A COMPETING RISK ANALYSIS OF ACADEMIC CAREERS WITH STUDENTS' ABILITY AND SPEED AS PREDICTORS

Michela Battauz [1]

[1] Department of Economics and Statistics, University of Udine, (e-mail: michela.battauz@uniud.it

ABSTRACT: A competing risk model in discrete time is employed to analyze the outcomes of students' academic careers, which are degree attainment, drop out or transfer to another course. As covariates, besides using the variables available from the administrative database, we consider also the performance of the students in terms of ability and speed, which are predicted from an IRT model applied to the grades obtained in the exams. An application shows that these variables are good predictors of the outcomes.

KEYWORDS: academic performance, latent variables, survival analysis.

1 Introduction

Competing risks models in discrete time (Tutz & Schmid, 2016) are particularly suitable for the analysis of the students' careers at university since they consider all the possible events that can occur in time (see Scott & Kennedy, 2005, Clerici *et al.* , 2015). Such events are degree attainment, drop out or transfer to another course. The novelty of our proposal is given by the predictors included in the model. In fact, we use an Item Response Theory (IRT) model for the grades obtained by the university students and the time needed to pass the exams that accounts for two sources of censoring: dropout and lack of grades for non-passed exams. Using this model we predict two latent variables for each student, which can be interpreted as ability and speed. More details on this model can be found in Battauz (2023). They are then used as predictors in the competing risk model together with other observed covariates.

2 Models and Methods

To extend the analysis of time-to-event data from the case of one possible event, as usually done in survival analysis, to the case of multiple events, it is necessary to define a hazard function for each target event

$$\lambda_r(t|\mathbf{x}) = P(T = t, R = r|T \geq t, \mathbf{x}), \tag{1}$$

where $R \in 1,\dots,m$ denotes the event, $T \in 1,\dots,t_{max}$ the time, and $\mathbf{x}$ a set of covariates. It is possible to show that the survival function is given by

$$S(t|\mathbf{x}) = P(T \geq t|\mathbf{x}) = \prod_{i=1}^{t}(1-\lambda(i|\mathbf{x})) \tag{2}$$

and the event probability results

$$P(T = t, R = r|\mathbf{x}) = \lambda_r(t|\mathbf{x})S(t-1|\mathbf{x}). \tag{3}$$

In discrete time, the hazard function is frequently modelled using the multinomial model

$$\lambda_r(t|\mathbf{x}) = \frac{\exp(\beta_{0tr} + \mathbf{x}^\top \beta_r)}{1+\sum_{i=1}^{m}\exp(\beta_{0ti} + \mathbf{x}^\top \beta_i)}. \tag{4}$$

Once the parameters have been estimated, it is possible to compute the event probabilities for a vector of covariate values $\mathbf{x}$ by means of Equations (2) and (3). The estimation was performed by maximum likelihood.

3 Application

The model was applied to the 2017 cohort of students enrolled in Business or Economics at the University of Udine, composed of 353 people at baseline. These two bachelor's degrees share many courses, especially in the first and second year, thus permitting us to fit the IRT model with censoring to the grades obtained by these students together. From the IRT model we predict two latent variables for each student: the ability (θ) and the speed (τ). The predicted values are then standardized and used as covariates in the competing risk model together with other variables available from the administrative database, which are age at enrolment, grade obtained from high school, type of high school, gender and residence. Table 1 reports the estimates of the coefficients of the variables selected on the basis of statistical significance. Hence, the only variable still significant when θ and τ are included in the model is age, with younger students having higher probabilities of attaining the degree and lower probabilities of dropping out. Both ability and speed have a positive effect on the probability of attaining the degree and also a positive joint effect. Figure 1 shows the cumulative predicted probabilities for different values of θ and τ. The age was fixed at 19, the most common value. It is apparent the important effect that these variables have on the outcome. Students with moderately high values of both ability and speed present a high probability of attaining the degree. Such probability results definitely lower for students with low values of ability and moderately high speed, while having low values of speed and moderately high ability affects the time of degree attainment with a less severe impact on the probability of obtaining this outcome. Finally, the case of low levels on both the latent variables determines very low probabilities of attaining the degree.

Table 1. *Estimates of the coefficients of the model (coef.) and their standard errors (s.e.).*

	Degree		Drop Out		Transfer	
Variable	coef.	s.e.	coef.	s.e.	coef.	s.e.
time 1	-26.13	0.00	-1.79	0.21	-1.28	0.20
time 2	-9.93	1.45	-3.91	0.38	-3.48	0.41
time 3	-0.35	0.24	-3.37	0.39	-4.74	1.02
time 4	0.71	0.34	-2.94	0.43	-13.01	86.69
time 5	-0.80	0.74	-1.30	0.42	-13.00	133.39
θ	1.87	0.32	-1.77	0.24	-0.52	0.23
τ	1.84	0.27	-0.63	0.19	-0.27	0.23
age at enrollment - 19	-0.20	0.12	0.06	0.03	-0.00	0.04
$\theta \times \tau$	0.55	0.39	-0.62	0.20	-0.15	0.22

4 Conclusion and ongoing work

This paper shows that students' ability and speed, measured through the grades obtained in the exams, play an important role in their career. These two variables are predicted using an IRT model with censoring, and hence the predicted values can be considered as measures with error of the latent variables. Since error-in-variables introduces bias in parameter estimation, in future research, we aim at jointly modelling the grades and the events of students' career so that the measurement error is properly taken into account in the model.

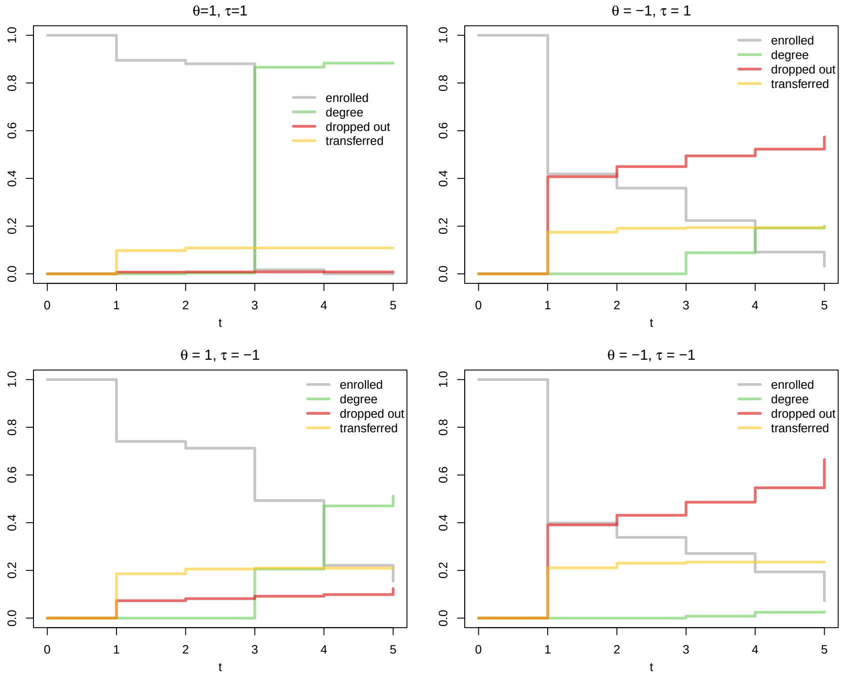

Figure 1. *Cumulative event probabilities obtained from the model for different values of* θ *and* τ*.*

References

BATTAUZ, M. 2023. Analysis of university grades: An IRT model for responses and response times with censoring. *Submitted to the Conference SIS 2023 - Statistical Learning, Sustainability and Impact Evaluation*.

CLERICI, R., GIRALDO, A., & MEGGIOLARO, S. 2015. The determinants of academic outcomes in a competing risks approach: Evidence from Italy. *Studies in Higher Education*, **40**, 1535–1549.

SCOTT, M. A., & KENNEDY, B. B. 2005. Pitfalls in Pathways: Some Perspectives on Competing Risks Event History Analysis in Education Research. *Journal of Educational and Behavioral Statistics*, **30**, 413–442.

TUTZ, G., & SCHMID, M. 2016. *Modeling Discrete Time-to-Event Data*. New York: Springer.