Bayesian survival modelling of university outcomes

Catalina A. Vallejos, Mark F.J. Steel [1]
Department of Statistics, University of Warwick

Summary

Dropouts and delayed graduations are critical issues in Higher Education Systems worldwide. A key task in this context is to identify risk factors associated to these events, providing potential targets for mitigating policies. For this purpose, we employ a discrete time competing risks survival model, dealing simultaneously with university outcomes and its associated temporal component. We define survival times as the duration of the student's enrolment at university and possible outcomes as graduation or two types of dropout (voluntary and involuntary), exploring the information recorded at admission time (e.g. educational level of the parents) as potential predictors. While similar strategies have been previously implemented, we extend the previous methods by handling covariate selection within a Bayesian variable selection framework, where model uncertainty is formally addressed through Bayesian Model Averaging. Our methodology is general, however here we focus on undergraduate students enrolled at three selected degree programmes of the Pontificia Universidad Católica de Chile during the period 2000-2011. Our analysis reveals interesting insights, highlighting the main covariates that influence students' risk of dropout and delayed graduation.

Keywords: University dropout; Delayed graduation; Bayesian model averaging; Competing risks; Proportional Odds model

1 Introduction

During the last several decades, the higher education system has seen substantial growth in Chile, evolving from around 165,000 students in the early 1980's to over 1 million enrolled in 2012 (see http://www.mineduc.cl/). Nowadays, the access to higher education is no longer restricted to an elite group. Among other reasons, this is due to a bigger role for education as a tool for social mobility, the opening of new institutions and a more accessible system of student loans and scholarships. However, currently, more than half of the students enrolled at Chilean higher education institutions do not complete their degree. This includes students expelled for academic or disciplinary reasons and those who voluntarily withdrew (dropout that is not instigated by the university but is also not necessarily the student's decision; e.g. forced by financial hardship). Another issue is the high proportion of late graduations, where obtaining the degree requires more

[1]Corresponding author: Mark Steel, Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK; email: M.F.Steel@stats.warwick.ac.uk. During this research, Catalina Vallejos acknowledges funding from the University of Warwick and the Pontificia Universidad Católica de Chile (PUC). We thank two referees and the Associate Editor for very insightful and stimulating comments and are grateful to the PUC for access to the dataset analyzed in this article. We also thank Lorena Correa and Guillermo Marshall for motivating the analysis and Valeria Leiva for support in accessing the data.

time than the nominal duration of the programme (Chilean universities allow more flexibility than the e.g. British educational system, so students can repeat failed modules and/or have a reduced academic load in some semesters). Dropout and delays in graduation involve a waste of time and resources from the perspective of the students, their families, universities and the society. While the focus here is on Chile, these issues are also critical in many other parts of the world.

There is a large literature devoted to university dropout. It includes conceptual models based on psychological, economic and sociological theories (e.g. Tinto, 1975; Bean, 1980). Instead, we focus on empirical models. Previous research often considered the dropout as a dichotomous problem, neglecting the temporal component and focusing on whether or not a student has dropped out at a given time. Nonetheless, ignoring when the dropout occurs is a serious waste of information (Willett and Singer, 1991) — potential high risk periods will not be identified and no distinction between early and late dropout will be made. An alternative is to use (standard) survival models for the time to dropout (e.g. Murtaugh et al., 1999), labelling graduations as censored observations. However, this contradicts the idea of censoring because dropout is a possibility only whilst the student is enrolled at university (those who graduate will never dropout). Instead, graduation must be considered as a competing event and incorporated into the model. The drawbacks of treating competing risks events as if they were censored observations are discussed in Andersen et al. (2012).

We wish to identify risk factors associated with dropouts and delayed graduations, focusing on undergraduate students of the Pontificia Universidad Católica de Chile (PUC). This is one of the most prestigious universities in Chile (and the best university in Latin America, according to QS Ranking 2014, see http://www.topuniversities.com/). Despite having one of the lowest dropout rates in the county (far below the national level), dropout is still an important issue for some degrees of the PUC. Similarly, while delays in graduation are rarely observed in some degree programs (e.g. Medicine), some others are particularly affected by this issue. Therefore, it is critical for university authorities to identify risk factors associated with these events as this might inspire policies mitigating late graduations and dropouts.

A competing risks model is proposed for the length of stay at university, where the possible events are: graduation, voluntary dropout and involuntary dropout. These are defined as the final academic situation recorded by the university at the end of 2011 (students that have not experienced any of these events by then are recorded as right-censored observations and censoring is assumed to be non-informative). Survival times are defined as the length of stay at university, measured in semesters from admission (which is the frequency at which the university updates its records), producing survival times that are inherently discrete. It is an advantage of this approach that it deals jointly with graduations and dropouts. We explore the information recorded at admission time (e.g. sex, educational level of the parents) as potential risks factors associated with these outcomes. While similar strategies have been previously implemented, we extend the previous methods by handling covariate selection within a Bayesian framework, where model uncertainty is formally addressed through Bayesian Model Averaging (BMA). Additionally, we illustrate how a Bayesian scheme is particularly useful here, where the nature of the data precludes maximum likelihood inference for the model that is typically adopted in this setting.

The construction and the main features of the PUC dataset are summarized in Section 2, showing high heterogeneity (in terms of academic outcomes and the student population characteristics) among programmes. Section 3 introduces a competing risks model for university outcomes, which can be estimated via a multinomial logistic regression. Section 4 proposes a suitable prior structure and introduces a Markov chain Monte Carlo (MCMC) algorithm, exploiting a hierarchical representation of the multinomial logistic likelihood (based on Holmes and Held, 2006; Polson et al., 2013). We also propose BMA as a tool for detecting risk factors associated with dropouts and delayed graduations, exploring all covariates configurations, directly accounting for the uncertainty linked to the covariate selection. Empirical results are summarised in Section 5, focusing on three science programmes that are severely affected by dropout and late graduations. Our analysis reveals interesting insights and highlights the relevant covariates that characterise students with higher risk of dropout and delayed graduation, providing potential candidates for mitigating policies. Finally, Section 6 concludes.

2 The PUC dataset

The PUC provided anonymised information about 34,543 undergraduate students enrolled during the period 2000-2011 via the ordinary admission process (based on high school marks and an standardised selection test, applied at a national level). We only analyse the degree programmes that existed during the entire period and students who: *(i)* were enrolled for at least 1 semester (the dropout produced right after enrolment might have a different nature), *(ii)* were enrolled in a single programme (students doing parallel degrees usually need more time to graduate and have less risk of dropout), *(iii)* did not have validated previously passed modules from other degree programmes (which could reduce the time to graduation), *(iv)* were alive by the end of 2011 (0.1% of the students had died by then) and *(v)* had full covariate information. Overall, 78.7% of the students satisfied these criteria. The Supplementary Material breaks this number down by program. Throughout, we will only consider this subset of the data, pertaining to 27,189 students.

By the end of 2011, 41.9% of the students were still enrolled (right censored), 37.2% had graduated, 6.6% were expelled (involuntary dropout, mostly related to poor academic performances), 10.7% withdrew (voluntary dropout), and 3.7% abandoned the university without an official withdrawal (after not being enrolled for any modules for at least 2 consecutive semesters). Following university policy, the latter group is classified as voluntary dropout. The high percentage of censoring mostly relates to students from later years of entry, who were not yet able to graduate by the end of 2011. The performance of students is not homogenous across programmes (Figure 1). To illustrate, the following statistics refer to the cohorts starting in 2000-2004 (avoiding later cohorts where a large number of students are right censored). In terms of total dropout, Medicine (4.6%) and Chemistry (71.2%) have the lowest and highest rates, respectively. The highest rates of involuntary and voluntary dropout are for Chemistry (28.8%) and Astronomy (54.5%), respectively. Dropouts are mostly observed during the first semesters of enrolment. In contrast, graduation times are concentrated on large values, typically above the official length of the programme (which varies between 8 and 14 semesters, with a typical value of 10 semesters). As shown in Figure 1, programmes also exhibit strong heterogeneity in terms of timely graduation, the proportion of which varies from 88% (Medicine) to 11% (Education Elementary School).

Table 1: Information recorded at admission time. Options for categorical variables in parentheses

Demographic factors
Sex (female, male)
Region of residence (Metropolitan area, others)
Socioeconomic factors
Parents education (at least one with a technical or university degree, no degrees)
High school type (private, subsidized private, public)
Funding (scholarship and loan, loan only, scholarship only, none)
Admission-related factors
Selection score
Application preference (first, others)
Gap between high school graduation and admission to PUC (1 year or more, none)

Demographic, socioeconomic and variables related to the admission process are recorded (see Table 1). For these covariates, substantial differences are observed between programmes (see Supplementary Material, Section A). In terms of demographic factors, some degrees have a very high percentage of female students (e.g. all education-related programmes) while e.g. most of the Engineering students are male. The proportion of students who live outside the Metropolitan area is more stable across programmes (a particularly high percentage is observed in the Education for Elementary School degree taught in the Villarrica campus, which is located in the south of

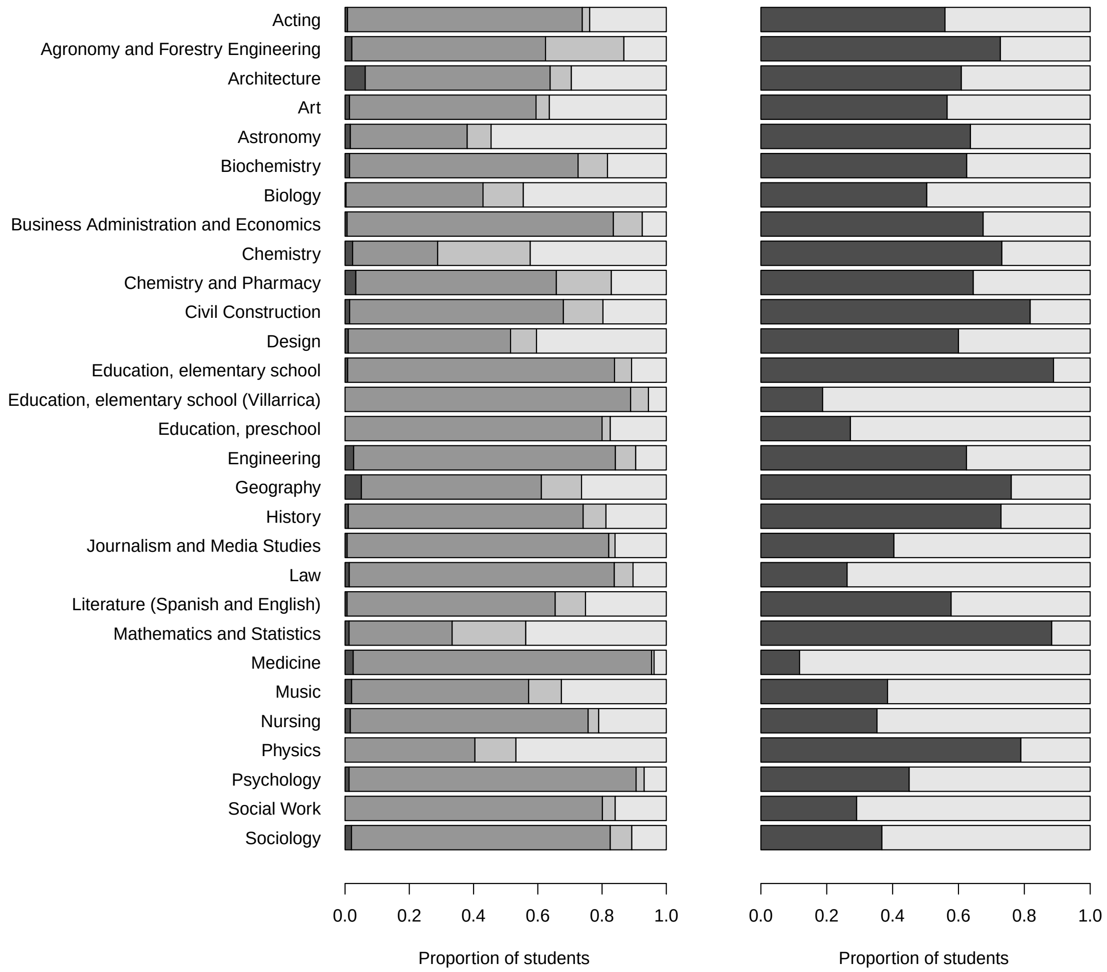

Figure 1: Cohorts 2000-2004. Left: distribution of students according to final academic situation. From darkest to lightest, shaded areas represent the proportion of censored observations, graduation, involuntary dropout and voluntary dropout, respectively. Right: among graduated students, proportion of students graduated within the nominal duration of the programme (lighter area).

Chile). Strong differences are also detected for the socioeconomic characteristics of the students. Chilean schools are classified according to their funding system as public (fully funded by the government), subsidised private (the state covers part of the tuition fees) and private (no funding aid). This classification can be used as a proxy for the socioeconomic situation of the student (low, middle and upper class, respectively). The educational level of the parents is usually a good indicator of socioeconomic status as well. Some degrees have a very low percentage of students that graduated from public schools (e.g. Business Administration and Economics) and others have a high percentage of students whose parents do not have a higher degree (e.g. Education for Elementary School in Villarrica). In addition, a few programmes have low rates of students with a scholarship or student loan (e.g. Business Administration and Economics). Finally, "top" programmes (e.g. Medicine, Engineering) only admit students with the highest selection scores. For instance, in 2011, the lowest selection score in Arts was 604 but Medicine did not enrol any students with a score below 787. In the same spirit, these highly selective programmes only enrolled students that applied to it as a first preference.

The substantial heterogeneity (in terms of outcomes and covariates) precludes meaningful modelling across programmes. Thus, the analysis will be done separately for each degree.

3 Discrete time competing risks models

Standard survival models only allow for a unique event of interest, often recording occurrences of alternative events as censored observations. In the context of university outcomes, graduations have been treated as censored observations when the event of interest is dropout (e.g. Murtaugh et al., 1999). However, those students who graduated are obviously no longer at risk of dropout (from the same degree, within the same enrolment). Instead, competing risks models are more appropriate when multiple event types can occur and there is a reason to believe they are the result of different mechanisms. These models simultaneously incorporate both the survival times and the event types (or "causes"). Most of the previous literature focuses on continuous survival times (e.g. Crowder, 2001). Instead, in the context of university outcomes (where survival times are usually measured in numbers of academic terms), a discrete time approach is more appropriate. In a discrete-time competing risks setting, the variable of interest is (R, T), where $R \in \{1, \ldots, \mathcal{R}\}$ denotes the type of the observed event and $T \in \{1, 2, \ldots\}$ is the survival time. Analogously to the single-event case, a model can be specified via the *cause-specific* cumulative incidence or hazard functions, defined respectively as

$$F(r, t) = P(R = r, T \leq t) \text{ and } h(r, t) = \frac{P(R = r, T = t)}{P(T \geq t)}. \tag{1}$$

In turn, these quantities represent the proportion of subjects for which an event type r has been observed by time t and the conditional probability of observing an event of type r at period t given that no event occurred before. Like the Kaplan-Meier estimator in the discrete case, the non-parametric maximum likelihood estimator of $h(r, t)$ is the ratio between the number of events of type r observed at time t and the total number of subjects at risk at time t (Crowder, 2001). An alternative approach would be to define a competing risks model based on the distribution of the latent survival times $T_1, \ldots, T_{\mathcal{R}}$ (with $T = \min\{T_1, \ldots, T_{\mathcal{R}}\}$). Nonetheless, as shown in Tsiatis

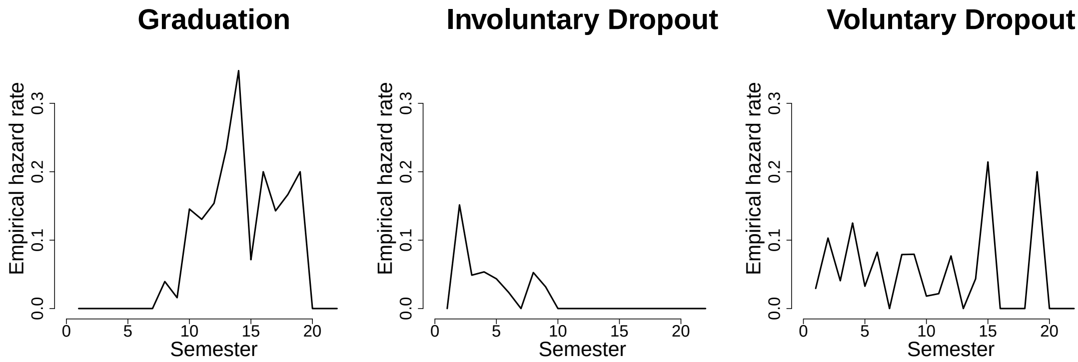

Figure 2: Non-parametric estimates of cause-specific hazard rates for Chemistry students (cohorts 2000-2004).

(1975), this approach leads to an unidentifiability problem as the observed data itself does not allow us to distinguish between independent and dependent latent survival times.

Sometimes, a simple (cause-specific) parametric model can be adopted. However, such models are not suitable for the PUC dataset, where the cause-specific hazard rates have a rather erratic behaviour over time (Figure 2 illustrates this for Chemistry students). In particular, no graduations are observed during the first semesters of enrolment, inducing a zero graduation hazard at those times. Graduations only start about a year before the official duration of the programme (10 semesters). For this programme, the highest risk of being expelled from university is at the end of the second semester. In addition, during the first years of enrolment, the hazard of voluntary dropout has spikes located at the end of each academic year (even semesters). Therefore, more flexible models are required in order to accommodate these hazard trajectories.

3.1 Proportional Odds model for competing risks data

Cox (1972) proposed a Proportional Odds (PO) model for discrete times and a single event type. It is a discrete variation of the well-known Cox Proportional Hazard model, proposed in the same seminal paper. Let $x_i \in \mathbb{R}^k$ be a row vector containing the covariate values for individual i and $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbb{R}^k$ a vector of regression parameters. The Cox PO model is given by

$$\log\left(\frac{h(t|\delta_t, \beta; x_i)}{1 - h(t|\delta_t, \beta; x_i)}\right) = \delta_t + x_i'\beta, \quad i = 1, \ldots, n, \tag{2}$$

where $\{\delta_1, \delta_2, \ldots\}$ denote the baseline log-odds (with respect to no event) at times $t = 1, 2, \ldots$. The model in (2) can be estimated via a binary logistic regression with *time-specific* intercepts as $\text{logit}\,[h(t|\delta_t, \beta; x_i)] = \sum_{s=1}^{\infty} \delta_s D_{is}(t) + x_i'\beta$, with $D_{is}(t) = 1$ if $s = t$ and 0 otherwise (Singer and

Willett, 1993). In fact, in (2), the likelihood contribution of an individual i is

$$L_i = \left[\frac{h(t_i|\delta_{t_i}, \beta; x_i)}{1 - h(t_i|\delta_{t_i}, \beta; x_i)}\right]^{1-c_i} \prod_{s=1}^{t_i}[1 - h(s|\delta_s, \beta; x_i)], \tag{3}$$

with $c_i = 1$ if the survival time is censored, 0 otherwise. This coincides with the likelihood associated to independent Bernoulli trials, with $Y_{it} = 1$ if the event is observed at time t for individual i and 0 otherwise ($t \leq t_i$) and $P(Y_{it} = y_{it}) = [h(t|\delta_t, \beta; x_i)]^{y_{it}} [1 - h(t|\delta_t, \beta; x_i)]^{1-y_{it}}$.

In the competing risks case, with $\mathcal{R}$ possible events, define δ_{rt} as the baseline log-odd of observing the event r (with respect to no event) at time t and let $\beta_{(r)}$ ($\in \mathbb{R}^k$) be the associated vector of regression parameters. Define $B = \left\{\beta_{(1)}, \ldots, \beta_{(\mathcal{R})}\right\}$ and $\delta = \{\delta_{11}, \ldots, \delta_{\mathcal{R}1}, \delta_{12}, \ldots, \delta_{\mathcal{R}2}, \ldots\}$, the model in (2) can then be extended to accommodate the $\mathcal{R}$ competing events as

$$\log\left(\frac{h(r, t|\delta, B; x_i)}{h(0, t|\delta, B; x_i)}\right) = \delta_{rt} + x_i'\beta_{(r)}, \quad r = 1, \ldots, \mathcal{R}; i = 1, \ldots, n, \tag{4}$$

$$\text{where } h(0, t|\delta, B; x_i) = 1 - \sum_{r=1}^{\mathcal{R}} h(r, t|\delta, B; x_i) \tag{5}$$

is the hazard of no event being observed at time t ($t = 1, 2, \ldots$). The latter is equivalent to

$$h(r, t|\delta, B; x_i) = \frac{e^{\delta_{rt} + x_i'\beta_{(r)}}}{1 + \sum_{s=1}^{\mathcal{R}} e^{\delta_{st} + x_i'\beta_{(s)}}}. \tag{6}$$

Here, $\{h(r, t|\delta, B; x_i)\}$'s represent *cause-specific hazard rates*, defined as the instant probability of observing an event of type r given that no event (type r or any other) has been observed before. Unless we assume independence between the cause-specific survival processes, the latter does not necessarily coincide with the *marginal hazard rate*, i.e. the instant probability of observing an event of type r given that an event of type r has not happened before (Crowder, 1996).

Similar to (2), the latter model can be estimated by means of a multinomial logistic regression where the δ_{rt}'s are estimated as binary time- and cause-specific intercepts. The latter notation implies that the same predictors are used for each cause-specific component (but this can be generalised). In (4), covariates influence both the marginal probability of the event $P(R = r)$ and the rate at which the event occurs. Positive values of the cause-specific coefficients indicate that (at any time point) the hazard of the corresponding event increases with the associated covariate values and the effect of covariates on log odds is constant over time. In the context of the PUC analysis, if a covariate has a negative coefficient related to graduations, that would constitute a risk factor associated to delayed graduations (as the corresponding students will tend to graduate less and more slowly). Similarly, positive values of the coefficients associated to voluntary and involuntary dropout will characterise risk factors linked to these events.

For university outcomes, (4) has been used by Scott and Kennedy (2005), Arias Ortis and Dehon (2011) and Clerici et al. (2014), among others. Nonetheless, its use has some drawbacks. Firstly, it involves a large number of parameters (if $\mathcal{T}$ is the largest recorded time, there are $\mathcal{R} \times \mathcal{T}$ different δ_{rt}'s). Scott and Kennedy (2005) overcome this by assigning a unique cause-specific log-odds δ_{rt_0} to the period $[t_0, \infty)$ (for fixed t_0). The choice of t_0 is rather arbitrary but it is reasonable to choose t_0 such that most individuals already experienced one of the events (or censoring) by time t_0. Secondly, maximum likelihood inference on (4) is problematic in our context because the range of the survival times associated with graduations and dropouts do not have a large overlap (while no graduations can be observed during the first semesters of enrollment, only few dropouts are observed at later times). The multinomial logistic literature refers to this as (quasi) complete separation of the outcomes with respect to the predictors, i.e. some outcomes are not (or rarely) observed for particular covariate configurations (Albert and Anderson, 1984). In other words, the predictors can (almost) perfectly predict the outcomes. In (4), these predictors include binary variables linked to the δ_{rt}'s. Therefore, (quasi) complete separation occurs if the event types are (almost) entirely defined by the survival times. For example, in our context, no graduations can be observed during the second semester of enrollment. Therefore, the likelihood function will be maximized when the cause-specific hazard related to graduations (defined in (6)) is equal to zero at time $t = 2$. Thus, from equation (4) it follows that the maximum likelihood estimate for the corresponding cause-specific log-odds parameter δ_{r2} would be equal to $-\infty$. This problem will typically arise when dealing with educational data, although it might diminish in shorter programmes (e.g. 2 years college programmes).

Singer and Willett (2003) use polynomial baseline odds to overcome the separation issue. This option is less flexible than (4), and its use is only attractive when a low-degree polynomial can adequately represent the baseline hazard odds. This is not the case for the PUC dataset, where cause-specific hazard rates have a rather complicated behaviour (see Figure 2) and not even high-order polynomials provided a good fit (not shown). Another alternative is to parametrize the model in terms of $\delta^*_{rt} = e^{\delta_{rt}}$ and to set $\delta^*_{rt} = 0$ in those periods where the separation occurs. Nonetheless, this requires non trivial prior elicitation as the periods affected by the separation issue are not completely defined by university regulations. For example, students can graduate at any time as long as they approve all graduation requirements. In general, we would not expect this to happen before the official duration of each programme. However, students are allowed to take extra credits each semester, possibly leading to early graduations. Therefore, while we could fix the hazards for the 1st (and perhaps the 2nd) semester of studies, later periods can not be fixed a priori and the separation will remain an issue. Moreover, this strategy avoids the use of estimation algorithms that are based on a multinomial logistic scheme, requiring the development of tailored estimation procedures. In contrast, as described in Section 4.2, posterior inference for the model in (4) under the prior described below can be implemented as a simple extension of available procedures, e.g. Polson et al. (2013).

Here, the model in (4) is adopted for the analysis of the PUC dataset, using Bayesian methods to handle the separation issue. As in Scott and Kennedy (2005), we define the last period as $[t_0, \infty)$, using $t_0 = 16$ semesters (after which few students were still enrolled). For identifiability, we reparametrize the log-odds parameters δ_{rt}'s, defining the δ_{r1}'s (first semester) as overall cause-specific intercepts (the corresponding binary variables are equal to 1 for all periods) and interpreting the remaining δ_{rt}'s as cause-specific log-odds changes (w.r.t. the first semester).

4 Bayesian Proportional Odds competing risks regression

4.1 Prior specification

An alternative solution to the separation issue lies in the Bayesian paradigm, allowing the extraction of information from the data via an appropriate prior distribution for the log-odds parameters δ_{rt}'s (Gelman et al., 2008). The Jeffreys prior can be used for this purpose (Firth, 1993). This is attractive when reliable prior information is absent. In a binary logistic case, the Jeffreys prior is proper and its marginals are symmetric with respect to the origin (Ibrahim and Laud, 1991). These properties have no easy generalisation for the multinomial case, where an expression for the Jeffreys prior is very complicated (Poirier, 1994). Instead, Gelman et al. (2008) suggested weakly informative independent Cauchy priors (with scale equal to 2.5) for a re-scaled version of the regression coefficients. When the outcome is binary, these Cauchy (and any Student t) priors are symmetric like the Jeffreys prior but produce fatter tails (Chen et al., 2008). The prior in Gelman et al. (2008) assumes that the regression coefficients fall within a restricted range. For the model in (4), it shrinks δ_{rt}'s estimates away from $-\infty$ (and ∞), i.e. assigning small probability to near-zero cause-specific hazard rates at all periods. Such a prior is convenient if the separation of the outcomes relates to a small sample size (where increasing the sample size will eventually eliminate this issue). This is not the case for the PUC dataset, or other typical data on four-year-university outcomes, where the separation arises from structural restrictions (e.g. it is not possible to graduate during the first periods of enrolment). Hence, we expect δ_{rt}'s to have a large negative value in those periods where event r is very unlikely to be observed (inducing a nearly zero cause-specific hazard rate). Defining $\delta_{(r)} = (\delta_{r1}, \dots, \delta_{rt_0})'$, we suggest the prior

$$\delta_{(r)} \sim \text{Cauchy}_{t_0}(\mathbf{0}_{t_0}, \omega^2 I_{t_0}), \quad r = 1, \dots, \mathcal{R}, \tag{7}$$

where I_{t_0} denotes the identity matrix of dimension t_0 and $\mathbf{0}_{t_0}$ is a vector of t_0 zeros. Equivalently,

$$\pi(\delta_{(r)}|\Lambda_r) \sim \text{Normal}_{t_0}(\mathbf{0}_{t_0}, \Lambda_r^{-1}\omega^2 I_{t_0}), \quad \Lambda_r \sim \text{Gamma}(1/2, 1/2), \quad r = 1, \dots, \mathcal{R}. \tag{8}$$

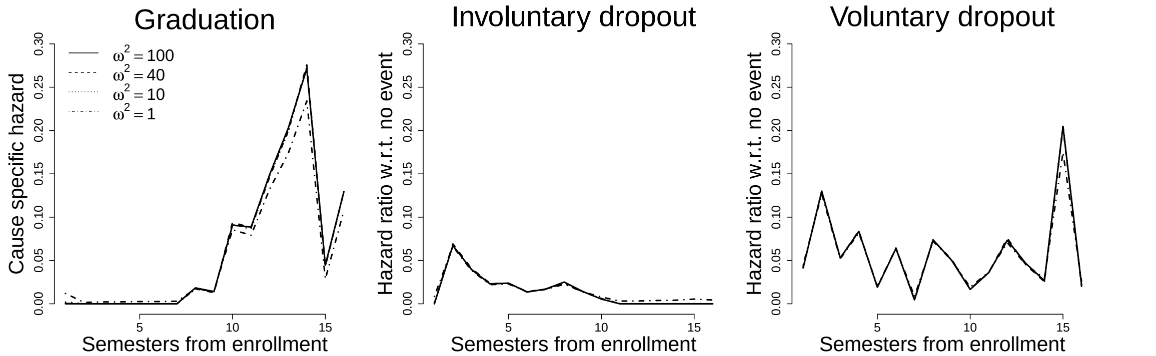

Figure 3: Chemistry students: posterior median trajectory of the hazard rate for each competing event, using the model in (4) under $\delta_{(r)} \sim \text{Cauchy}_{t_0}(\mathbf{0}_{t_0}, \omega^2 I_{t_0})$.

In contrast to the prior in Gelman et al. (2008), the multivariate Cauchy prior in (7) does not lead to independence between the δ_{rt}'s. Small values of ω^2 result in tight priors (as in Gelman et al., 2008). Instead, large values of ω^2 assign non-negligible probability to large negative (and positive) values of the δ_{rt}'s. For the analysis of the PUC dataset, $\omega^2 = 100$ is adopted matching the suggestion given in Gelman et al. (2008) for the intercept term. Of course, we could also define this hyper-parameter from prior information. However, this would require non-trivial prior elicitation because, as discussed early, it is not entirely clear a priori which δ_{rt}'s are affected by the separation issue. For Chemistry students, Figure 3 illustrates that our choice of setting $\omega^2 = 100$ leads to essentially identical posterior estimates for the cause-specific baseline cause-specific hazard rates as, say, $\omega^2 = 10$. For simplicity, covariates are excluded for this comparison. We conclude that: (i) choosing a value of ω^2 is not critical for those periods where the separation is not a problem (where the data is more informative and dominates posterior inference) and (ii) the actual value of ω^2 is not critical as long as it is large enough to allow near-zero hazard rates on the periods where this is required (e.g. semesters 1 to 7 for graduations in Chemistry). Including covariates with the prior as below, we found (not reported) that posterior inference on the regression coefficients in B is also robust with respect to the choice of ω^2.

We define $X = (x_1, \ldots, x_n)'$ and independently assign to each vector of cause-specific regression coefficients the g-prior described in Sabanés Bové and Held (2011), i.e.

$$\beta_{(r)} | g_r \sim \text{Normal}_k(\mathbf{0}_k, 4 g_r (X'X)^{-1}), \qquad r = 1, \ldots, \mathcal{R}. \tag{9}$$

This is an extension of the prior introduced in Zellner (1986) in the context of Generalized Linear Models (in the logistic case, the only difference is the scaling factor 4 in the covariance matrix). The family of g-priors is a popular choice in Bayesian model selection and averaging under uncertainty regarding the inclusion of covariates (e.g. Fernández et al., 2001). In particular, they are invariant to scale transformations of the covariates and incorporate the correlation structure among the covariates. However, the choice of values for $\{g_1, \ldots, g_{\mathcal{R}}\}$ can fundamentally affect the posterior inference and is quite challenging (Liang et al., 2008; Ley and Steel, 2009).

For a binary logistic regression, Hanson et al. (2014) elicit g_r using averaged prior information (across different covariate configurations). Alternatively, and in line with the Bayesian treatment of uncertainty, a hyper-prior can be assigned to each g_r, inducing a hierarchical prior structure (Liang et al., 2008). Here, we adopt the hyper-g/n prior of Liang et al. (2008), i.e.

$$\pi(g_r) = \frac{1}{n}\left(1 + \frac{g_r}{n}\right)^{-2}. \tag{10}$$

This hyper-prior was also investigated in Sabanés Bové and Held (2011), who show it leads to similar results as the hyper-prior implicitly adopted by Zellner and Siow (1980).

To assess the robustness of our results to the prior for the regression coefficients, we also investigated alternative prior choices such as the Zellner (1986) g-prior in combination with a benchmark Beta hyper prior for g (Ley and Steel, 2012). We concluded that this prior leads to very similar results (Section C.1 of the Supplementary Material).

4.2 Markov chain Monte Carlo implementation

Bayesian inference for a multinomial (or binary) logistic regression is not straightforward. There is no conjugate prior and sampling from the posterior distribution is cumbersome (Holmes and Held, 2006). A popular approach is to use an alternative representation of the multinomial logistic likelihood. For instance, Forster (2010) exploits the relationship between a multinomial logistic regression and a Poisson Generalized Linear Model. Holmes and Held (2006) adopt a hierarchical structure where the logistic link is represented as an infinite scale mixture of normals. Alternatively, Frühwirth-Schnatter and Frühwirth (2010) approximated the logistic link via a finite mixture of normal distributions. Instead, we adapt the hierarchical structure used in Polson et al. (2013) in order to construct a Gibbs sampling scheme for the multinomial logistic model, under the prior described in Section 4.1. Unlike the original implementation in Polson et al. (2013), we do not assume a Gaussian prior for the regression coefficients (which in our case includes δ_{rt}'s and $\beta_{(r)}$'s). In contrast, our prior is a product of independent multivariate Cauchy and hyper-g prior components. However, as both of these components can be represented as a scale mixture of normal distributions (see (8) and (9)), the sampler in Polson et al. (2013) can be extended by adding extra steps to the sampler, where values for $\Lambda_1, \ldots, \Lambda_{\mathcal{R}}, g_1, \ldots, g_{\mathcal{R}}$ in (8) and (9) are generated. In brief, at each iteration and for each $r \in \{1, \ldots, \mathcal{R}\}$, we sample

- $(\delta_{(r)}, \beta_{(r)})'|\Lambda_r, g_r$ according to the sampler described in Section 3.1 of Polson et al. (2013)
- $\Lambda_r|\beta_{(r)}, \delta_{(r)}, g_r \sim$ Gamma $\left(\frac{t_0+1}{2}, \frac{\delta'_{(r)}\delta_{(r)}}{2\omega^2}\right)$,
- $g_r|\beta_{(r)}, \delta_{(r)}, \Lambda_{(r)} \sim \pi(g_r|\beta_{(r)}) \propto g_r^{-k/2} \exp\left\{-\frac{\beta'_{(r)}X'X\beta_{(r)}}{2g_r}\right\}\pi(g_r)$.

As the full conditional for g_r is not a known distribution, an adaptive Metropolis-Hastings step (see Section 3 in Roberts and Rosenthal, 2009) is implemented for these parameters. A full description of this MCMC sampler is provided in Section B of the Supplementary Material. Our freely available R code (https://github.com/catavallejos/UniversitySurvival) is also documented in the Supplementary Material (Section D).

4.3 Bayesian model averaging

Our main goal is to identify the important risk (or prevention) factors associated with delayed graduations and both types of dropout. The latter can be translated into a model comparison problem, where candidate models are defined by different covariates. If k^* is the number of available covariates (k^* might differ from the number of regression coefficients k because categorical covariates may have more than two levels), there are $\mathcal{M} = 2^{k^*}$ candidate models. These are denoted by $M_1, \ldots, M_{\mathcal{M}}$ and we assume that if a categorical covariate is included, all its levels are represented. A Bayesian response to model uncertainty is Bayesian Model Averaging (BMA), which averages inference over all possible models, weighted according to the posterior model probabilities (Hoeting et al., 1999). BMA is a formal probabilistic way of dealing with model uncertainty and has been shown to lead to better predictive performance than choosing a single model (Raftery et al., 1997; Fernández et al., 2001). Let T_{obs} and R_{obs} be the observed times and event types, respectively. Using BMA, the posterior density of each cause-specific vector of regression coefficients is

$$\pi(\beta_{(r)}|T_{obs}, R_{obs}) = \sum_{m=1}^{\mathcal{M}} \pi_m(\beta_{(r)}|T_{obs}, R_{obs})\pi(M_m|T_{obs}, R_{obs}), \tag{11}$$

where $\pi_m(\beta_{(r)}|T_{obs}, R_{obs})$ is the posterior density of $\beta_{(r)}$ under model M_m and $\pi(M_m|T_{obs}, R_{obs})$ corresponds to the posterior probability assigned to M_m. Both can be derived using Bayes theorem. In particular, the latter corresponds to

$$\pi(M_m|T_{obs}, R_{obs}) = \frac{L(T_{obs}, R_{obs}|M_m)\pi(M_m)}{\sum_{j=1}^{\mathcal{M}} L(T_{obs}, R_{obs}|M_j)\pi(M_j)}, \quad \text{with} \quad \sum_{j=1}^{\mathcal{M}} \pi(M_j) = 1, \tag{12}$$

where $L(T_{obs}, R_{obs}|M_m)$ is the marginal likelihood of the data under the m-th model, after integrating out all model parameters with their prior. For $\delta_{(r)}$ we use the prior (7) and for $\beta_{(r)}$ the prior is defined as in (9), replacing X by the design matrix corresponding to the covariates included in model m (the elements of $\beta_{(r)}$ corresponding to excluded covariates have a prior point-mass at zero). In (12), $\pi(M_1), \ldots, \pi(M_{\mathcal{M}})$ represent the prior on model space. For the latter, we adopt a uniform prior assigning the same probability to every model

$$\pi(M_m) = \frac{1}{\mathcal{M}}, \quad m = 1, \ldots, \mathcal{M}. \tag{13}$$

This prior has an equivalent formulation in terms of covariate inclusion indicators γ_j (with $\gamma_j = 1$ if the j-th covariate is included in the model and $\gamma_j = 0$ otherwise) by independently setting $p_j = Pr(\gamma_j = 1) = 0.5$. Alternatively, and assuming p_j is the same for all covariates, we can assign a hyperprior to this common inclusion probability, which leads to less informative priors in terms of model size (Ley and Steel, 2009). This prior downweighs models with size around $k^*/2 = 4$, which is perhaps less intuitive for this application.

If k^* is small, the expressions involved in (11) could be estimated via a complete enumeration of the model space. In such a case, for each model, the MCMC sampler described in Section 4.2 can be run and the associated marginal likelihoods can be estimated e.g. using the Bridge sampling method proposed in Meng and Wong (1996). However, this strategy is time-consuming and does not scale well as k^* increases (even for moderate values). Instead, we extend the MCMC sampler described in Section 4.2 so that it also explores the model space and posterior probabilities associated to each model can be directly estimated from the MCMC chain. Here we implement a simple approach which is able to deal effectively with small numbers of covariates (such as in the case of the PUC dataset). This extends the implementation described in Section 4.2 by adding an extra step where covariate-inclusion indicators γ_j are sampled. At each iteration, the sampler sketched in Section 4.2 is implemented conditional on the currently sampled value of $\gamma = (\gamma_1, \ldots, \gamma_{k^*})'$. Then, new values for γ are sequentially generated from the full conditionals

$$\pi(\gamma_j|\gamma_{-j}, \delta, B, \Lambda, g) \propto L(\gamma) \times \left[\prod_{r=1}^{\mathcal{R}} \pi(\beta_{(r)}|g_r; X_\gamma)\right], \tag{14}$$

with $\gamma_{-j} = \{\gamma_1 \ldots, \gamma_{j-1}, \gamma_{j+1}, \ldots, \gamma_{k^*}\}$, $\Lambda = \{\lambda_1, \ldots, \lambda_{\mathcal{R}}\}$ and $g = \{g_1, \ldots, g_{\mathcal{R}}\}$. In (14), $L(\gamma)$ represents the likelihood function associated to the model defined by (6) and the covariate configuration induced by γ. Additionally, $\pi(\beta_{(r)}|g_r; X_\gamma)$ is defined as in (9), replacing X by X_γ, i.e. the design matrix based on the covariate configuration defined by γ. As (14) does not correspond to a known distribution, we implemented Metropolis-Hastings updates as further detailed in the Supplementary Material (Section B). BMA results obtained via this extended sampler were very similar to those obtained via a complete enumeration over the model space.

BMA estimates can be directly obtained by post-processing the sampled chain (after removing a burn-in period). In particular, $\pi(M_m|T_{obs}, R_{obs})$ can be estimated as the proportion of times that M_m was visited. In addition, the Marginal Posterior Probability of Inclusion (MPPI), summarising the importance of each covariate, is estimated as the proportion of times that the sampled value of γ_j is equal to 1.

5 Empirical results for the PUC data

The PUC dataset is analysed through the model in (4) using the prior and the algorithm described in Section 4. Due to marked differences (in terms of covariates and outcomes) between different degree programmes (see Section 2), we fit separate models for each programme. This strategy alleviates multiple comparison issues as for each degree programme we make separate assumptions (e.g. regarding covariate inclusion). In addition, this decision is supported by the results presented in this section, where the BMA approach selects different sets of covariates across degree programmes. In this article, we focus on some of the science programmes for which the rates of dropout and/or late graduations are normally higher. In particular, we consider Chemistry (379 students), Mathematics and Statistics (598 students) and Physics (237 students). For all programmes, 8 covariates are available (see Table 1), inducing $2^8 = 256$ possible models (using the same covariates for each cause-specific hazard). Selection scores cannot be directly compared across admission years (the test varies from year to year). Hence, the selection score is replaced by an indicator of being in the top 10% of the enrolled students (for each programme and admission year). The following regression coefficients are defined for each cause (the subscript r is omitted for ease of notation): β_1 (sex: female), β_2 (region: metropolitan area), β_3 (parents' education: with degree), β_4 (high school: private), β_5 (high school: subsidised private), β_6 (funding: scholarship only), β_7 (funding: scholarship and loan), β_8 (funding: loan only), β_9 (selection score: top 10%), β_{10} (application preference: first) and β_{11} (gap after high school graduation: yes). For all degree programmes, we run the MCMC algorithm for 200,000 iterations, storing draws every 10 iterations. The results presented here exclude the first 100,000 iterations as a conservative burn-in period. Trace plots and the usual convergence criteria strongly suggest good mixing and convergence of the chains (see Section C of the supplementary material).

5.1 Temporal behaviour

Figure 4 shows the estimated trajectory of the baseline cause-specific hazard rates (without additional covariates), capturing the temporal behaviour of the analysed outcomes. The first row of panels in Figure 4 roughly recovers the same patterns as in Figure 2, suggesting these estimates

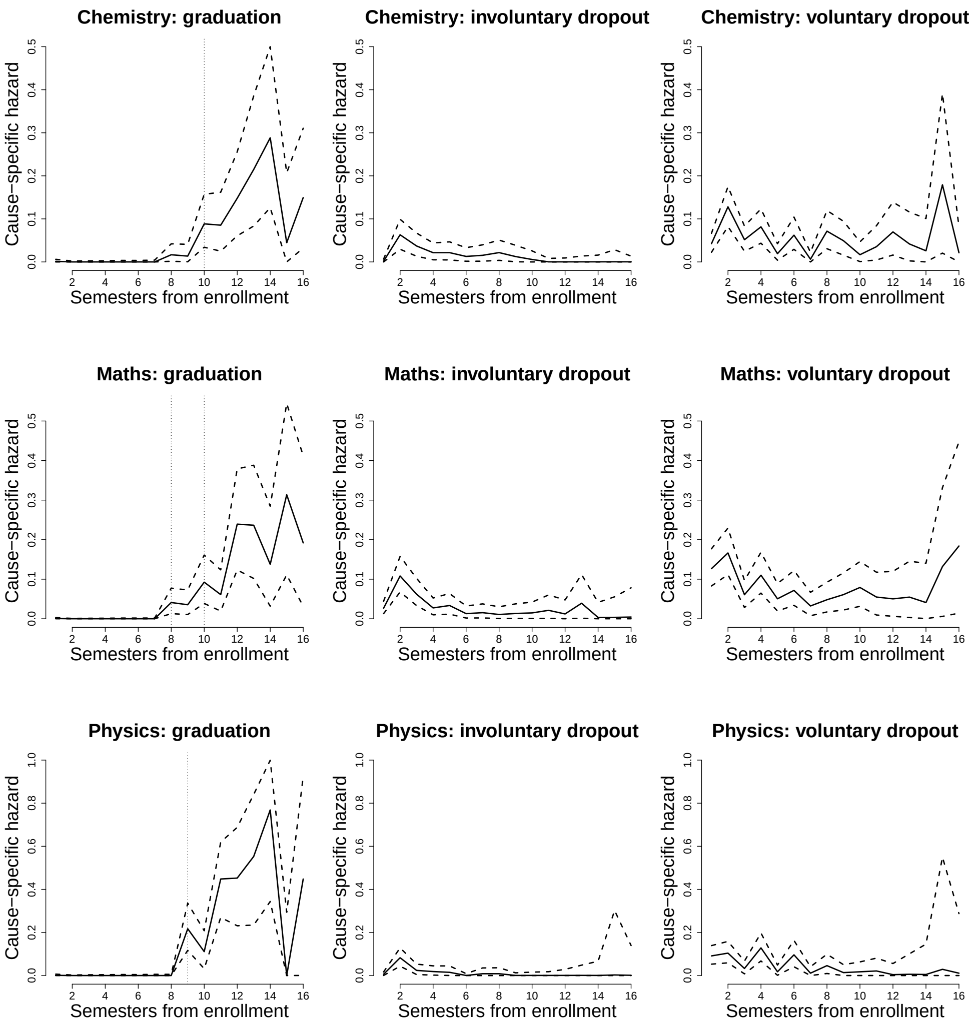

Figure 4: BMA estimates for baseline cause-specific hazards (without additional covariates). Solid lines represent the trajectory of the posterior medians and dashed lines represent the associated limits for the 95% highest posterior density interval. For graduation hazards, dotted vertical lines are located at the official duration of the programme (in Mathematics and Statistics, students in Statistics can take two additional semesters to get a professional degree).

are in line with the data. Some similarities appear between these programmes. For example, the highest risk of involuntary dropout occurs by the end of the second semester, possibly related to poor performances during the first year of studies (where the students' performance can be affected by a sudden transition from high school to university standards). In addition, during the 4 first years of enrolment, the hazard rate associated to voluntary dropouts has spikes located at even semesters. This is not surprising as withdrawing at the end of the academic year allows students to re-enroll in a different programme without having a gap in their academic careers. In terms of graduation, mild increases are located at the official duration of the programmes. Nonetheless, for these programmes, the highest hazards of graduation occur with about 4 semesters of delay (the spikes at the last period are due to a cumulative effect, as δ_{r16} represents the period $[16, \infty)$).

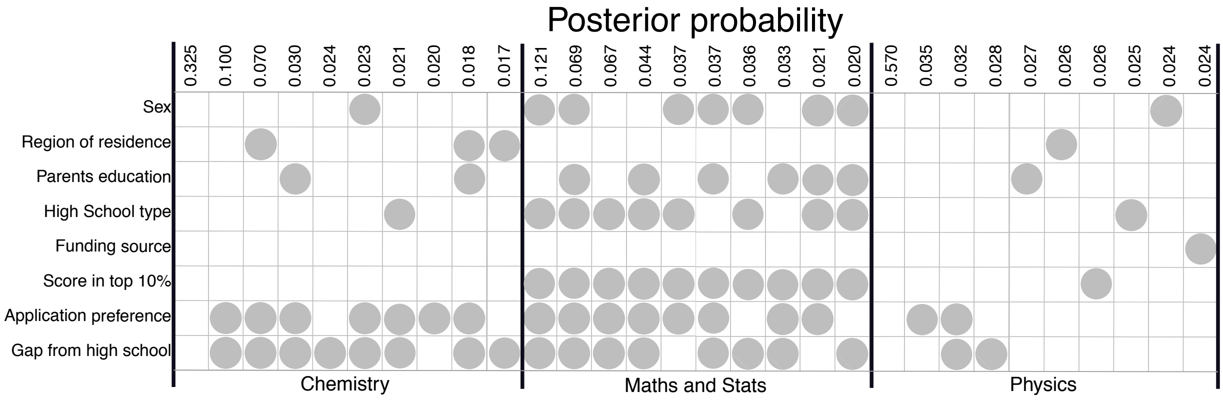

Figure 5: BMA posterior inference. Top 10 models according to posterior probability, for each degree programme. Gray circles indicate covariate inclusion.

Table 2: MPPI of each covariate for the three analysed degree programmes.

Programme	Sex	Region	Parents	School	Funding	Top 10%	Pref.	Gap
Chemistry	0.11	0.27	0.14	0.12	0.06	0.12	0.47	0.51
Maths and Stats	0.63	0.17	0.46	0.73	0.14	0.90	0.81	0.76
Physics	0.08	0.09	0.09	0.09	0.03	0.05	0.23	0.17

5.2 Covariate selection

The main objective of our analysis is to identify risk factors related to delayed graduations and dropout events. We tackle this as a variable selection problem, using a BMA approach to uncover the most relevant covariates. Figure 5 shows the top 10 models, with the highest posterior probabilities, for each degree programme. Importantly, across all degree programmes, the maximum posterior probability associated to a single model is 0.570 (the null model with no covariates, in the case of Physics). This suggests that a BMA strategy — where the uncertainty about covariate selection is formally accounted for — seems more appropriate than selecting a single model. For each covariate, marginal inclusion probabilities (MPPIs) are displayed in Table 2. Overall, the highest MPPIs relate to the student's application preference and the gap indicator (associated with β_{10} and β_{11}, respectively). This is in line with the results in Figure 5, where these covariates are typically included in the models with the highest posterior probabilities. Moreover, and perhaps surprisingly, the source of funding (linked to β_6, β_7 and β_8) appears to be the least important covariate in this context. This is likely due to an overlap of information between this variable and other socio-economic indicators (see Table 1). However, despite these similarities, results show important variations across degree programmes. While most covariates appear to be relevant for the Mathematics and Statistics programme, Chemistry and Physics favour sparser models.

5.3 Interpretation

The MPPIs displayed in Table 2 indicate which covariates have a clear effect on the university outcomes and their associated temporal behaviour. Together with the covariate selection, it is also important to consider the sign and the magnitude of the regression coefficients to assess specific covariate effects on graduation and both types of dropouts. This determines whether a covariate corresponds to a risk factor regarding the analysed outcomes. In particular, for graduations, negative regression coefficients indicate risk factors (students graduate less and slower). For the dropout events, those covariates with positive effects are risk factors (students tend to drop out more and earlier).

From (11), the posterior distribution of each β_{rj} is given by a point mass at zero (equal to the probability of excluding the j-th covariate) and a continuous component (a mixture over the posterior distributions of β_{rj} given each model where the corresponding covariate is included). Figure 6 summarises the continuous component of the posterior distribution by BMA estimates for all regression coefficients (posterior medians and 95% highest posterior density intervals) and all degree programmes. Some effects are consistent across different degree programmes. For instance, in line with Figure 5 and Table 2, one of the strongest effects relate to the student's application preference (see β_{10}). In general, students who applied as a first preference to these degrees graduate more and faster (i.e. for graduations, the posterior distribution of β_{10} is concentrated on positive values). They also exhibit a lower rate of voluntary dropout, possibly linked to a higher initial motivation. To put the magnitude of the effect into perspective, the odds for outcome $r = 1, 2, 3$ versus no event are multiplied by a factor $\exp(\beta_{r\ 10})$ if the programme was the student's first preference. Whether or not the student had a gap between high school graduation and university admission (β_{11}) also has a strong influence on the academic outcomes for these programmes. These gaps can, e.g., correspond to periods in which the student was preparing for the admission test (after a low score in a previous year) or enrolled in a different programme. Overall, this gap is linked to fewer and slower graduations. In addition, in each semester, this gap corresponds to a higher risk of being expelled from Chemistry. The effects of other covariates are less stable across the programmes. Whereas the effect of the student's sex (β_1) is almost negligible in Chemistry and Physics, female students in Mathematics and Statistics have a lower risk of being expelled. This is consistent with the MPPI estimates in Table 2, which are low for Chemistry and Physics.

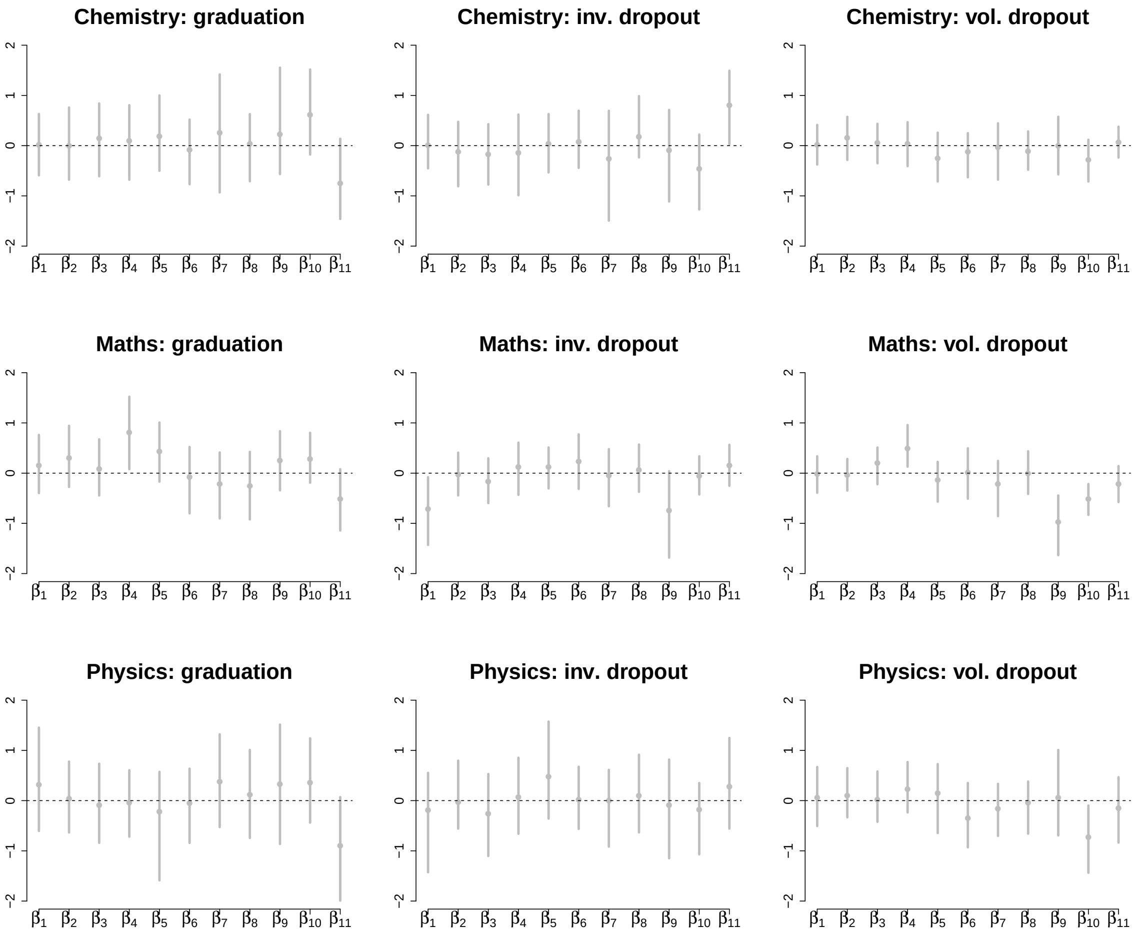

Figure 6: BMA estimates for regression coefficients, per degree programme and event type. Dots are located at posterior medians and vertical segments represent the 95% highest posterior density interval. The following regression coefficients are defined for each event type (the subscript r is omitted for ease of notation): β_1 (sex: female), β_2 (region: metropolitan area), β_3 (parents' education: with degree), β_4 (high school: private), β_5 (high school: subsidised private), β_6 (funding: scholarship only), β_7 (funding: scholarship and loan), β_8 (funding: loan only), β_9 (selection score: top 10%), β_{10} (application preference: first) and β_{11} (gap after high school graduation: yes).

6 Concluding remarks

In this article, a simple but flexible competing risks survival model is employed for the modelling of university outcomes (graduation or dropout). This is based on the Proportional Odds model introduced in Cox (1972) and can be estimated by means of a multinomial logistic regression. The suggested sampling model has been previously employed in the context of university outcomes, but the structure of typical university outcomes data seriously complicates a maximum likelihood analysis. Here, we use a Bayesian setting, where an appropriate prior distribution allows the extraction of sensible information from the data. Adopting a hierarchical structure allows for the derivation of a reasonably simple MCMC sampler for inference. The proposed methodology is applied to a dataset on undergraduate students enrolled in the Pontificia Universidad Católica de Chile (PUC) over the period 2000-2011.

As illustrated in Sections 2 and 5, there are strong levels of heterogeneity between different programmes of the PUC. Hence, building a common model for the entire university is not recommended. For brevity, this article only presents the analysis of three science programmes for which late graduations and dropouts are a major issue, but the methodology presented here can be applied to all programmes. We formally consider model uncertainty in terms of the covariates included in the model. In view of the posterior distribution on the model space, we conclude that choosing a single model is not appropriate and that BMA provides more meaningful inference in our context. In summary, our analysis uncovers strong evidence in favour of high application preferences acting as a risk-mitigating factor in terms of delayed graduations and voluntary dropout in the three analysed programmes. In contrast, having a gap between high school graduation and university admission generally increases the risks associated to these events. The performance in the selection test is also an important determinant for Mathematics and Statistics, where good performance mitigates the risk of both types of dropout. Other factors, such as gender and the region of residence, only appear to matter for some of the programmes. Covariates related to funding sources do not appear to influence outcomes of any of the three programmes.

One limitation of our approach are potential violations to the proportional odds (PO) assumption. While this assumption does not seem too unreasonable for some covariates and degree programmes (e.g. preference and gap covariates in Mathematics and Statistics), this is less clear in other cases (see Section E in the Supplementary Material for an approximate nonparametric check). This is perhaps not that critical for those covariates that are not robustly associated with the analysed outcomes (e.g. for Physics students, where the null model concentrates more than 50% of the posterior probability). Potential reasons for deviations from PO are unobserved confounders and time-varying covariate effects (e.g. if some of the variables recorded at admission might have a diminishing effect throughout time). In such cases, possible solutions would be to keep the PO specification but to add an interaction effect between time and covariates (e.g. different effect magnitudes during the first year of admission) and to incorporate random effects in order to account for unobserved sources of heterogeneity. This will be investigated in follow-up analyses. However, it should be borne in mind that inference in such more general models might well be challenging with the available sample sizes.

An issue that was not dealt with explicitly is that of missing covariate information. In this analysis, we implicitly assumed that the missingness pattern can be ignored. It would be interesting to deal with this issue more carefully in subsequent analysis.

An obvious extension of the model presented here is to allow for different covariates in the modelling of the three risks within the same programme. This would substantially increase the number of models in the model space, so the simple sampler implemented here is expected to perform poorly as it was only designed to cope with moderately large model spaces. In such a case, more sophisticated sampling schemes such as the evolutionary stochastic search implemented in Bottolo and Richardson (2010) could be employed.

References

Albert, A. and J. Anderson (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika 71*, 1–10.

Andersen, P. K., R. B. Geskus, T. de Witte, and H. Putter (2012). Competing risks in epidemiology: Possibilities and pitfalls. *International Journal of Epidemiology 41*(3), 861–870.

Arias Ortis, E. and C. Dehon (2011). The roads to success: Analyzing dropout and degree completion at university. Working Papers ECARES 2011-025, ULB - Universite Libre de Bruxelles.

Bean, J. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education 12*, 155–187.

Bottolo, L. and S. Richardson (2010). Evolutionary stochastic search for Bayesian model exploration. *Bayesian Analysis 5*(3), 583–618.

Chen, M.-H., J. Ibrahim, and S. Kim (2008). Properties and implementation of Jeffreys's prior in binomial regression models. *Journal of the American Statistical Association 103*, 1659–1664.

Clerici, R., A. Giraldo, and S. Meggiolaro (2014). The determinants of academic outcomes in a competing risks approach: evidence from Italy. *Studies in Higher Education 40*(9), 1535–1549.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B 34*, 187–220.

Crowder, M. (1996). On assessing independence of competing risks when failure times are discrete. *Lifetime Data Analysis 2*(2), 195–209.

Crowder, M. (2001). *Classical competing risks*. Chapman & Hall/CRC.

Fernández, C., E. Ley, and M. Steel (2001). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics 16*, 563–576.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika 80*, 27–38.

Forster, J. (2010). Bayesian inference for Poisson and multinomial log-linear models. *Statistical Methodology 7*, 210–224.

Frühwirth-Schnatter, S. and R. Frühwirth (2010). Data augmentation and MCMC for binary and multinomial logit models. In T. Kneib and G. Tutz (Eds.), *Statistical Modelling and Regression Structures*, pp. 111–132. Springer.

Gelman, A., A. Jakulin, M. Pittau, and Y. Su (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics 2*, 1360–1383.

Hanson, T., A. Branscum, and W. Johnson (2014). Informative g-priors for logistic regression. *Bayesian Analysis 9*(3), 597–612.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical Science 14*, 382–401.

Holmes, C. and L. Held (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis 1*, 145–168.

Ibrahim, J. and P. Laud (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. *Journal of the American Statistical Association 86*, 981–986.

Ley, E. and M. Steel (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics 24*, 651–674.

Ley, E. and M. Steel (2012). Mixtures of g-priors for Bayesian model averaging with economic applications. *Journal of Econometrics 171*, 251–266.

Liang, F., R. Paulo, G. Molina, M. Clyde, and J. Berger (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association 103*, 410–423.

Meng, X. and W. Wong (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica 6*, 831–860.

Murtaugh, P., L. Burns, and J. Schuster (1999). Predicting the retention of university students. *Research in Higher Education 40*, 355–371.

Poirier, D. (1994). Jeffreys' prior for logit models. *Journal of Econometrics 63*, 327–339.

Polson, N., J. Scott, and J. Windle (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. *Journal of the American Statistical Association 108*, 1339–1349.

Raftery, A., D. Madigan, and J. Hoeting (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association 92*, 179–191.

Roberts, G. and J. Rosenthal (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics 18*, 349–367.

Sabanés Bové, D. and L. Held (2011). Hyper-g priors for generalized linear models. *Bayesian Analysis 6*, 387–410.

Scott, M. and B. Kennedy (2005). Pitfalls in pathways: Some perspectives on competing risks event history analysis in education research. *Journal of Educational and Behavioral Statistics 30*, 413–442.

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

Singer, J. and J. Willett (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press, USA.

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

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences 72*(1), 20–22.

Willett, J. and J. Singer (1991). From whether to when: New methods for studying student dropout and teacher attrition. *Review of Educational Research 61*, 407–450.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In P. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*, North-Holland: Amsterdam, pp. 233–243.

Zellner, A. and A. Siow (1980). Posterior odds ratios for selected regression hypotheses (with discussion). In J. Bernardo, M. DeGroot, D. Lindley, and A. Smith (Eds.), *Bayesian Statistics*, Valencia: University Press, pp. 585–603.