Noname manuscript No.
(will be inserted by the editor)

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

Elena Arias Ortiz · Catherine Dehon

Abstract In this paper we study the factors that influence both dropout and degree completion (5 years to earn a degree) throughout the whole path 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) in order to identify the time at which events are more likely to occur. The model focuses on the characteristics of students and their socioeconomic background as determinants of dropout and timely graduation. Using the competing risks model, we show that 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. In general, in spite of low entry barriers, students coming from lower socioeconomic background are more vulnerable to dropout along the whole academic path because of financial constraints that prevent them to re-enroll. Also, our results reveal that after a failed year, a significant higher proportion of students who re-enroll in a different field obtain a degree than among those that re-enroll in the same field. Thus, to promote degree completion and decrease dropout, universities should increase the tools provided to students in order to manage failure. 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.

Keywords Competing risks · Degree Completion · Dropout · Event history · Higher Education · Survival Analysis

Elena Arias
Inter-American Development Bank, United States
E-mail: earias@ulb.ac.be

Catherine Dehon
Université libre de Bruxelles, ECARES, SBS-EM, 50, avenue F. Roosevelt, CP 114/04, 1050 Brussels
Tel.: +32-2-6503858
Fax: +32-2-6504012
E-mail: cdehon@ulb.ac.be

1 Objective

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)[1]. 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. The complete description of the samples and the degree programs are displayed in Table 1.

Table 1 Career paths at university classified by the minimum required years to graduate

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

Throughout the paper, we focus 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. The object of the present companion appendix is to analyze in detail the 5-year program sample and to present a detailed discussion of the cases in which students in this sample present particular or interesting behavior with respect to those in the 4-year program sample the results for the 5 year sample.

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

2 The 5 year sample and main descriptive statistics

The composition of the two samples is relatively similar (see Table 2). 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 traditional school enrolled in 5-year degree programs.

Table 2 Descriptive statistics of the 4 year program and the 5 year program samples

	Percentage	
	4 Year Sample	5 Year Sample
Generation 2001	49	52
Gender (Female=1)	55	50
Foreigner	20	15
Mother - Higher Education	55	58
Father - Low level employee	28	24
High level employee	63	67
Father -No profession/unemployed	10	9
Traditionnel	9	14
Weak Mathematical profile	67	44
On time	53	65
Reduced fee (first year)	35	32
N	2973	2073

ULB students take many different paths to get to graduation and 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 (Table 3). 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%

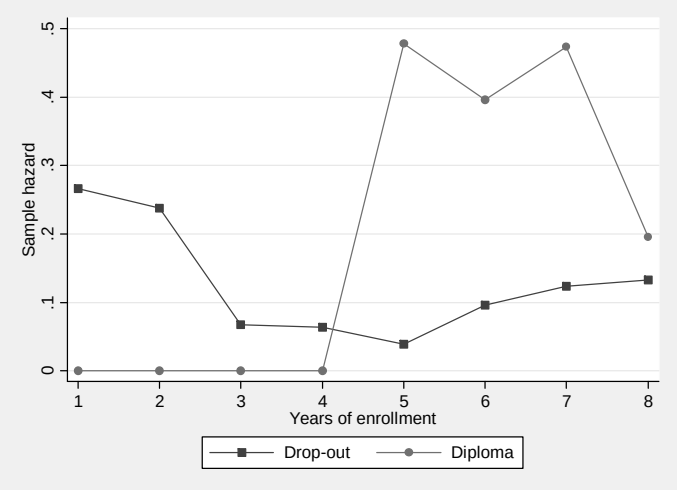

Fig. 1 Drop-Out and Degree hazard functions (5-year program sample)

3 Analyzing educational data using survival analysis in the 5 year sample

3.1 Hazard probabilities

As in the paper, we start by computing the hazard function for the 5 year sample and in both cases we observe a similar pattern than in the 4-year sample. The only difference is that the functions are less smooth because the number of students enrolled in the 5 year programs is smaller than in the 4 year programs.

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 more than 20% at the end of their second year, conditional

upon still being enrolled (see Figure 1). 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, as for the 4 year sample, 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, almost 50% of the students at risk obtain their degree on time (in this case 5 year), a higher rate than the one we observed for the 4 year sample (35%). Another difference we observe with respect to the 4 year sample is that the probability of graduating one year late, conditional upon surviving the first 5 years, is lower than that of graduating on time (40% against almost 50% among those at risk, respectively). The probability of graduating rises again after 7 years but at that point in time we have very few observations, and thus, the estimation becomes less precise.

The hazard functions also reveal that some sub-populations have the different behavior whether we consider 4-year programs or 5-year programs, as shown by the example in Figure 2 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.

3.2 Cumulative function

The cumulative function for each sample is displayed in Figure 3 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 figure 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 approx-

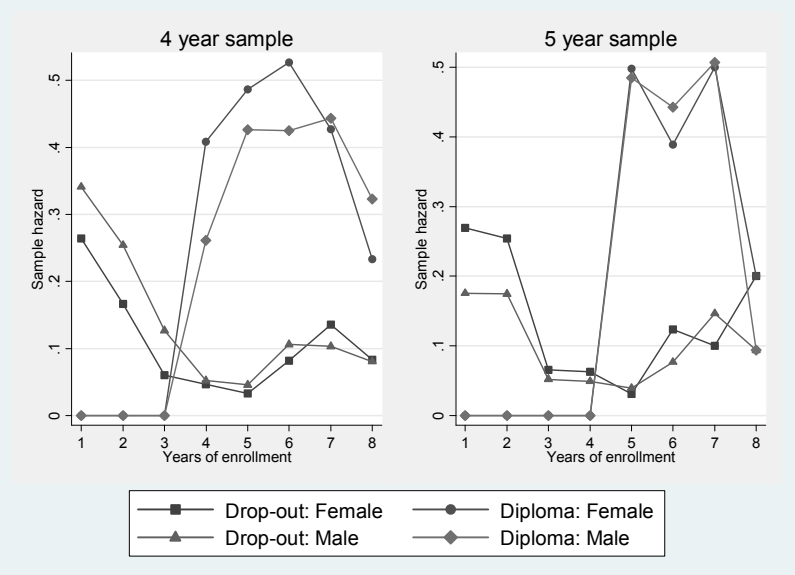

Fig. 2 Drop Out and Degree hazard functions by gender

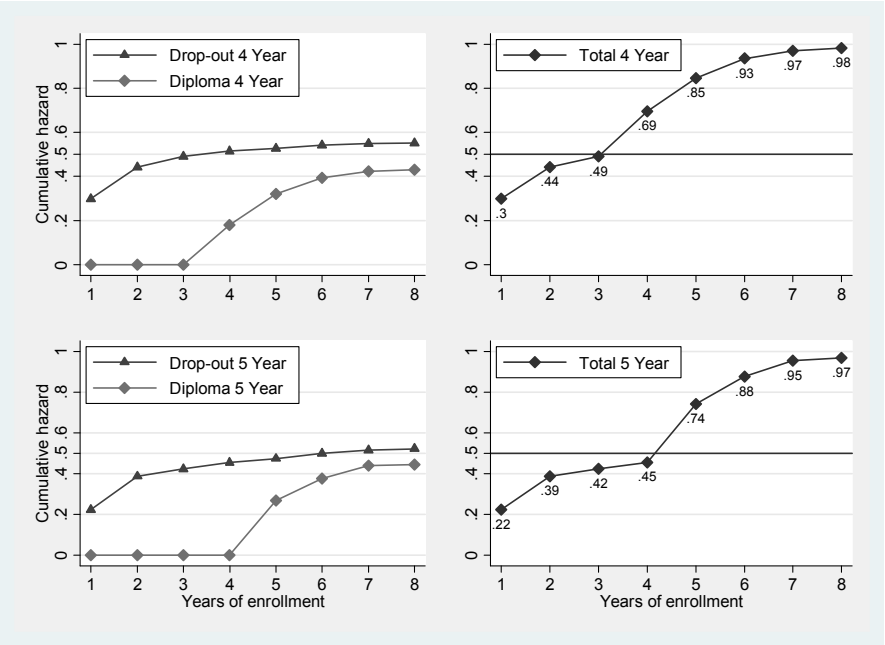

Fig. 3 Cumulative hazard functions

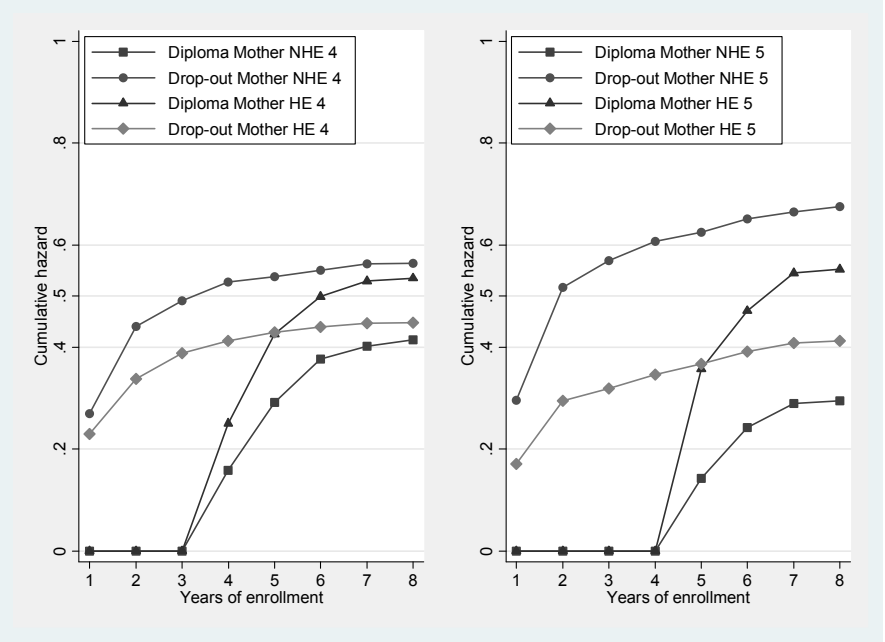

Fig. 4 Cumulative hazard functions by educational level of the mother

imately 3 years of enrollment for the 4-year program sample meaning that 50% of the students experience an outcome before that time and the other 50% will experience an outcome after the 3 years.[2] 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 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 4 shows that, unsurprisingly, the level of education of the mother has a strong impact on the cumulative hazard functions of both samples. Having a mother with a higher education degree decreases the cumulative probability of dropping out and increases that of getting a degree.

[2] In our case, we have to bear in mind that surviving means experiencing no outcome, so only still being enrolled.

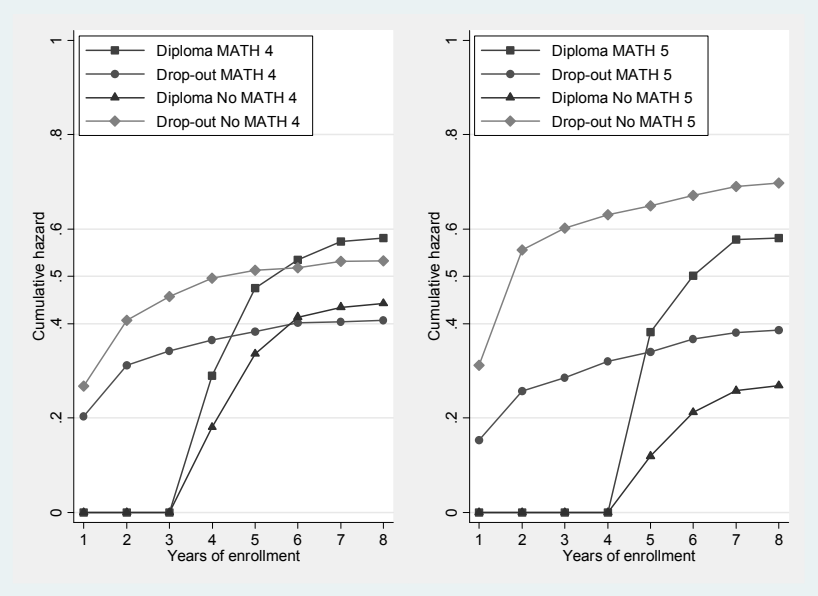

Fig. 5 Cumulative hazard functions by mathematical profile

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 5 . 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.

4 Competing risks: a multivariate analysis

As for the 4-year sample, 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.

The estimation results of the model where the hazards probabilities depend only on the intercepts and the time dummies are displayed in Table 4. 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 the model (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 below 0.10 from years 3 to 6. In the case of the graduation outcome, we cannot interpret the estimated time coefficients for periods 5 and 6 given that they are not significant.

Table 4 Results of fitting the baseline model for the 5-year sample

	Years of enrollment							
Event type	1	2	3	4	5	6	7	8
Coefficients								
Dropout	-1.253	-1.315	-2.784	-2.840	-2.576	-1.602	-1.079	-1.587
Diploma	-51.400	-50.432	-50.920	-50.930	0.036	-0.144	0.307	-1.587
Hazard ratio (x100)								
Dropout	28.56	26.85	6.18	5.84	7.61	20.15	33.99	20.45
Diploma	0.00	0.00	0.00	0.00	103.67	86.59	135.93	20.45
Outcome hazard (x100)								
Dropout	22.22	21.17	5.82	5.52	3.60	9.75	12.59	14.52
Diploma	0.00	0.00	0.00	0.00	49.07	41.88	50.36	14.52

Note: all coefficient are significant at 0.01 level except those for periods 5 and 6 of graduation who not significant

The results of the full model for the 5-year program population containing all the explanatory variables plus the interactions between the variables $Math$ with time[3] are displayed in Table 5. 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.

Several differences can be noted with respect to the estimation of the multivariate model using the 4-year sample. First, the difference in success at university between foreign and Belgian students do not hold in this sample. This could be explained by the fact that as we saw in the descriptive statistics, the share of foreign students in lower in the 5 year programs than in the 4 year programs. However, while the proportion of female students in both sample is similar, unlike the 4 year sample, gender does not influence significantly the probability of dropping out. At the same time, female students have almost 40% higher odds of graduating from university relative to staying enrolled.

[3] In the case of the 5-year sample, the variable Math was the only variable that did not fulfill the proportionality assumption.

Table 5 Estimation results from the discrete-time hazard model: multinomial logit (5-year sample)

	OUTCOME			
	Dropout		Diploma	
Variable	Odds	SE	Odds	SE
Generation 2001	0.98	0.1	0.7**	0.1
Science	0.4***	0.06	1.47**	0.24
Health Sciences	0.73	0.15	0.65	0.18
Gender (female==1)	1.02	0.11	1.39**	0.2
Not Belgian	0.92	0.17	0.69	0.18
Mother-Higher Education	0.69***	0.07	1.12	0.17
Father-High level employee	0.55***	0.06	1.63***	0.3
Father-No profession/unemployed	0.89	0.18	1.09	0.36
Reduced entry fee	1.04	0.12	0.82	0.14
Traditional	0.58***	0.09	1.46**	0.25
On Time	0.54***	0.06	1.67	30
Math X Period 1	1.52**	0.25		
Math X Period 2	2.58***	0.49		
Math X Period 3	2.02**	0.70		
Math X Period 4	1.07	0.40		
Math X Period 5	0.92	0.43	0.39***	0.08
Math X Period 6	0.65	0.29	0.92	0.26
Math X Period 7	1.03	0.60	0.44**	0.18
Math X Period 8	1.16	0.88	2.03	1.6
Time Dummies	Yes		Yes	
N	1062			

As far as the socioeconomic factors are concerned, the results are very similar as well. The only difference is that while we observe that having a mother with a higher education degree reduces significantly the probability of dropping out in both samples, in the case of the 5 year programs, it does not increases the probability of graduation. To some extent, the results in the case of the 5 year sample might be capturing more a role model effect: having a father who holds a high level position in the public or the private sector a strong and significant influence on the odds of graduating and also it reduces the probability of dropping out in a sample where most of the domains of study lead to independent business or very high rank positions: law, business, dentist and engineering. The results suggest that the university could not be enhancing "professional" mobility as much as expected.

Our competing risk setting also reveals that the effect of attending a high school system with a more rigid curriculum in terms of optional courses (*traditional* system) instead of a more flexible secondary school curriculum (*renovation* system) does not disappear after the first year (Arias and Dehon (2011)): it significantly reduces the probability of dropping out at university throughout the entire academic path and in the case of the 5 year sample, it also increases the chances of getting a degree. 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 three years of enrollment. In the same way, having a weak mathematical profile

reduces by 60% the probability of timely degree completion relative to staying enrolled and the effect is much more stronger than for the 4 year sample. Completing high school late influences negatively the probability of dropping out , but then its influence becomes insignificant. Finally, we see that there is no significant effect of this variable on the degree completion outcome.

5 Conclusion

All the methodology described in the main paper stay valid for other samples and can be applied in a straightforward way. Concerning the interpretations, there exist some small differences between the 4-year program and the 5-year program samples but the main conclusions stays the same.