To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–19

Bayesian binary quantile regression for the analysis of Bachelor-to-Master transition

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

The multi-cycle organization of modern university systems stimulates the interest in studying the progression to higher level degree courses during the academic career. In particular, after the achievement of the first level qualification (Bachelor degree), students have to decide whether to continue their university studies, by enrolling in a second level (Master) programme, or to conclude their training experience. In this work we propose a binary quantile regression approach to analyze the Bachelor-to-Master transition phenomenon with the adoption of the Bayesian inferential perspective. In addition to the traditional predictors of academic outcomes, such as the personal characteristics and the field of study, different aspects of student's performance are considered. Moreover, the role of a new contextual variable, representing the type of university regulations experienced during the academic path, is investigated. The utility of the Bayesian binary quantile regression to characterize the non-continuation decision after the first cycle studies is illustrated with an application to administrative data of Bachelor graduates at the School of Economics of Sapienza University of Rome. **The method favorably compares with more conventional model specifications concerning the conditional mean of the binary response.**

Keywords: Binary quantile regression; Asymmetric Laplace distribution; Data augmentation; Gibbs sampling; Bachelor-to-Master transition; university drop-out

1. Introduction

The set of regulations launched in 1999 by the Bologna Process deeply reformed the organization of the higher education systems in Europe, with the aim at expanding the participation to tertiary education and improving its outcomes.

In accordance with the reforms of the Bologna Declaration, the university systems of the signatory countries are currently structured in three hierarchical levels of academic education: the 3-year first cycle degree (*Bachelor degree*), the 2-year second cycle degree (*Master degree*) and the 3-year third cycle research doctorate. The multi-cycle organization of modern higher education systems offers the possibility to examine new academic outcomes of students' training path and also to assess well-studied phenomena in new critical steps of the university career, such as the delicate phase of transition between consecutive levels of the academic studies. The progression from the first to the second cycle studies, referred to as *Bachelor-to-Master transition*, is a crucial step of the university experience, but its characterization still remains an open problem from both an institutional and a statistical perspective. When considering the entire degree path, involving both the first and the second level qualification, the non-transition to Master level determines a specific type of students' non-completion behavior. In particular, after the achievement of the Bachelor degree, students have to decide whether to continue their university studies, by enrolling in a second cycle programme, or to stop their training experience. From this perspective, the decision to interrupt the academic career after the completion of the first cycle course leads to the drop-out of the university, that is, to the withdrawal of the Bachelor graduate (also known as *undergraduate*) from the institution. In this paper we focus our attention on the non-progression decision after the first level qualification attainment and investigate the factors that can influence the choice of not embarking on a second level specializing degree course within the same institution of enrollment. Besides the classical predictors of academic outcomes, such as the personal characteristics and the field of study, a special attention is turned to different aspects of student's performance and on a contextual variable, indicating the type of academic regulations that the students experienced during their first level career.

From a statistical point of view, the university drop-out analysis has been traditionally addressed with regression methods within the generalized linear model family (GLM) and extensions thereof, by either adopting a binary definition of the response variable to discriminate between drop-out and continuation (see for example [20], [8] and references therein) or considering a more complex setting with multiple alternatives as in [9]. In this paper we consider a binary quantile regression (BQR) approach for the analysis of the non-continuation outcomes with the adoption of the Bayesian inferential framework. Bayesian methods are very useful and flexible tools accounting for parameter uncertainty by combining data with prior information.

Quantile regression (QR) provides a very useful device to explore as different location measures of the response distribution are affected by the predictors, in order to gain a more in-depth understanding of the relation between the outcome of interest and the explanatory variables. In so doing, it is possible to capture also 'extreme' behaviors, not detectable by those models that take into account only the mean of the response variable. The Bayesian QR framework usually relies the inference on the Asymmetric Laplace distribution (ALD), as described in the seminal works by [53] and [30] and in subsequent contributions, such as [36] and [13]. By using a data augmentation method, more recently QR modeling has been extended for the treatment of binary response variables, see [11] and [10]. Here we illustrate the utility of the Bayesian BQR approach to describe the non-continuation decision with an application to administrative data on students who attained their first level qualification at the School of Economics of Sapienza University of Rome during the period from the academic year (a.y.) 2009-2010 to 2012-2013. Up to our knowledge this represents the first attempt to investigate the Bachelor-to-Master transition phenomenon by means of a Bayesian binary regression model.

The remainder of the paper proceeds as follows. In Section 2 we provide an outline of the higher education reform launched in 1999 and describe its effects on the current Italian university system. Section 3 overviews the existing analyses of the Bachelor-to-Master transition, whereas Section 4 contains a review of the BQR model with the related MCMC methods to accomplish for a fully Bayesian estimation. The administrative data set and the results of the Bayesian BQR are presented in Section 5. The paper ends with concluding remarks and proposals of future developments in Section 6.

2. The university reform in Italy

The Bologna Declaration in 1999 marked the beginning of a radical and ambitious reform process of higher education in Europe, known as *Bologna Process*. With the aim at creating an European Area of Higher Education and promoting its international competitiveness, the major novelties introduced by the Bologna Process include: (i) the harmonization of the European higher education qualification systems, (ii) the adoption of tertiary education study programmes based on a multi-cycle structure and (iii) the introduction of a common system of credits to facilitate qualification comparisons and encourage students' mobility.

In Italy the adaptation to the objectives of the Bologna Process started with the Ministerial Decree (MD) 509/99, whose reforms became effective in the a.y. 2001-2002 and continued with the subsequent MD 270/04, that was applied starting from the a.y. 2008-2009. The reforms led to a substantial remodeling of the architecture of the Italian university system. The fundamental change concerns the replacement of the single-cycle 4/5-year degree programme with a study path organized in three hierarchical levels of academic education:

- the 3-year first cycle degree (equivalent to the Bachelor degree), that provides both the basic knowledge and the necessary tools for the acquisition of adequate professional competence, in order to facilitate the entry in the world of work;
- the 2-year second cycle degree (equivalent to the Master degree), that offers an advanced level of training to prepare students for the exercise of specific professions or highly qualified activities requiring specific skills;
- the 3-year third cycle research doctorate, that represents the maximum level of academic education and the key step to take up a career based on the research activity.

Due to the duration of the first two academic levels, in Italy the reform is also referred to as *'3+2' degree system.*

As mentioned above, one of the most important reforms established by the Bologna Process is the introduction of the European Credit Transfer and Accumulation System (ECTS), that supplies a standardized criterion to describe higher education qualifications in the European Union through the adoption of a common measure of the study commitment needed for their achievement. The ECTS aims at increasing transparency of qualification comparisons and at facilitating students' mobility among the European education institutions thanks to the transfer of the ECTS credits (hereinafter, simply referred to as credits). Students can progressively accumulate credits by passing the exams scheduled by their study plan. Each exam, in fact, is associated with a certain number of credits commensurate with the student workload required to achieve the expected learning outcomes of the course.

The latter MD 270/04 brought a series of modifications and integrations to the former MD 509/99 with the declared aim at increasing the autonomy of the single academic institutions, mostly in the organization of their training services. Apart from some formal changes regarding the designation of the degree qualifications and the groupings of the study courses into homogeneous degree classes, two important points deserve to be remarked for their potential impact on students' choice to start a second cycle programme. The former entails the specification of a maximum number of exams for the qualification achievement, in order to assist a more coherent attribution of the credits with respect to the required workload and to avoid an extreme fragmentation of the training activities. This limit varies according to the qualification level and is equal, respectively, to 20 exams for the Bachelor degree and 12 exams for the Master degree. The latter point concerns the set of reforms that regulate the access to a Master degree course and establishes a stronger separation between the first and the second cycle career. In particular, although the Bachelor degree is still a fundamental requisite, with the MD 270/04 the admission to a second level programme is constrained to the possession of curricular requirements and of an adequate personal preparation described by the academic regulations of the specific Master course and, hence, is not necessarily dictated by the first level study path. This aspect facilitates undergraduates' mobility, in the sense that they have a major chance to start a Master level experience in a school different from that where they have earned the Bachelor degree. Moreover, unlike the MD 509/99, with the new MD 270/04 the exam grades of the first level career do not contribute to determine the final grade of the Master degree. This set of reforms makes the prerequisites for accessing to a second cycle programme less restrictive and dependent on the previous first cycle studies. The greater flexibility and the more careful recalibration of the number of activities in the

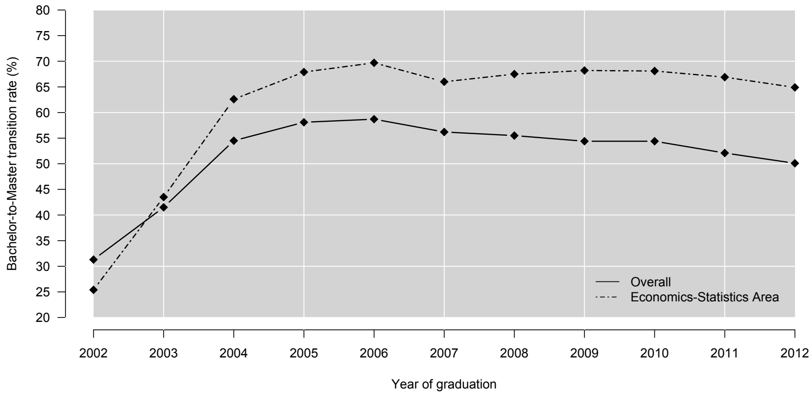

Figure 1.: Bachelor-to-Master transition rate (%) in Italy: overall percentage (solid line) and the one for the Economics-Statistics area of study (dashed line). Source: [5].

study programme introduced by the MD 270/04 could represent an incentive for the undergraduates to continue their academic studies at Master level. For this reason, in addition to other explanatory variables, we will analyze the effect of the MD, that is, the role of the academic regulation context that the students faced during the first level career on their continuation decisions.

3. An overview of the Bachelor-to-Master transition literature

Whereas the interruption of attendance by enrolled students (university drop-out) and the transition from university to work have been widely investigated in the higher education literature, see for example [47], [49], [20], [8], [9], [33], [14], [22], [7] and [43], the progression to the second cycle academic studies has received much less attention. Nevertheless, the athenaeum is strongly interested in promoting and encouraging the completion of the '3+2' degree programme, in order to train highly qualified professional figures, who can better face the selection of the labour market, and to increase its competitiveness. Moreover, an in-depth understanding of the Bachelor-to-Master transition mechanism is motivated also by organizational purposes, to assist the authorities in planning and updating the training offer, as well as by economic reasons concerning the optimization of resource allocation.

From an institutional point of view, the attention on the progression to the second cycle studies of the Italian Bachelor graduates is proved by the monitoring of the National Agency for the Evaluation of Universities and Research Institutes (ANVUR) of the Italian Ministry of Education, University and Research. Figure 1 shows the trend of the Bachelor-to-Master transition rate for the decade 2002-2012, as reported in [5], where the solid line indicates the national average and the dashed one is specific for the Economics-Statistics area of study. As remarked in the report, starting from 2005-2006 data are more stable and reliable, since in those years the '3+2' system became fully operating and the first genuine Bachelor cohorts faced with the transition decision. The proportion of undergraduates who progress to a Master degree course reached around 59% in 2006, but slowly drops in the second part of the period and shrinks up to 50% in 2012. The continuation rate for the Economics-Statistics area is manifestly above the national values, ranging between 65-70%, and follows an analogous decreasing tendency at the end of the period. For organizational purposes, identifying the determinants of the decision to enroll in a Master degree course can concretely suggest to the academic institution how to incentivize and support the progression to higher level studies.

Indeed, a factual research approach for the analysis of the transition between consecutive study programmes, based on the application of statistical procedures or model-based methods, has not been yet deeply developed. In contrast, despite the multi-cycle structure of the university system is effective for more than a decade, the investigation appears to be limited to the computation of descriptive summaries in the national reports on the state of the higher education systems, that clearly does not allow for a constructive characterization of the transition phenomenon. With regard to the Italian experience, the main reason lies in the lack of exhaustive and unified national archives of students' individual careers, that record the evolution and the outcomes of their academic paths.

In the international literature, very few attempts have been implemented to go further basic descriptive analyses. For example, by means of a factor analysis [18] explored the motivations that guide Bachelor graduates at universities of applied sciences to move to research-led universities for entering into a Master programme. [51] and [39], instead, interpret the transition/exit of first level graduates as a selection mechanism and apply a logistic regression model to identify the role of several institutional factors on the continuation outcome in the Hungarian higher education context. In [46] a collection of studies from several European countries concerning Bachelor graduates is presented. [15] contribute to such a collection with the results of the most recent surveys on the Italian graduates carried out by the AlmaLaurea Inter-University Consortium, with a special focus on international mobility, employment condition and the study continuation decision of the Bachelor graduates. Although their analysis relies on rich and integrated data bases, it still remains at an exploratory level.

In the attempt to go beyond a partial picture of the phenomenon and unveil the determinants of the transition dynamic, a multivariate setting that describes the causal relation between the outcome and a set of explanatory variables should be invoked. Our work contributes to fill this gap in the Bachelor-to-Master transition literature by making use of a recent proposal of regression modeling for binary responses in the Bayesian QR framework, that we review in the next section.

4. Bayesian quantile regression modeling for binary responses

In order to formalize the non-transition event, the response variable is represented by a binary indicator y, equal to 1 to denote the *Bachelor-to-Master drop-out*, that is, the case of student's withdrawal after the achievement of the Bachelor degree, and equal to 0 in the case of continuation of the academic studies at Master level. In this work we consider the BQR approach in a Bayesian framework, for which a continuous latent variable y^* **underlying the choice of dropping-out** after the Bachelor qualification is introduced. The BQR model allows for studying the effect of known predictors on location measures of the latent **propensity-to-drop-out** distribution other than the mean and can integrate the analysis with the description of extreme behaviors and attitudes concerning the non-continuation decision. In general, the application of the QR setup is also motivated as a more robust alternative to the traditional conditional mean model in the presence of outliers and heteroskedasticity [23]. Moreover, the Bayesian paradigm considered here is particularly useful for modeling binary response variables, since it solves some of the technical drawbacks faced when the frequentist approach is applied. **Frequentist methods, in fact, may exhibit difficulties in optimizing the regression parameters of the BQR model, as discussed in [21], and in building confidence intervals for the estimates. The latter point concerns the limiting distribution of the**

frequentist (maximum score) estimator introduced in the seminal works [37] and [38], whose complicated form prevents from a straightforward computation of the asymptotic standard errors, as described in [1]. See also [11] for a more in-depth review and discussion on frequentist approaches for BQR.

To our knowledge there is no previous application of the Bayesian BQR model for the analysis of the Bachelor-to-Master transition phenomenon. In this section we review the BQR model and detail how, from a Bayesian estimation perspective, the data augmentation method represents the key strategy to address the inferential issues related to the discreteness of the outcomes.

4.1 *Model setup*

Let $\mathbf{y} = (y_1, \ldots, y_n)$ be the vector of the observed binary outcomes, where the subscript $i = 1, \ldots, n$ indexes the sample units. For an arbitrary quantile level $\tau \in (0, 1)$, the BQR model originally proposed by [37] postulates that

$$y_i = I_{[y_i^* \geq 0]}, \tag{1}$$

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta}_\tau + \epsilon_i, \tag{2}$$

where $I_{[E]}$ denotes the indicator function of the event E, $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{iK})'$ is the $(K+1) \times 1$ vector of known predictors describing the individual profile with respect to the covariates, $\boldsymbol{\beta}_\tau = (\beta_{\tau 0}, \ldots, \beta_{\tau K})'$ is the $(K+1) \times 1$ vector of regression parameters depending on τ and ϵ_i is the random error component such that $\int_{-\infty}^{0} f_{\epsilon_i}(\epsilon_i)\, d\epsilon_i = \tau$. Assumption (1) illustrates the latent variable interpretation of the BQR, that relies on the introduction of an underlying quantitative variable y_i^* associated to each observed data point y_i. Equality (1), in fact, expresses the observed outcome y_i as the dichotomization of the corresponding latent continuous response y_i^*. As described by equality (2), an ordinary QR model is postulated on y_i^*, where the error variable ϵ_i is constrained to have the τ-th quantile equal to zero. **It follows that**

$$Q_{y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_\tau}(\tau) = \mathbf{x}_i' \boldsymbol{\beta}_\tau,$$

where $Q_{y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_\tau}(\cdot)$ denotes the conditional quantile function of y_i^*. When $\tau = 0.5$, equality (2) becomes the popular conditional median regression model. The QR model for quantitative responses was firstly introduced by [26]. A general introduction on the topic can be found in [27], whereas [52] and [25] offer a review of the literature. The theoretical justification of combining (1) and (2) to construct a QR model for the analysis of binary outcomes is provided by the *equivariance property* of quantile functions described in [42], stating that $Q_{g(y^*)}(\cdot) = g(Q_{y^*}(\cdot))$ when g is a monotone nondecreasing function. In the BQR context this property applies since one has $g(\cdot) = I_{[\cdot]}$, leading to

$$Q_{y_i | \mathbf{x}_i, \boldsymbol{\beta}_\tau}(\tau) = I_{\left[Q_{y_i^* | \mathbf{x}_i, \boldsymbol{\beta}_\tau}(\tau) \geq 0\right]} = I_{[\mathbf{x}_i' \boldsymbol{\beta}_\tau \geq 0]}.$$

4.2 *Data augmentation*

In the Bayesian domain, inference on the BQR is efficiently solved by means of the *data augmentation* strategy, that essentially allows to borrow in the discrete case the well-established estimation framework for the QR of continuous responses, as described in [11] and [10]. Auxiliary variables have been demonstrated to be convenient for both sample space completion and prior specification, since they facilitate the construction of sampling-based algorithms to conduct approximate inference [48].

Tracing back to the seminal work of [26], in the absence of parametric restrictions on the error terms, frequentist estimates $\hat{\boldsymbol{\beta}}_\tau$ for the QR model (2) can be obtained as the solution of the following optimization problem

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}_\tau} \sum_{i=1}^{n} \rho_\tau(y_i^* - \mathbf{x}_i'\boldsymbol{\beta}_\tau), \tag{3}$$

where $\rho_\tau(u) = u(\tau - I_{[u<0]}) = (|u| + (2\tau - 1)u)/2$ is called *loss* or *check function*. [28] and [53] pointed out that the resulting τ-th *regression quantile* $\hat{\boldsymbol{\beta}}_\tau$ coincides with the maximum likelihood estimate (MLE) under independent ALD for the unobserved error terms. This finding inspired a series of works on QR modeling in the Bayesian framework where, instead, the specification of the likelihood is needed. Recent contributions to the Bayesian literature can be found in [45], [35], [32], [36], [4], [24], [3] and [13]. To implement the Bayesian inference, a convenient choice is the use of the exponential-gaussian mixture representation of the ALD, see [31], [40] and [32]. In particular, let $\epsilon \sim \mathrm{ALD}(\mu, \sigma, \tau)$ with density function given by

$$f_{\mathrm{ADL}}(\epsilon|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{\epsilon - \mu}{\sigma}\right)\right\} \qquad \epsilon \in \mathbb{R}, \tag{4}$$

where $\tau \in (0, 1)$ is the skewness parameter, $-\infty < \mu < +\infty$ is the location parameter reflecting both the mode and the τ-th quantile and $\sigma > 0$ is the scale parameter. By following [31] and setting $\theta = \frac{1-2\tau}{\tau(1-\tau)}$ and $p^2 = \frac{2}{\tau(1-\tau)}$, the random variable $\epsilon \sim \mathrm{ALD}(0, 1, \tau)$ admits the representation in terms of location-scale mixture of normals given by

$$\epsilon = \theta u + p\sqrt{u}\, z, \tag{5}$$

where $z \sim \mathrm{N}(0, 1)$ and $u \sim \mathrm{Exp}(1)$ are mutually independent and $\mathrm{N}(\cdot)$ and $\mathrm{Exp}(\cdot)$ denote, respectively, the Gaussian and the Exponential distribution. The substitution of the equality (5) in (2) leads to

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta}_\tau + \theta u_i + p\sqrt{u_i}\, z_i, \tag{6}$$

implying that $y_i^*|u_i, \boldsymbol{\beta}_\tau, \mathbf{x}_i \sim \mathrm{N}(\mathbf{x}_i'\boldsymbol{\beta}_\tau + \theta u_i, p^2 u_i)$. Thus, the mixture representation (6) expands the likelihood specification into a favorable hierarchical structure, whose fundamental advantage is the possibility to transfer the normal linear model framework to the QR approach. With the addition of equality (1), the same inferential scheme can be easily extended into the Bayesian BQR estimation for the analysis of dichotomous responses, see [24]. In the discrete case, in fact, one can augment the observed binary data $\mathbf{y}$ with both the latent vectors $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)$ and $\mathbf{u} = (u_1, \ldots, u_n)$ and write the complete-data likelihood as follows

$$L_c(\boldsymbol{\beta}_\tau, \mathbf{y}^*, \mathbf{u}) = \prod_{i=1}^{n} (y_i I_{[y_i^* \geq 0]} + (1 - y_i) I_{[y_i^* < 0]}) f(y_i^*|u_i, \mathbf{x}_i, \boldsymbol{\beta}_\tau) f(u_i|\mathbf{x}_i, \boldsymbol{\beta}_\tau).$$

The unique role played by the binary response is the truncation of the conditional normal density of y_i^* on the suitable range.

4.3 *Prior specification and model inference*

To complete the Bayesian model specification, a prior distribution for the parameter vector $\boldsymbol{\beta}_\tau$ has to be elicited. The data augmentation, combined with the mixture representation (6), allows to mimic the conjugate prior setting of the Bayesian normal linear model in the QR framework. This implies that the natural choice for the prior distribution of the regression coefficients is $\boldsymbol{\beta}_\tau \sim \mathbf{N}_{K+1}(\mathbf{b}_0, \mathbf{B}_0)$, that is, a $(K+1)$-variate normal distribution with prior mean vector $\mathbf{b}_0$ and covariance matrix $\mathbf{B}_0$. The conjugate analysis has the advantage of leading to closed-form solutions for the full-conditionals involved in the Gibbs sampling (GS), see [45] and [32]. These are recalled in the Appendix. In our model we opted for a weakly informative prior to minimize its impact on the final estimates. More details about the hyperparameters of the prior distribution are given in Section 5.3.

5. Empirical application

In this paper we work with individual data regarding students who enrolled in a post-reform first level course in the School of Economics of Sapienza University of Rome and attained their Bachelor degree in one of the a.y. from 2009-2010 to 2012-2013. The data set was provided by the administrative offices of Sapienza University of Rome and collects several information on the four cohorts of undergraduates including their personal characteristics, educational background and academic career. In this section we first describe the variables considered in the regression framework and the preliminary evidence emerging from the exploratory analysis and then discuss the results of the Bayesian BQR model.

5.1 *Definition of the variables*

Here we are interested in identifying which factors impact on the decision of dropping-out of the university after the achievement of the 3-year first level qualification. For this purpose, we constructed a binary response variable y_i for all $i = 1, \ldots, n$ as indicator of the non-continuation event. In more detail, we set $y_i = 1$ if student i is not enrolled in a new second level study programme at Sapienza within the two a.y. following the Bachelor degree attainment and $y_i = 0$ otherwise. Thus, our main focus is on the Bachelor-to-Master drop-out of a specific academic institution, rather than on the abandonment of the entire university system. The case $y_i = 1$, in fact, occurs when either the undergraduate decides to transfer to another institution or to leave the university system. The case $y_i = 0$, instead, denotes the transition to a Master degree course and includes both the retention in the same school and the enrollment in second level programme of another school within Sapienza. For each considered undergraduate cohort, the occurrence of the withdrawal event has been assessed in the two a.y. following the first level graduation. Notice that [18] considered the same time interval to define the transition decision in their analysis. We argue the choice of the two-year period with the fact that the observed proportion of students who enroll in a second cycle after more than two a.y. from the completion of the the first cycle career is negligible. Moreover, this evidence agrees with the national statistics on the Bachelor-to-Master progression reported in [5], indicating that almost all the transitions occur without any break period between the two cycles.

As described in Section 4, the observed binary response y_i is regarded as the result of a censoring mechanism acting on the latent response of interest y_i^*, which represents the unobserved quantitative measure of *propensity* or

Table 1.: Summary statistics.

Variable	**%**	**Variable**	**%**	**Variable**	**mean (s.d.)**
Bachelor-to-Master drop-out		*ISEE*		*Age at graduation*	24.1 (3.53)
No	72.2	[0,10000)	18.4	*WAM*	24.2 (1.86)
Yes	27.8	[10000,20000)	26.9	*High school mark*	0.8 (0.12)
Gender		[20000,30000)	20.5		
Female	54.0	≥30000	34.3		
Male	46.0	*Degree class*			
A.y. of graduation		Busin. Manag.	86.9		
2009-2010	20.8	Economics	13.1		
2010-2011	25.7	*MD*			
2011-2012	27.0	509/99	26.6		
2012-2013	26.5	270/04	73.4		
Citizenship		*Lyceum*			
Italian	95.4	No	39.5		
Other	4.6	Yes	60.5		
Place of residence		*Out-of-course*			
Rome	70.8	No	31.7		
Other	29.2	Yes	68.3		

Source: administrative offices of Sapienza University of Rome

utility to drop-out **after the Bachelor degree. In addition, the ALD distribution assumption** (4) **induces the presence of an additional latent variable** u_i **resulting from the mixture representation described in equation** (6)**, that contributes to make the inferential process tractable.**

As potential predictors of the latent utility to drop-out we considered variables concerning different aspects of student's individual profile. In line with the existing literature, our model specification includes personal characteristics, such as gender (1 = Male, 0 = Female), age at first level graduation, citizenship (1 = Italian, 0 = Other) and place of residence (1 = Rome, 0 = Other). The administrative data provide also information on the family financial condition expressed by means of the Equivalent Economic Situation Indicator (ISEE). This index is suitably rescaled to allow comparisons between family units that differ in size and composition. Similarly to [8] and [9], in our analysis we categorized the ISEE values into four classes, respectively 0 € $\leq$ ISEE $<$ 10000 € (reference category), 10000 € $\leq$ ISEE $<$ 20000 €, 20000 € $\leq$ ISEE $<$ 30000 € and ISEE $\geq$ 30000 €. Unfortunately, further information on the household context are not collected by the administration. This prevented us to assess the effect of parental characteristics, such as the educational background and the occupational condition, although in previous works **these variables were found to be associated with university drop-out,** as discussed in [17], [19], [20] and [2].

Regarding student's educational background, we considered a binary variable equal to 1 if the student granted the diploma in a classical or scientific lyceum and equal to 0 for other types of high secondary schools. With this dichotomization we stress the fundamental distinction between secondary schools that provide a more general and theoretical preparedness, such as the lyceums, and the technical or professional institutions, that are mainly oriented to a vocational training for the future exercise of specific professions. Besides the type of high school attended before the enrollment in the university, students were also asked by the administration to specify their high school leaving mark, that in the present analysis has been rescaled in the interval [0.6, 1].

The remaining predictors, instead, concern different aspects of student's academic career, such as a contextual variable, the field of study and the academic performance. **Regarding the first aspect, since the new MD 270/04 came into effect in**

the a.y. 2008-2009 and impacted to a substantial extent on students' study plans in the School of Economics, we employed an indicator variable defined as 1 = MD 270/04 and 0 = MD 509/99 to account for the set of academic regulations that the student underwent. The MD variable is of primary interest in the present analysis because it reflects the set of regulations adopted at national level for the organization of the degree courses in the Italian university system and, consequently, could indicate how the subsequent changes made to the pioneering reform of the Bologna Declaration have impacted on the Bachelor-to-Master transition phenomenon.

The field of study is a classical predictor of academic outcomes. We stress that the School of Economics at Sapienza University of Rome offers a wide range of 3-year degree courses. By exploiting the grouping of the degree courses with similar learning objectives into homogeneous degree classes **(as prescribed by the MD 509/99 and subsequently maintained by the MD 270/04)**, the 3-year degree courses of the School of Economics are formally categorized into two main classes: Economics and Business Management. We adopted this classification to characterize the field of study through a binary variable such that 1 = Economics and 0 = Business Management.

The role of the academic performance on the choice to leave the university is widely debated in the drop-out literature, although it seems to depend on the specific context, as demonstrated by the heterogeneous evidence collected so far. With the achievement of the Bachelor degree, the final graduation mark could be regarded as the candidate measure of student's performance. Indeed, the final grade is obtained from the combination of two main components: the average mark of passed exams and the actual duration of the studies. The former variable involves the credits described in Section 2 and it is equal to the weighted average of the exam grades with weights given by the corresponding amount of credits. The latter variable is, instead, related to an aspect of great concern for the Italian academic institutions, known as *out-of-course phenomenon*. Formally, a student is classified by the administration as *out-of-course graduate* when he/she attains the degree qualification in a time greater than the legal duration of the study course. Unlike other European university systems, where the students have to pass the exams within the prescribed period (or at most with a short delay) in order to successfully conclude their studies, in Italy they can continue to be enrolled in the university and graduate after the scheduled period as long as they pay the due fees. Thus, the out-of-course condition can be coded with a dummy variable equal to 1 for out-of-course undergraduates and to 0 for students who earned the Bachelor degree within the scheduled duration. In order to better capture different aspects of student's ability, we preferred to work with the weighted average mark (WAM) of passed exams and the binary indicator of the out-of-course graduation, rather than with the single final degree mark. In this way we can discriminate how the level of acquired knowledge and the fulfillment of the scheduled time limit separately act on the latent response. Additionally, we considered an interaction term between the WAM and the out-of-course condition to account for the major difficulty of exhibiting a good academic performance and graduating on time rather than after a longer period of study.

Finally, we examined also the possible presence of a time trend, or 'cohort effect', by involving in the linear predictor a categorical variable with four levels indicating the a.y. of graduation, where the a.y. 2009-2010 is assumed as the reference category.

5.2 *Exploratory analysis*

The data set provided by the administrative offices of Sapienza University of Rome contains a total of $n = 2655$ individual profiles relative to four cohorts of Bachelor graduates. As shown by the summary statistics reported in Table 1, there is a slight majority of

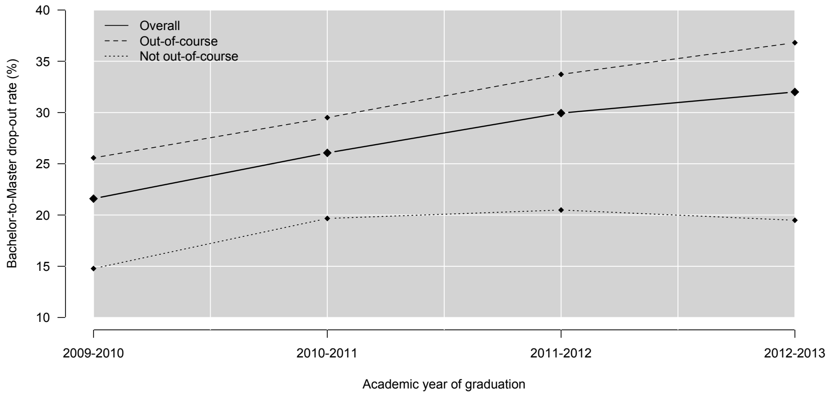

Figure 2.: Bachelor-to-Master drop-out rate (%) by academic year of graduation (solid line) and conditionally on the out-of-course condition (dashed lines).

females (54.0%) among undergraduates, the most part (70.8%) of them is resident in Rome and only 4.6% is a non-Italian citizen. Regarding the educational background, 60.5% of the undergraduates attended a lyceum (classic or scientific) and the rescaled average high school leaving mark is 0.8.

In the considered four-year period, the Business Management study field largely prevails on the Bachelor degree in Economics, with 86.9% of the students who achieved the qualification in this degree class. Moreover, 73.4% of the Bachelor graduates underwent the academic regulations established by the MD 270/04 during their first level career. The mean age at first level graduation is 24.1 years and less than one third (31.7%) of the students completed the first cycle within the legal duration of the degree programme. These values on the average age at graduation and the regularity of the first cycle studies are in line with the national indicators reported in [5].

As far as the response variable is concerned, the overall Bachelor-to-Master transition rate observed at the School of Economics in the considered period is 72.2%, that turns out to be comparable with the progression percentages estimated at national level for the Economics-Statistics area (Figure 1). Only few cases of school change within Sapienza have been observed, indicating a high correspondence of the second cycle studies with the first level career. An analogous tendency for the Economics-Statistics area of study

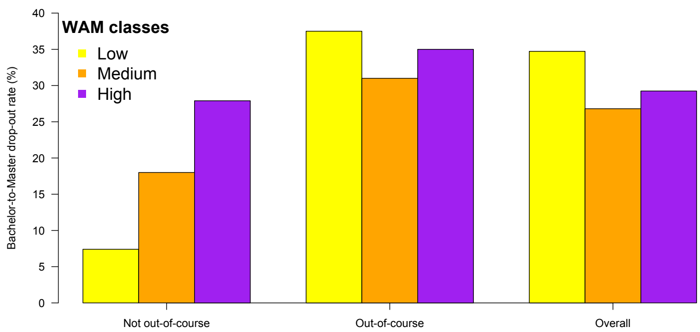

Figure 3.: Bachelor-to-Master drop-out rate (%) by WAM classes (right) and conditionally on the out-of-course condition (left and center).

is highlighted in [15], showing that a massive percentage of Bachelor graduates in this field of study progress at Master level in the same school.

Consequently, in the considered period more than one forth (27.8%) of the undergraduates decided to retire from Sapienza after the achievement of the Bachelor degree. The interest in the non-continuation decision is reinforced by the fact that the School of Economics experienced a manifest increment of the withdrawal phenomenon during the reference period, with a 10-points growth of the Bachelor-to-Master drop-out proportion from 21.6% for the cohort 2009-2010 up to 32.0% for the cohort 2012-2013 (Figure 2).

Table 2 shows the non-continuation proportions by specific undergraduates' characteristics, employed as preliminary tool to explore the association with the covariates. A first notable aspect to be stressed is that the drop-out decision appears to be strongly associated with the out-of-course condition, since students who graduated after the scheduled end of the course more frequently choose to interrupt their educational path in Sapienza after the Bachelor degree than those who regularly conclude the first cycle. This evidence is remarked by the large gap between the dashed lines in Figure 2, pointing out that this differential pattern characterizes all the four Bachelor cohorts and, indeed, the difference between out-of-course and regular undergraduates visibly increases over time.

Interestingly, a remarkable differential evidence emerges also for the variables describing the type of first level qualification. The non-continuation percentage is considerably higher for Bachelor graduates in Economics (36.2%) than for those in Business Management (26.5%). Additionally, students subject to the MD 509/99 regulations during their first level career exhibit a greater Bachelor-to-Master drop-out rate (34.7%) than students undergone the MD 270/04 (25.3%).

To facilitate the preliminary analysis of the relation between the continuous WAM variable and the binary outcome, we transformed the WAM of passed exams into an ordered factor with three levels reflecting, respectively, a low, medium and high academic performance and then computed the drop-out percentage within each class. The barplot on the right in Figure 3 (Overall) indicates higher withdrawal rates for the extreme WAM classes and, hence, the presence of a marginal non-monotonic relation with the outcome. Interestingly, by conditioning on the out-of-course condition, a similar pattern is observed also for the out-of-course undergraduates (barplot in the center), whereas a clear positive influence of the WAM is revealed for regular Bachelors (barplot on the left). This finding further supports the possible usefulness of an interaction term between the two variables in the regression equation to gain a more in-depth understanding of the academic performance as potential determinant of the non-continuation event.

Regarding the personal characteristics and the educational background, Table 2 points out a greater propensity to interrupt the studies among males, foreign students and among those who do not come from a lyceum.

By exploring also the association between the response variable and the household economic condition, one can observe that the first three ISEE classes have a comparable drop-out rate around 26%, whereas the non-continuation percentage among undergraduates whose ISEE is over 30.000 € is higher (31.3%).

Of course, all the above preliminary findings provide only a crude description of the interruption decision. To suitably enquire into the effect of the explanatory variables on the withdrawal event, a multiple regression analysis is needed.

5.3 *Results from the Bayesian BQR analysis*

For the analysis of the Bachelor-to-Master transition we estimated a series of Bayesian BQR models by varying the quantile level τ in the regular grid $\{0.05, 0.10, \ldots, 0.90, 0.95\}$. **Model estimation has been performed in R [44] with the recently released**

Table 2.: Bachelor-to-Master drop-out rates (%) by specific characteristics.

Variable	**%**	**Variable**	**%**
Out-of-course		*ISEE*	
No	18.6	[0,10000)	26.6
Yes	32.0	[10000,20000)	25.8
Gender		[20000,30000)	25.4
Female	25.3	≥30000	31.3
Male	30.7	*Degree class*	
A.y. of graduation		Busin. Manag.	26.5
2009-2010	21.6	Economics	36.2
2010-2011	26.1	*MD*	
2011-2012	29.9	509/99	34.7
2012-2013	32.0	270/04	25.3
Citizenship		*Lyceum*	
Italian	27.5	No	31.1
Other	33.9	Yes	25.6
Place of residence			
Rome	27.6		
Other	28.2		

Source: administrative offices of Sapienza University of Rome

package `bayesQR` developed by [12]. We adopted the default prior specification implemented by the command `bayesQR` with $\mathbf{b}_0 = \mathbf{0}$ and $\mathbf{B}_0 = 100\mathbf{I}_{K+1}$, corresponding to independent and diffuse normal distributions for the components of $\boldsymbol{\beta}_\tau$. In the GS application, algorithm convergence was assisted by the routines of the R package `coda` [41] and, specifically, by checking the mixing of the trace-plots and the extent of the sample autocorrelation for each parameter. In this respect, a certain amount of autocorrelation was detected in the posterior samples obtained with a preliminary GS application, likely due to the presence of latent variables in the model ([50] and [34]). So, we implemented the following procedure to reduce the extent of autocorrelation: first, we separately ran three MCMC chains and for each we retained 10000 drawings after a burn-in phase of 1000 iterations; subsequently, the three chains have been combined into a single sample of 30000 values, that was finally thinned by retaining a drawing every 3-th value (thinning). This resulted in a total of 10000 drawings from the posterior distribution employed for the computation of the final estimates.

The inferential results of the BQR can be conveniently summarized by means of the forest plots constructed on the normalized coefficient estimates of each explanatory variable, as displayed in Figure 4. As recall by [29], in fact, the regression parameter vector $\boldsymbol{\beta}_\tau$ in the BQR model is identified up to a scale, see also [38] for an in-depth explanation and treatment of the identifiability issue. Thus, in order to allow for comparisons among different conditional quantiles, [29] suggests to unify the scale by normalizing the slope parameters with the Euclidean norm $||\cdot||$. By denoting with $\boldsymbol{\beta}_\tau^{-c}$ the regression parameter vector deprived of the intercept $\beta_{\tau 0}$, we considered posterior inference on $\boldsymbol{\beta}_\tau^{-c}/||\boldsymbol{\beta}_\tau^{-c}||$. For each specific predictor, the forest plot simultaneously describes the point and interval estimate of the corresponding normalized regression coefficient as a function of the quantile level τ (Figure 4). Specifically, dots in the forest plots represent the posterior means, whereas segments illustrate the 95% Highest Posterior Density (HPD) credible intervals. **For a given credibility level $(1-\alpha) \in [0,1]$, the $100(1-\alpha)\%$ HPD interval provides the most probable parameter values of the posterior distribution that collectively cover $100(1-\alpha)\%$ of the posterior probability.**

Regarding the degree class, the coefficient trend illustrated in Figure 4 points out a positive and significant effect for all the considered quantiles, confirming an important difference between the two study courses: after attaining the Bachelor degree in Econometrics, students are more likely to drop-out than after the first level studies in Business Management. One possible interpretation could be the presence of competitive and appealing Master programmes **in Economics in other universities, motivating students to change institution.**

Regarding the academic performance, firstly we can note that the regularity of the study path is related with the continuation decision. Specifically, students who do not complete their studies on time have a higher probability to retire from the institution, as indicated by the interval estimates largely above the dashed line in the corresponding plot of Figure 4. Moreover, the addition of the interaction term in the BQR model provides a very enlightening interpretation on the impact of the WAM on the binary outcome, suggesting a different effect of the performance variable for out-of-course and regular students. The significant and positive estimates of the coefficient associated to the WAM regressor reveal that, for almost all quantile levels, this variable influences the interruption choice of regular students (reference group). This means that, among students who fulfill the regular duration of the programme, those who conclude their first level studies with a higher WAM have a greater probability of dropping out after the Bachelor degree. Several conjectures can be formulated about the estimated effect of student's performance. A greater Bachelor-to-Master drop-out rate for larger WAM values could be explained with the fact that brilliant students have a higher chance to find a job after the first cycle studies or could be even employed before concluding their 3-year degree course. On the other hand, best-performing students take their decisions with a more careful and critical evaluation; thus, if they realize that other institutions offer a training programme matching their expectations, they are more prone to move. By summing the WAM coefficient for the reference group with the interaction estimate, one obtains the effect of the WAM for out-of-course undergraduates, displayed at the bottom of Figure 4. The latter regression coefficient turns out to be not statistically significant over the full range of τ values, meaning that the level of acquired competence does not modify the drop-out decision of out-of-course students. In the absence of the interaction term, this important difference on the relevance of the academic performance in the two groups of students would have been missed.

Unlike previous studies emphasizing the role of the type of diploma and of the high school leaving mark, in our analysis these variables do not emerge as relevant predictors of the progression at Master level. It is possible, in fact, that the educational background exerts an effect on the continuation decision in the first phase of the university career, that in general immediately follows the high school experience, but not on the transition to the second cycle studies.

Regression coefficients associated to the cohort indicators clearly show the growing trend of Bachelor-to-Master drop-out rate in the considered period and, in particular, a significant increase for the last two Bachelor cohorts.

The coefficient for the age at graduation exhibits a significant positive value across all quantiles with a manifest growing trend. This means that age at graduation not only contributes to explain the drop-out decision, but its relevance is even stronger for students with a high propensity to leave the university after the first cycle. The positive effect of age could be argued with the fact that older students could desire to not further delay their entry in the labour market or could be already employed and, hence, less motivated to put effort into a new degree course.

The forest plot in Figure 4 regarding the estimates for the MD indicator is the one that better exemplifies the usefulness of the BQR approach. The MD 270/04 is estimated to have a negative effect on the entire latent non-continuation propensity, but its magnitude and significance considerably vary across quantiles. By moving from higher to lower quantiles, one has that the estimates associated to the MD 270/04 remarkably decrease and vary from not statistically significant values to significant effects, specifically on the quantiles below the median. Moreover, the influence of the MD 270/04 becomes more and more pronounced on the left tail of the underlying distribution. This implies that students with lower propensity to withdraw are sensible to the type of academic regulations when they have to choose whether to continue at Master level or not and, in particular, those who experienced the MD 270/04 are less likely to drop-out after the first level qualification. To a reduced extent, a similar monotone behavior of the posterior estimates can be also highlighted for the effect of the gender. Being male affects positively the drop-out probability but its impact is significant only for quantile levels above 0.75. This suggests that the difference between males and females mainly concerns students with high propensity to interrupt. The Italian citizenship is estimated to have a negative effect that becomes stronger over higher quantiles. However, the corresponding coefficient is significant only on the 90-th and 95-th percentile. This borderline evidence could be due to very low prevalence of foreign students in the present study. In accordance with the preliminary analysis, the BQR approach does not indicate a difference in the continuation behavior for students belonging to the first three ISEE classes, whereas undergraduates with a household economic indicator at least equal to 30000 € have a significant higher probability to drop-out. Finally, the place of residence turns out to be not significant in the description of students' continuation choice for the present study, although the negative sign is consistent with several drop-out analyses stating that being resident in the city where the university is located facilitates the continuation of the studies.

5.4 *Comparison with alternative methods for binary outcomes*

In order to show the usefulness of the BQR in capturing the Bachelor-to-Master drop-out features, we estimated also alternative binary regression models, all considering the conditional mean of the response variable as function of the same set of regressors. Specifically, the competing methods involve Bayesian and frequentist approaches given by: (i) Bayesian logistic regression model with diffuse priors; (ii) binary regression model with cauchit link; (iii) heteroskedastic probit model with the regressor "Age at graduation" acting on the variance of the error distribution and (iv) power logit model based on the asymmetric logit link proposed by [6].

Interestingly, the corresponding inferential results reported in Table 3 show that none of the competing conditional mean models highlights a significant role of the type of academic regulations on the continuation decision. These findings confirm the importance of having more information on the whole response distribution, rather than on the mean value only, to better characterize the phenomenon of interest.

6. Concluding remarks and future developments

The reforms of the Bologna process have transformed the academic training in a stepwise path, whose multi-cycle structure is the core feature. Although a wide literature concerns the problem of university drop-out referred to the very first years after the enrollment, very little has been explored with respect to the transition from the first to the second cycle academic studies. In this paper we described a BQR model from the Bayesian

Table 3.: Estimation results of alternative binary regression models fitted to the data set of Bachelor graduates at the School of Economics of Sapienza University of Rome: posterior means and 95% HPD credible intervals for the Bayesian logistic regression model, MLE and 95% confidence intervals (CI) for the frequentist methods.

	Bayesian Logit		**Cauchit**	
	$\hat{\beta}$	95% HPD	$\hat{\beta}$	95% CI
Intercept	-9.081	[-11.750,-6.357]*	-11.757	[-15.932,-7.871]*
A.y. of graduation: 2010-2011	0.355	[0.048,0.643]*	0.392	[0.042,0.757]*
A.y. of graduation: 2011-2012	0.620	[0.301,0.929]*	0.670	[0.305,1.055]*
A.y. of graduation: 2012-2013	0.711	[0.378,1.029]*	0.806	[0.430,1.201]*
Gender: Male	0.221	[0.037,0.418]*	0.236	[0.027,0.447]*
Out-of-course: Yes	4.565	[1.736,7.515]*	5.856	[2.068,9.817]*
Degree class: Economics	0.429	[0.175,0.680]*	0.413	[0.144,0.674]*
Age at graduation	0.157	[0.122,0.191]*	0.204	[0.156,0.259]*
Citizenship: Italian	-0.265	[-0.714,0.186]	-0.124	[-0.633,0.463]
Place of Residence: Rome	-0.204	[-0.401,0.003]	-0.257	[-0.475,-0.036]*
Lyceum: Yes	-0.094	[-0.298,0.110]	-0.075	[-0.303,0.155]
ISEE: [10000,20000)	-0.017	[-0.291,0.273]	-0.021	[-0.333,0.302]
ISEE: [20000,30000)	-0.020	[-0.337,0.280]	-0.082	[-0.439,0.274]
ISEE: $\geq$30000	0.356	[0.085,0.643]*	0.347	[0.044,0.669]*
MD: 270/04	-0.274	[-0.524,0.001]	-0.267	[-0.544,0.013]
High school mark	0.190	[-0.667,1.109]	0.330	[-0.680,1.339]
WAM	0.150	[0.053,0.255]*	0.199	[0.060,0.341]*
Out-of-course: Yes × WAM	-0.168	[-0.287,-0.054]*	-0.214	[-0.368,-0.064]*
	Heteroskedastic Probit		**Power Logit**	
	$\hat{\beta}$	95% CI	$\hat{\beta}$	95% CI
Intercept	-16.184	[-27.128,-5.240]*	-10.110	[-13.066,-7.213]*
A.y. of graduation: 2010-2011	0.416	[-0.090,0.921]	0.363	[0.043,0.686]*
A.y. of graduation: 2011-2012	0.801	[0.161,1.442]*	0.652	[0.313,0.996]*
A.y. of graduation: 2012-2013	0.857	[0.181,1.533]*	0.733	[0.381,1.089]*
Gender: Male	0.361	[0.005,0.717]*	0.246	[0.041,0.451]*
Out-of-course: Yes	6.541	[1.046,12.036]*	4.905	[1.851,7.985]*
Degree class: Economics	0.662	[0.119,1.205]*	0.473	[0.193,0.751]*
Age at graduation	0.325	[0.095,0.554]*	0.187	[0.147,0.230]*
Citizenship: Italian	-0.551	[-1.326,0.224]	-0.327	[-0.817,0.170]
Place of Residence: Rome	-0.327	[-0.692,0.038]	-0.223	[-0.440,-0.005]*
Lyceum: Yes	-0.120	[-0.449,0.209]	-0.097	[-0.322,0.130]
ISEE: [10000,20000)	-0.012	[-0.452,0.428]	-0.016	[-0.321,0.292]
ISEE: [20000,30000)	0.030	[-0.444,0.504]	-0.012	[-0.343,0.320]
ISEE: $\geq$30000	0.589	[0.028,1.151]*	0.383	[0.080,0.690]*
MD: 270/04	-0.118	[-0.559,0.322]	-0.253	[-0.546,0.041]
High school mark	0.442	[-0.982,1.866]	0.225	[-0.738,1.190]
WAM	0.243	[0.041,0.444]*	0.166	[0.060,0.272]*
Out-of-course: Yes × WAM	-0.245	[-0.461,-0.029]*	-0.181	[-0.305,-0.058]*

* indicates interval estimates not including zero

estimation perspective as effective device to describe the withdrawal event after the first level qualification. We approached the failed Bachelor-to-Master transition as a drop-out problem and tried to understand the main determinants of the non-continuation choice by exploiting the QR model setup. This method permits to highlight different student behaviors by means of a latent propensity scale measuring the utility to drop-out.

Especially in latent variable settings, Bayesian inference is addressed in a more straightforward manner and, unlike the frequentist approach, does not rely on asymptotic properties and computational demanding methods to assess estimation uncertainty. In the Bayesian framework, in fact, the final knowledge on the parameters of interest is completely formalized by the posterior distribution and one can take advantage of an easy derivation of credible intervals to supply well understandable measures of statistical un-

certainty. Moreover, as proved by [16], the frequentist estimation of extreme quantiles (τ very close to 0 or 1) could require specific inferential methods. This is because the number of observations in the tails of the response distribution could be insufficient to justify the use of traditional normal approximations, making Bayesian estimation a profitable alternative.

From an empirical point of view, one of the main contributions of this work concerns the assessment of a critical aspect of the Italian university context, that is, the out-of-course phenomenon, and the identification of the significant benefits of graduating on time on the continuation decision at Master level. Consequently, some factual interventions to reduce the out-of-course rate should be planned, such as a more accurate revision of the course programmes and of the corresponding credits and/or a different organization of the exam sessions during the a.y.. Interestingly, our BQR analysis also points out that the changes introduced by the MD 270/04 are associated with a reduction of the non-continuation rate after the Bachelor degree, that instead does not emerge from **alternative binary regression models describing the conditional mean of the response distribution**. The MD variable accounts for the evolution of the radical reform of 1999 in the Italian academic context and for its impact on students' decisions. Our results provide an important feedback at institutional level, since they concretely suggest to the relevant authorities how to improve the academic regulations and plan the future interventions. Obviously, the MD variable defined for the present Italian case can be suitably reinterpreted in other national contexts to assess the impact of the academic regulations in other countries.

A natural extension of the present Bachelor-to-Master drop-out analysis could be the application of the methods to the data of the entire university, as well as to the regional or national ones. In this regard, a generalization of the Bayesian BQR for handling the multilevel structure of the individual observations should be developed, due to the typical nested organization of the universities where degree courses are nested into departments which are in turn nested into schools. Some advantages of the multilevel modeling include the possibility to capture the heterogeneity among different degree courses or departments and to derive relative performance measures for comparisons purposes. Moreover, the multilevel approach allows to introduce course- or department-level covariates and to assess their impact on students' decisions.

Appendix

Gibbs sampling for the Bayesian binary quantile regression model

At the generic $(l+1)$-th GS iteration, the full-conditionals are

$$y_i^{*(l+1)}|u_i^{(l)}, y_i, \mathbf{x}_i, \boldsymbol{\beta}_\tau^{(l)} \sim \begin{cases} \mathbf{N}(\mathbf{x}_i'\boldsymbol{\beta}_\tau^{(l)} + \theta u_i^{(l)}, p^2 u_i^{(l)}) I_{[y_i^* \geq 0]} & \textbf{if } y_i = 1, \\ \mathbf{N}(\mathbf{x}_i'\boldsymbol{\beta}_\tau^{(l)} + \theta u_i^{(l)}, p^2 u_i^{(l)}) I_{[y_i^* < 0]} & \textbf{otherwise}; \end{cases}$$

$$u_i^{(l+1)}|y_i^{*(l+1)}, y_i, \mathbf{x}_i, \boldsymbol{\beta}_\tau^{(l)} \sim \mathbf{GIG}(1/2, (y_i^{*(l+1)} - \mathbf{x}_i'\boldsymbol{\beta}_\tau^{(l)})^2/p^2, 2 + \theta^2/p^2);$$

$$\boldsymbol{\beta}_\tau^{(l+1)}|\mathbf{y}^{*(l+1)}, \mathbf{u}^{(l+1)}, \mathbf{y}, \mathbf{X} \sim \mathbf{N}_{K+1}(\hat{\mathbf{b}}, \hat{\mathbf{B}}),$$

where GIG denotes the Generalized Inverse Gaussian distribution and

$$\hat{\mathbf{B}}^{-1} = \tau^{-2}\mathbf{X}'\mathbf{U}^{-1}\mathbf{X} + \mathbf{B}_0^{-1} \qquad \hat{\mathbf{b}} = \hat{\mathbf{B}}\left(\tau^{-2}\mathbf{X}'\mathbf{U}^{-1}(\mathbf{y}^{*(l+1)} - \theta\mathbf{u}^{(l+1)}) + \mathbf{B}_0^{-1}\mathbf{b}_0\right)$$

with $\mathbf{U} = \mathbf{diag}(\mathbf{u}^{(l+1)})$.

References

[1] J. Abrevaya and J. Huang, *On the bootstrap of the maximum score estimator*, Econometrica 73 (2005), pp. 1175–1204.
[2] C. Aina, *Parental background and university dropout in Italy*, Higher Education 65 (2013), pp. 437–456.
[3] R. Alhamzawi and K. Yu, *Conjugate priors and variable selection for Bayesian quantile regression*, Computational Statistics & Data Analysis 64 (2013), pp. 209–219.
[4] R. Alhamzawi, K. Yu, and D.F. Benoit, *Bayesian adaptive Lasso quantile regression*, Statistical Modelling 12 (2012), pp. 279–297.
[5] ANVUR, *Rapporto sullo stato del sistema universitario e della ricerca*, Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, 2014.
[6] F.J. Aranda-Ordaz, *On two families of transformations to additivity for binary response data*, Biometrika 68 (1981), pp. 357–363.
[7] G. Ballarino and M. Bratti, *Field of study and university graduates' early employment outcomes in Italy during 1995–2004*, Labour 23 (2009), pp. 421–457.
[8] F. Belloc, A. Maruotti, and L. Petrella, *University drop-out: an italian experience*, Higher Education 60 (2010), pp. 127–138.
[9] F. Belloc, A. Maruotti, and L. Petrella, *How individual characteristics affect university students drop-out: a semiparametric mixed-effects model for an Italian case study*, Journal of Applied Statistics 38 (2011), pp. 2225–2239.
[10] D.F. Benoit, R. Alhamzawi, and K. Yu, *Bayesian Lasso binary quantile regression*, Computational Statistics 28 (2013), pp. 2861–2873.
[11] D.F. Benoit and D. Van den Poel, *Binary quantile regression: a Bayesian approach based on the asymmetric laplace distribution*, Journal of Applied Econometrics 27 (2012), pp. 1174–1188.
[12] D.F. Benoit, R. Alhamzawi, K. Yu, and D. Van den Poel, *bayesQR: Bayesian quantile regression*, R package version 2.2 (2014), Available at http://cran.r-project.org/web/packages/bayesQR/bayesQR.pdf.
[13] M. Bernardi, G. Gayraud, and L. Petrella, *Bayesian tail risk interdependence using quantile regression*, Bayesian Analysis 10 (2015), pp. 553–603.
[14] L. Biggeri, M. Bini, and L. Grilli, *The transition from university to work: a multilevel approach to the analysis of the time to obtain the first job*, Journal of the Royal Statistical Society: Series A (Statistics in Society) 164 (2001), pp. 293–305.
[15] A. Cammelli, G. Antonelli, A. Di Francia, G. Gasperoni, and M. Sgarzi, *Mixed outcomes of the Bologna process in Italy*, in *Employability and mobility of Bachelor graduates in Europe*, Springer, 2011, pp. 143–170.
[16] V. Chernozhukov, *Inference for extremal conditional quantile models (extreme value inference for quantile regression)* (2002).
[17] P. Cipollone and F. Cingano, *University drop-out-the case of italy*, Bank of Italy Temi di Discussione (Working Paper) No 626 (2007).
[18] H. de Boer, R. Kolster, and H. Vossensteyn, *Motives underlying Bachelors–Masters transitions: The case of Dutch degree stackers*, Higher Education Policy 23 (2010), pp. 381–396.
[19] B. d'Hombres, *The impact of university reforms on dropout rates and students? status: Evidence from Italy*, Tech. Rep., JRC European Commission, 2007.
[20] G. Di Pietro and A. Cutillo, *Degree flexibility and university drop-out: The Italian experience*, Economics of Education Review 27 (2008), pp. 546–555.
[21] K. Florios and S. Skouras, *Exact computation of max weighted score estimators*, Journal of Econometrics 146 (2008), pp. 86–91.
[22] L. Grilli, *Multilevel models for the analysis of the transition from university to work*, in *VIII International Meeting on Quantitative Methods for Applied Sciences*, 2006.
[23] P.J. Huber, *Robust Statistics*, Wiley New York, 1981.
[24] Y. Ji, N. Lin, and B. Zhang, *Model selection in binary and Tobit quantile regression using the Gibbs sampler*, Computational Statistics & Data Analysis 56 (2012), pp. 827–839.
[25] R. Koenker, *Quantile regression*, 38, Cambridge University Press, 2005.
[26] R. Koenker and G.J. Bassett, *Regression quantiles*, Econometrica: Journal of the Econometric Society (1978), pp. 33–50.
[27] R. Koenker and K. Hallock, *Quantile regression: An introduction*, Journal of Economic Perspectives

15 (2001), pp. 43–56.

[28] R. Koenker and J.A. Machado, *Goodness of fit and related inference processes for quantile regression*, Journal of the American Statistical Association 94 (1999), pp. 1296–1310.

[29] G. Kordas, *Smoothed binary regression quantiles*, Journal of Applied Econometrics 21 (2006), pp. 387–407.

[30] A. Kottas and A.E. Gelfand, *Bayesian semiparametric median regression modeling*, Journal of the American Statistical Association 96 (2001), pp. 1458–1468.

[31] S. Kotz, T.J. Kozubowski, and K. Podgórski, *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*, Birkhäuser Boston, Inc., Boston, MA, 2001.

[32] H. Kozumi and G. Kobayashi, *Gibbs sampling methods for Bayesian quantile regression*, Journal of Statistical Computation and Simulation 81 (2011), pp. 1565–1578.

[33] M.S. Larsen, K.P. Kornbeck, R.M. Kristensen, M.R. Larsen, and H.B. Sommersel, *Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by the universities to prevent or reduce it? A systematic review*, Tech. Rep., Department of Education, Aarhus University, 2013.

[34] R. Leon-Gonzalez, *Data augmentation in the bayesian multivariate probit model*, Tech. Rep., Sheffield Economic Research Paper Series 2004001, University of Sheffield, Department of Economics, 2004.

[35] Q. Li, R. Xi, N. Lin, *et al.*, *Bayesian regularized quantile regression*, Bayesian Analysis 5 (2010), pp. 533–556.

[36] K. Lum, A.E. Gelfand, *et al.*, *Spatial quantile multiple regression using the Asymmetric Laplace Process*, Bayesian Analysis 7 (2012), pp. 235–258.

[37] C.F. Manski, *Maximum score estimation of the stochastic utility model of choice*, Journal of Econometrics 3 (1975), pp. 205–228.

[38] C.F. Manski, *Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator*, Journal of Econometrics 27 (1985), pp. 313–333.

[39] S. Nyüsti, V. Veroszta, *et al.*, *Institutional effects on Bachelor-Master-level transition*, International Journal of Social Sciences 4 (2015), pp. 39–61.

[40] T. Park and G. Casella, *The Bayesian Lasso*, Journal of the American Statistical Association 103 (2008), pp. 681–686.

[41] M. Plummer, N. Best, K. Cowles, and K. Vines, *CODA: Convergence diagnosis and output analysis for MCMC*, R news 6 (2006), pp. 7–11.

[42] J.L. Powell, *Least absolute deviations estimation for the censored regression model*, Journal of Econometrics 25 (1984), pp. 303–325.

[43] D. Pozzoli, *The transition to work for Italian university graduates*, Labour 23 (2009), pp. 131–169.

[44] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2014), Available at http://www.R-project.org/.

[45] C. Reed and K. Yu, *A Partially Collapsed Gibbs sampler for Bayesian quantile regression*, Tech. Rep., Department of Mathematical Sciences, Brunel University, 2009.

[46] H. Schomburg and U. Teichler, *Employability and mobility of Bachelor graduates in Europe*, Springer, 2011.

[47] J.P. Smith and R.A. Naylor, *Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students*, Journal of the Royal Statistical Society: Series A (Statistics in Society) 164 (2001), pp. 389–405.

[48] M.A. Tanner and W.H. Wong, *The calculation of posterior distributions by data augmentation*, Journal of the American Statistical Association 82 (1987), pp. 528–540.

[49] V. Tinto, *Research and practice of student retention: what next?*, Journal of College Student Retention: Research, Theory & Practice 8 (2006), pp. 1–19.

[50] D.A. Van Dyk and X.L. Meng, *The art of data augmentation*, Journal of Computational and Graphical Statistics 10 (2001), pp. 1–50.

[51] Z. Veroszta, *The way to master programmes – an examination of the selection mechanisms in the bachelor/master transition in higher education*, in *Hungarian Graduates 2011*, Educatio Public Services Non-profit LLC, Department of Higher Education, 2013, pp. 9–35.

[52] K. Yu, Z. Lu, and J. Stander, *Quantile regression: applications and current research areas*, Journal of the Royal Statistical Society: Series D (The Statistician) 52 (2003), pp. 331–350.

[53] K. Yu and R.A. Moyeed, *Bayesian quantile regression*, Statistics & Probability Letters 54 (2001), pp. 437–447.

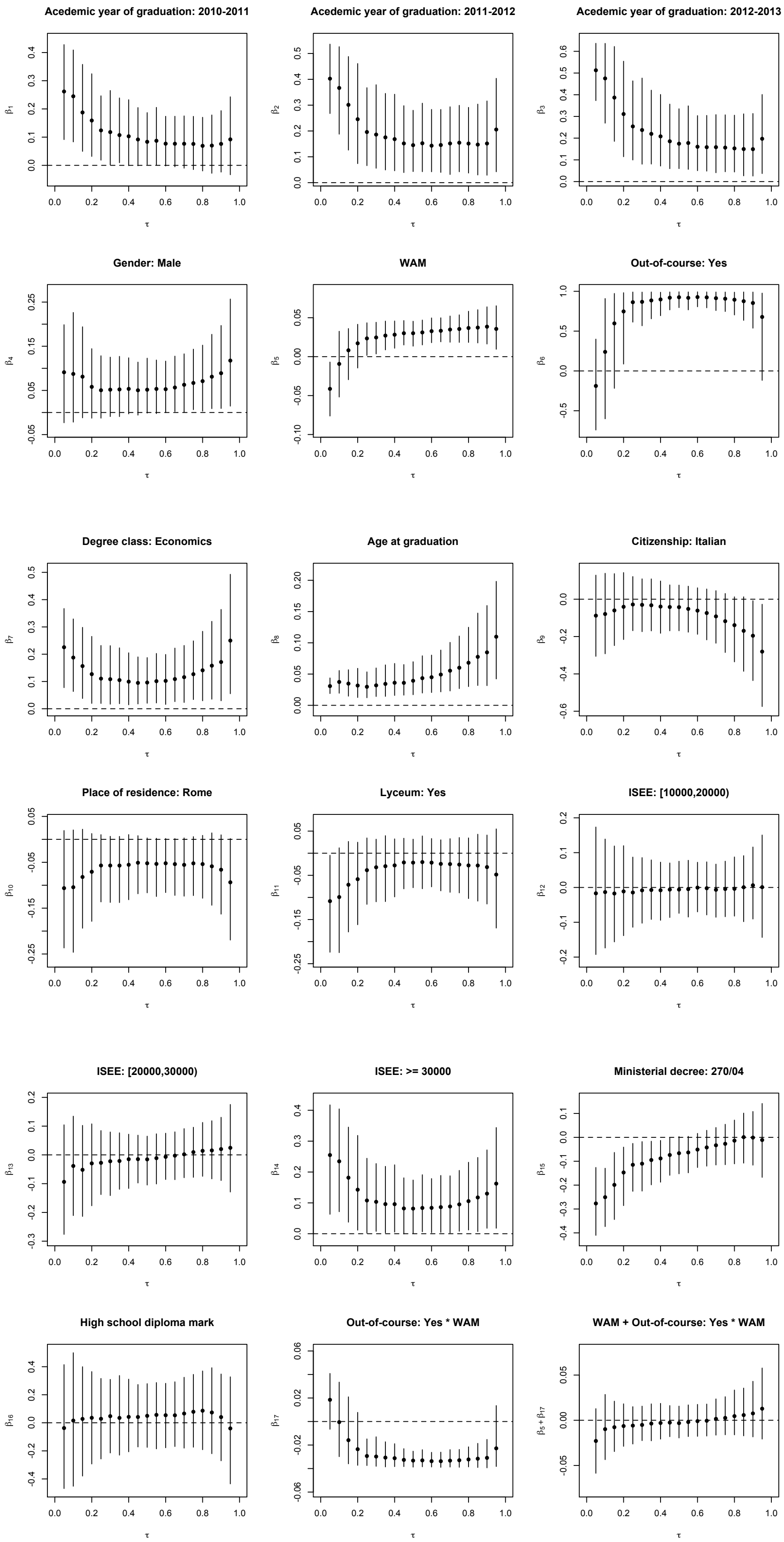

Figure 4.: Forest plots of the posterior estimates of the BQR parameters for selected values of the quantile level τ. Dots indicate the posterior means and segments represent the corresponding 95% HPD credible intervals.