probability of dropping out of university (in the sample of students enrolled in a 4-year program). At the same time, female students have a significantly higher probability of graduating from university. Again, this last result is observable thanks to our competing risks setting. [12]

We notice that students enrolled in a scientific degree program (science or health science) have a higher probability of dropping out during their career at university than students in social sciences or humanities. But we can also mention that students enrolled in health sciences have a higher probability of getting a degree during their academic path than students in social sciences or humanities.

As far as the socioeconomic factors are concerned, we see that having a mother with a higher education degree reduces significantly the probability of dropping out and at the same time, it increases the probability of graduation. We believe that these results may be capturing two very different types of information about attainment at the ULB. On the one hand, having a mother with a higher education degree might influence the probability of getting a degree through the home environment or help with studying. On the other hand, the fact that it reduces significantly the probability of dropping out may be due to the fact that students from higher socioeconomic backgrounds can afford to stay enrolled at university for more years. This is somewhat supported by the result obtained for another socioeconomic control variable included in the model: the professional activity of the student's father. We see that having an unemployed father increases the probability of dropping out, showing that indeed financial constraints could largely determine whether a student drops out of university. In a system of free entry subsidized by publics funds, this raises the question of whether to a certain extent, students from more privileged socioeconomic backgrounds that can afford to stay longer at university are the ones who really benefit from such a system.

[12]Note that if we use a simple logit model where the dependent variable is getting a degree and students who dropout are treated as censored, the results yield the opposite effect for the variable gender, demonstrating that modeling incomplete information about student outcomes can lead to erroneous results.

Table 6: Estimation results from the discrete-time hazard model: multinomial logit (4-year program sample)

	OUTCOME			
	Dropout		Degree	
Variable	Odds	SE	Odds	SE
Generation 2001	1.20*	0.12	0.84	0.10
Gender (Female=1)	0.74**	0.07	1.58***	0.19
Not Belgian	0.81	0.15	0.53**	0.12
Science	1.53**	0.22	1.17	0.20
Health Sciences	3.34***	0.61	2.27**	0.76
Mother-Higher Education	0.83*	0.08	1.30**	0.16
Father-High level employee	0.89	0.10	1.21	0.16
Father-No profession/unemployed	1.36*	0.24	1.03	0.25
Traditionnel	0.55***	0.09	0.92	0.16
Weak Math X Period 1	1.83***	0.30		
Weak Math X Period 2	1.73**	0.36		
Weak Math X Period 3	2.61**	0.93		
Weak Math X Period 4	2.09*	0.80	0.74*	0.13
Weak Math X Period 5	1.48	0.76	0.80	0.18
Weak Math X Period 6	0.34*	0.22	1.19	0.42
Weak Math X Period 7			0.40*	0.22
Weak Math X Period 8	0.18	0.24	0.75	0.55
On Time X Period 1	0.68**	0.10		
On Time X Period 2	0.37***	0.07		
On Time X Period 3	0.46**	0.14		
On Time X Period 4	0.42**	0.14	1.79**	0.36
On Time X Period 5	0.76	0.37	1.13	0.27
On Time X Period 6	0.47	0.31	0.67	0.24
On Time X Period 7	4.75	5.37	1.17	0.70
On Time X Period 8	1.66	2.27	1.47	1.18
Time Dummies	Yes		Yes	
N	1196			

Our competing risk setting also reveals that the effect of attending a *traditionnel* school does not disappear after the first year: it significantly reduces the probability of dropping out at university throughout the entire academic path. However, contrary to the predictions of the cumulative hazard function analysis, students from the *traditionnel* do not have higher chances of getting a degree when we control for the complete set of explanatory variables.

With respect to the variable $WeakMath$, the interaction terms reveal that the impact of this variable is stronger in the early stages, as it increases significantly the probability of dropping out in the first four years of enrollment. In the same way, having a weak mathematical profile reduces the probability of timely degree completion, but the effect is much weaker. Completing high school late influences negatively the probability of dropping out in the first 4 years of enrollment, but then its influence becomes insignificant. Finally, we see that completing high school on time influences positively the probability of timely degree completion, but other than that, there is no significant effect of this variable on the degree outcome.

Note that interaction terms in a nonlinear setting must be interpreted with caution. Indeed, when using ordinary least squares (OLS), the marginal effect of the interaction is given by the double derivative of the regression with respect to each of the interacted variables. In the linear case, the result of this double derivative is the coefficient associated with the interacted explanatory variables. Ai and Norton (2003) show that in a nonlinear model, the double derivative yields a quantity that is completely different from the coefficient associated with the interaction term, disabling any natural interpretation of the estimated coefficient. However, this does not apply for our multinomial logit model given that we are not estimating the marginal effect of the variable $WeakMath$ (or the variable $OnTime$) in addition to the interaction. The only effect we estimate for this variable is the one that depends on a particular time period and these coefficients can be interpreted using only a single derivative.

5 Conclusion

We have analyzed factors that influence student behavior throughout the whole path at university. We applied the set of discrete-time methods for competing risks event history analysis described in Scott and Kennedy (2005) in order to identify the time at which events are more likely to occur. The model of student departure focuses on the characteristics of students and their socioeconomic background as determinants of dropout and timely graduation using a database of newly enrolled students at the ULB, one of the biggest universities in the Belgian French community.

The analysis of the hazard functions revealed that the result obtained by the student at the end of the first year at university is a very good predictor of the rest of the academic path. Indeed, students who pass the first year (whether they succeed in the first or second session of examinations) have a higher probability of completing and a lower probability of dropping out throughout the whole academic path. We also analyzed the cumulative hazard functions and found that the mean student survival time at university is 3 years in the 4-year degree program sample and around 4 years in the 5-year program sample. If 50% of the students experience an outcome before this threshold, then at least 50% of the students enrolled will leave the university without a degree.

As far as the competing risks model is concerned, we have reported several important results. Firstly, Belgian students have a higher probability of getting a degree than foreign students, but they do not have a different profile in terms of dropout. Secondly, having a mother that holds a higher education degree influences both outcomes: students are less likely to dropout and more likely to graduate. In addition, we have shown that both the effect of having a strong mathematical profile and finishing secondary schooling on time vary across time. Indeed, the impact of these variables on the probability of dropping out is stronger at early ages of enrollment. This could be due to either a selection effect (students with a weak mathematical profile drop out at the beginning, leaving a more homogenous group in subsequent years), or a learning effect (what a student studied during high school has less effect after spending several years at university). Finally, we have shown that certain student characteristics or socioeconomic factors can influence differently the probability of graduating or dropping out. Indeed, students with a *traditionnel* school background are less likely to drop out, but they are not more likely to graduate than students from *rénové* schools.

References

[1] Ai, C. & E. Norton (2003),"Interaction Terms in Logit and Probit Models", *Economics Letters*, 80, 123-129.

[2] Allison, P., D. (1982), "Discrete-time methods for the analysis of event histories", in S. Leinhardt (Ed.), Sociological methodology (61-98). San Francisco: Jossey-Bass.

[3] Arias Ortiz, E. & C. Dehon (2008), "What are the Factors that Influence Success at University? A Case Study in Belgium", *CESifo Economic Studies*, 54(**2**), 121-148.

[4] Arias Ortiz, E. & C. Dehon (2010), "High School Curriculum and Success at University: the Effect of Student Choices", *mimeo*.

[5] DesJardins, S., D. Ahlburg & B. McCall (2002), "A Temporal Investigation of Factors Related to Timely Degree Completion", The Journal of Higher Education, 73(**5**), 555-581.

[6] Jakobsen, V. & M. Rosholm (2003), "Dropping out of School? A Competing Risks Analysis of Young Immigrants' Progress in the Educational System", *IZA Discussion Paper series*, No. 918.

[7] Kleinbaum, D. & M. Klein (2005), *Survival analysis: a self-learning text*, Second edition, Springer, New York.

[8] Murtaugh, P., L. Burns & J. Schuster (1999), "Predicting the Retention of University Students", *Research in Higher Education*, 40(**3**), 355-371.

[9] Organization for Economic Cooperation and Development (2007), *Education at Glance: OECD indicators*, OECD, Paris.

[10] Scott, M. & B. Kennedy (2005), "Pitfalls in Pathways: Some Perspectives on Competing Risks Event History Analysis in Education Research", *Journal of Educational and Behavioral Statistics*, 30(**4**), 413-442.

[11] Singer, J. & J. Willett (1993), "It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events", Journal of Educational Statistics, 18(**2**), 155-195.

[12] Singer, J. & J. Willett (2003), "Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence", Oxford University Press, Oxford.

[13] Willett, J. & J. Singer (1991), "From Whether to When: New Methods for Studying Student Dropout and Teacher Attrition", *Review of Educational Research*, 61(**4**), 407-450.

APPENDIX

Table 7: Career paths at university classified by the minimum required years to graduate

4-year careers $(N_1 = 2973)$		
Fields	Number	Percentage
Philosophy	1343	45.17
Biomedical sciences	38	1.28
Sciences	330	11.10
Polt. Sc., sociology and Econ.	101	33.97
Physical Education	224	7.53
Other	28	0.94
5-year careers $(N_2 = 2073)$		
Fields	Number	Percentage
Engineering	441	21.27
Dentist/Pharmaceuticals	130	6.27
Law School	676	32.61
Business School	402	19.39
Psychology	424	20.45