Degree course change and student performance: a mixed-effect model approach

Giada Adelfio and Giovanni Boscaino*

Dipartimento di Scienze Economiche, Aziendali e Statistiche, Università degli Studi di Palermo, Palermo, Italia

This paper focuses on students credits earning speed over time and its determinants, dealing with the huge percentage of students who do not take the degree within the legal duration in the Italian University System. A new indicator for the performance of the student career is proposed on real data, concerning the cohort of students enrolled at a Faculty of the University of Palermo (followed for 7 years). The new indicator highlights a typical zero-inflated distribution and suggests to investigate the effect of the degree course (DC) change on the student career. A mixed-effect model for overdispersed data is considered, with the aim of taking into account the individual variability as well, due to the longitudinal nature of data. Results show the significant positive effect of the DC change on the student performance.

Keywords: university credits; expected years to the graduation; overdispersion; ZIP model; longitudinal data; motion chart

1. Introduction

Academic student performance is a crucial issue for university policy-makers, today more than ever. Universities monitor students careers in order to fix some issues related to the educational process and to apply policies that can improve students performance. That is crucial especially for those University Systems where there is not a time limit to graduation. In fact, most of those systems are often affected by a huge number of students that graduate over the legal duration. That has a negative effect both on the whole society (e.g. students emancipate and embark on job market at older age) and on the university management (e.g. in many systems, universities are financially penalized because the students take the degree over the legal duration). Generally, previous studies showed that performance is a complex phenomenon often affected by aspects related to students personal inclinations and to their subjective choices during the path to graduation. Recent literature offers several papers about student performance and its determinants, and results are not always in the same direction. A first rough classification of papers could distinguish between papers accounting for students social and demographic characteristics and those

*Corresponding author. Email: giovanni.boscaino@unipa.it

accounting for their previous performance and/or psychological and subjective features. Just to quote some, in the first group Cheesman *et al.* [6] applied a regression analysis to describe students performance – measured by four categories of graduation marks – singling out that gender, enrolment status, faculty, finance assistance, and residence are likely determinants. Birch and Miller [4] used a quantile regression approach identifying in the Australian Tertiary Entrance Rank, Gender, and High School the most important determinants of high and low performance. Boscaino *et al.* [5] focused on students who never earned credits after four years, using a zero-inflated model and singling out different social demographic profile for different performance levels, based on gender, high school, and income level. Grilli *et al.* [10] introduced pre-enrolment assessment test outcomes, together with some personal student characteristics, on earned credits at the end of the first year, using different models (hurdle and binomial mixture model): they highlighted the poor role on the pre-enrolment test as predictor of number of earned credits.

In the second group, for example, Tattersall *et al.* [23] measured the educational efficiency in terms of comparison between inputs and outputs. The output–input ratio was analysed including several aspects of the students path to graduation, for example, learning interruption and Changes of the Study Programme. Van Bragt *et al.* [24] followed a more social psychological root making a deeper analysis of performances: they also included an ad hoc survey to investigate the impact of the big five personality characteristics, personal learning orientations, and students study approach on their performance, using a logistic model. Results show a positive effect of conscientiousness and a negative one of ambivalence and lack of regulation. Horn *et al.* [14] performed an exploratory analysis on the determinants of success of second-year students. The authors asked if the factors that lead to the success at the end of the first year could rule the performance of the second year: they discovered that Lectures and Tutorial attendance were still significant factors, and the most important determinant was the performance during the first year. Attanasio *et al.* [3] highlighted the crucial role of the credits earned at the end of the first year as a good and simple predictor of the success, in a retrospective exploratory study. Adelfio *et al.* [2] proposed a new measure for student performance. Using a quantile regression, results showed no significant effect of the social and demographic variables. The only significant one was the attended degree course (DC).

In conclusion, literature results generally seem to suggest that the impact of the determinants varies (in terms of extent and direction) according to the context (economic, social, political, demographic, space-time, etc.) and results should hold just in that context. In fact, with respect to the University of Palermo (Italy), our recent studies highlighted that social and demographic characteristics do not seem to affect students careers, which seem only to be affected by their personal inclinations [1].

In the light of these results, this paper follows a different point of view, investigating the role of the DC change. In other words, our question is: Does the student who change DC improve his performance? Indeed, in the Italian University System (IUS), students can change the DC whenever they want: on the one hand, students can fix a wrong decision about their educational path (reducing the time to graduation), on the other hand the DC change could affect negatively the time to graduation (e.g. because they have to restart their career). In particular, in this paper we focus on the undergraduate first-level students that are still enrolled over the legal duration of studies (i.e. over the third academic year), hereafter out-of-time (OOT) students, in order to investigate the DC change effect on the expected time to graduation. In IUS, OOT percentage is very huge and seems to increase over time: in spite of several reforms, between 2005 and 2010 the amount of OOT graduated at the first-level DC is always over the 50% of the graduated, rising to 70% in 2010 [19].

In the first part of the paper, rationale and data are introduced. In Section 2, a new indicator is introduced, in order to study how the performance of students varies during time and by effect of the DC change. In Section 3, data are analysed following a descriptive perspective. In the second part of the paper data are analysed using a mixed-effect model for overdispersed data, since we deal with longitudinal data. In particular, in Section 4 the referred model is reported, while Section 5 is devoted to estimation and interpretation of results, together with some concluding remarks.

2. Rationale

The measure of performance of students career is usually based on the amount of credits earned at the end of each academic year and/or on the number of years they need to get the degree. In this paper, a student can be seen as an object which moves along a path, that is his career. In Physics, the object motion is measured in terms of travelled space, needed time, and therefore its velocity. Speed is the ratio between space and time. In our case, space is measured in credits (CR) and time is the academic year (AY). According to this perspective, we are following Attanasio *et al.* [3] that considered the mean velocity in the physic way, that is the total distance travelled divided by the total time elapsed:

$$\upsilon_i(J) = \frac{CR_i(J)}{J}, \tag{1}$$

where $\upsilon_i(J)$ is the ith student average speed kept from the enrolment up to the jth AY (i.e. for first J AY), measured in CR/AY. Equation (1) is then used to estimate the expected duration of the whole student career (EY_i) after J years, in terms of academic years:

$$EY_i(J) = \frac{180 - CR_i(J)}{\upsilon_i(J)} + J, \tag{2}$$

where 180 is the usual total amount of CR that a first-level student has to earn to get the degree, and it can be considered as the length of the path that the student has to travel.

Equation (2) could be a tool for those who want to monitor the students career, trying to answer to the question 'If the student keeps his average speed recorded during the first J years, how long his whole career will be?'. In our opinion, since the mean is not able to pick-up the spread (balancing good and bad performances recorded at different years), the question 'If the student keeps the velocity recorded during the last year, how many further years does he need to get the degree?' should be more interesting.

Therefore, in this paper, we focus on the expected remaining time on the basis of the 'instantaneous' velocity recorded at the end of each jth AY:

$$\upsilon_i^*(j) = \frac{\Delta CR_i(J)}{\Delta J} = \frac{CR_i(J) - CR_i(J-1)}{J-(J-1)} = CR_i(j), \tag{3}$$

where $\upsilon_i^*(j)$ is again measured in CR/AY.

EY_i now becomes

$$EY_i^*(j) = \frac{180 - CR_i(J)}{\upsilon_i^*(j)} \tag{4}$$

that represents the expected number of years that the ith student has to wait to get the degree if he maintains the same velocity equals to the instantaneous speed recorded during the jth year. Obviously, EY^* is updated every year and, besides catching the variation in the expected years to the graduation after j years, it reveals if the average speed has changed.

An example is reported in Table 1. Following the Attanasio perspective, if we consider that a given student keeps constant the average speed recorded for the first two AY ($v_i(J=2) = 25$ CR/AY), his whole career duration estimate is 7.20 AY. Following the new perspective, if for the next AY the student will keep constant the instantaneous velocity recorded at the second

Table 1. An example of υ and EY variation over years for the two methods.

AY (j)	$CR_i(j)$	$CR_i(J)$	$\upsilon_i(J)_{CR/AY}$	$EY_i(J)$	$\upsilon_i^*(j)_{CR/AY}$	$EY_i^*(j)$
1	10	10	10	18.00	10	17.00
2	40	50	25	7.20	40	3.25
3	20	70	23	7.78	20	5.50
4	60	130	32	5.56	60	0.83
5	50	180	36	5.00	50	0.00

AY ($v_i^*(j = 2) = 40$ CR/AY), it is expected that he needs further three AY and four months to complete his path. Therefore, the variation in $EY_i(J)$ and in $EY_i^*(j)$ are both informative of the student career performance, following two different perspectives. Indeed, we are proposing EY^* as an alternative measure to EY and not as a better one. As we stated, EY^* seems to be a more useful tool for the policy-makers (either tutor or student mentor) in order to monitor the performance of the student. In our opinion, in the earning credits process, keeping the velocity of a good-performance-year is easier than keeping constant the mean velocity.

3. Data and descriptive analysis

In previous sections, we have stressed how the monitoring of students career is crucial for improving students performance. In addiction, the indicator (4) is chosen as a measure of the performance over time: $EY_i^*(j)$ can be a useful tool to check the students performance 'in real time'.

From this section, we refer to real data, concerning the cohort of the 160 students enrolled in the 2002/03 academic year and followed up to the 2008/09 academic year, in the first-level DCs of the Faculty of Pharmacy of the University of Palermo (one of the faculties affected by the DC change event (12.5%)).

3.1 *Individual student performance and the DC change*

In order to give evidence to the connection between student performance and DC change, we used the Google Motion Chart tool (https://developers.google.com/chart/). The package `googleVis` [9] provides an interface between software R [21] and these tools.

In particular, we used the Bubble chart because it is very useful to display several indicators over time in a dynamic way. In addition we strongly suggest the policy-makers (either tutor or student mentor) to use that tool if they want to monitor the student path to graduation. In this paper, for the sake of simplicity, we report just three snapshots of the chart, conditioned to the first, the third, and the seventh AY (Figure 1). We focus on the cumulative of CR (*cumCR* on X-axis) and the expected remaining AY to get the degree (EY^* on Y -axis). Each bubble represents a student and its size is proportional to the total number of CR of the student study plan. In fact, although in Section 2 we have stated that this value is 180, in IUS some deviations are actually allowed, in order to lightly customize the Study Plan. Therefore, it is important to consider the exact total number of CR in order to better relate the student speed to the real 'length' of his own path.

If all the students were good students, as time goes by, all the bubbles of the charts in Figure 1 should go towards the right-bottom part of the plot. However, for several students it does not happen: the three plots in Figure 1 show very different patterns of the students performance, after three and seven years. In the same figure, it is possible to notice several bubbles on the top area of the charts, in particular when $EY^* = 60$. It is a conventional value, bigger than the

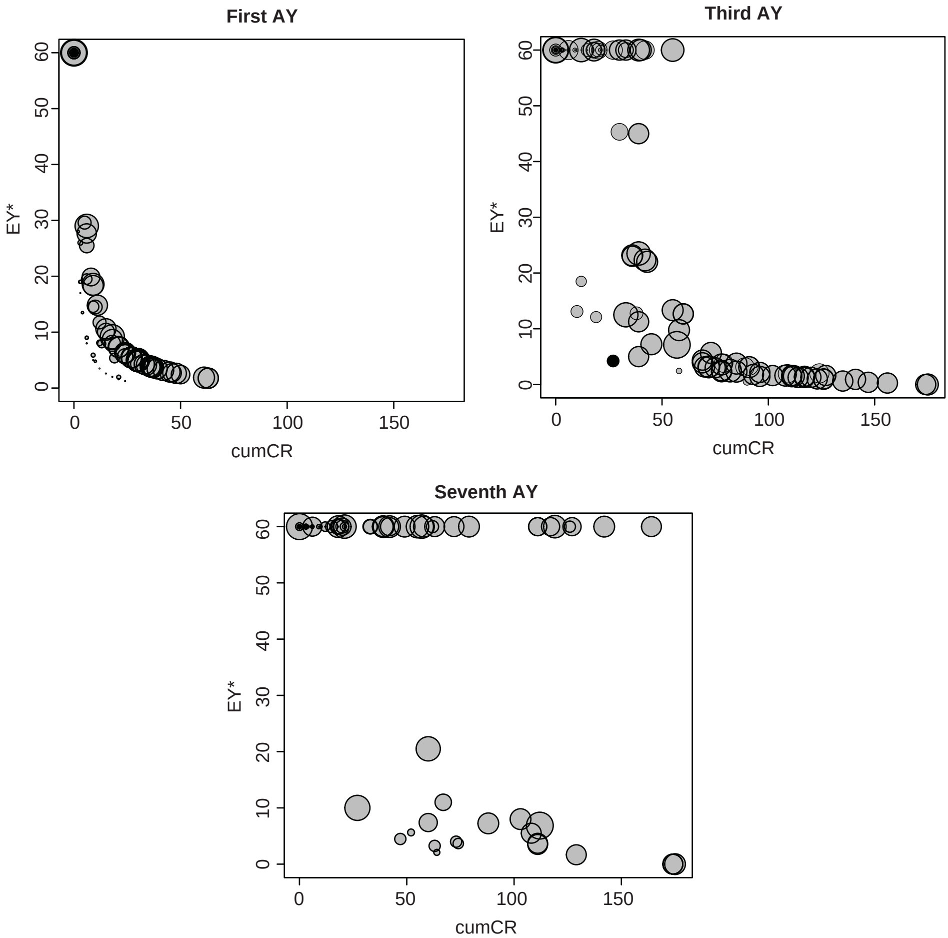

Figure 1. Google Bubble chart snapshots for *cumCR* and EY^* conditioned to first, third, and seventh AY, respectively.

maximum computed value of EY^* (i.e. 55), to represent those students that take 0 CR at the end of a year j on the chart. In fact, if a student does not take any CR, his instantaneous velocity is 0 and EY^* tends to infinity. It is coherent with the definition of EY^*: if a student keeps the instant velocity of 0 CR/AY for the next years, he never gets the degree.

In addition, in the same plots, black bubbles are reported: they correspond to those students who attend a different DC with respect to the previous AY. In this way, it is possible to check the variation on the students performance due to the DC change.

For a better description of the possible student path, in Figure 2 four typical students performance profiles are reported. The top-left plot shows a singular bubble in the top-left area of the chart that actually corresponds to a time series for one of the worst students: he never takes any CR over the seven-year period, therefore the expected number of years to take the degree tends to infinity. The top-right plot regards one of the best students: he earns 63 CR at the end of the first year and therefore he would need almost two years more to get the degree (first grey bubble on the left), if he keeps constant his 63 CR/AY velocity over time. But, at the end of the second AY (second bubble from the left) his path shows a lower instant velocity (58 CR/AY), and therefore his EY^* is one year and four months instead of just one year as expected at the end of

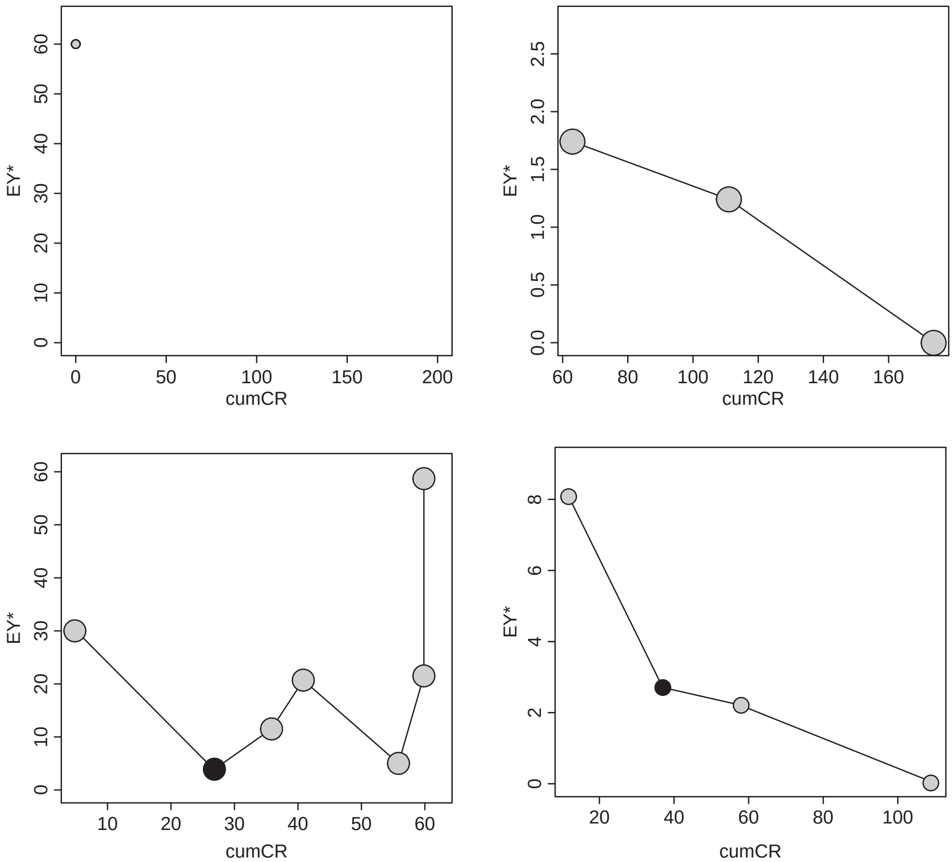

Figure 2. Four ‘typical’ student profiles time series.

the first year. Finally, during the third year, the student speeds up to 63 CR/AY and he gets the degree (first bubble from the right). Between these two extreme situations, two further profiles are reported at the bottom of Figure 2. The left one concerns the time series for a student with a seesawing career. He starts very bad, earning just 5 CR in 2002, he speeds up during the second year (once he has changed DC – black bubble), he decelerates during the next two AY, then he speeds up again but, finally, he performs very bad reaching the top of the chart at the end of the last year. We call him the ‘waving’ student due to his performance path and the typical W shape path over time. Finally, the right-bottom plot reports the time series for a student that changes DC and improves his performance. His career does not start well: he earns just 12 CR at the end of the first year. The DC change in the second year seems to be the right choice: after that moment, his career speeds up and the student gets the degree at the end of the forth AY instead of the ninth AY, as expected at the end of the first year.

3.2 *Students performance in summary*

In the previous subsection, we have showed a graphical way to analyse the career of each student. Now, we consider students performance as a collective. In Table 2, the first two columns refer to the mean and median values of EY^* conditioned to the observed AY. They show very high values due to the inflation of students who do not earn any CR, since their EY^* tends to infinity (60 here).

Table 2. Summary of EY^* and EY_s^* distributions, for each AY.

AY	$\bar{EY}^*$	Me(EY^*)	$\bar{EY}_s^*$	Me(EY_s^*)
1	27.173	10.334	7.996	5.444
2	37.850	60.000	13.369	6.680
3	38.783	60.000	6.958	3.116
4	37.878	60.000	7.832	3.409
5	41.122	60.000	6.995	2.864
6	45.268	60.000	5.452	3.174
7	50.461	60.000	4.931	3.853

Data reflect the effect of ‘not-earning CR’ on the mean and the median values.

Therefore, in order to get a more informative result, we focus the same analysis on those 110 students who never get 0 CR in a single AY. We use here EY_s^* to indicate that we are only considering the subset of those 110 students.

In the other two columns of Table 2, results are quite different, although the mean of the expected further years to graduation remains still high. This is due to the negative weight of the OOT students. On the average, EY_s^* data show that at the end of the first-year students are expected to need eight years more to get the degree (50% of students needs more than five years), and this value rises to 13 at the second year (more than six years for the 50% of the students). These results reflect a typical IUS situation [7]: the first two AY are the hardest for the students, affecting the length of their career path (which legal length is three years).

4. Modelling the DC change effect on EY*

In Section 3, the descriptive analyses have highlighted a possible effect of the DC change variable (DCc) on students performance and the strong asymmetry of the EY^* marginal distribution. Therefore, in order to answer the question ‘Does the student who changes DC improve his performance?’, we also have to take into account the inflation of the 60 values of EY^* and the longitudinal nature of data. From now on, graduates (39 students) are not considered since we want to focus on those students that after 7 years are still in IUS, and since those 39 students have never changed their DC. Hence, we introduce EY_0^* as follows:

$$EY_0^* = \begin{cases} EY^* & \text{when } EY^* < 60, \\ 0 & \text{when } EY^* = 60. \end{cases} \tag{5}$$

In this way, considering EY_0^* as integer (i.e. it has been approximated to the nearest integer) just for sake of simplicity and in order to maintain its ‘nature’ of ‘number of years’ (as AY does), its marginal distribution (Figure 3) clearly reveals a positive asymmetry with a zero mode, suggesting a typical zero-inflated distribution, that will be introduced in the next section.

4.1 *Model for overdispersed data*

Since we follow students over time, data concern repeated measurement on each student. Each student is a cluster of obviously correlated measures, that may produce an overdispersion if not properly accounted.

Moreover, dealing with count data (as EY_0^*), a Poisson model fits the data reasonably when the residual deviance is roughly equal to the residual degrees of freedom. If residual deviance is large, it suggests that the conditional variation of the expected number of clusters exceeds the variation of a Poisson distributed variable, for which the variance equals the mean. This

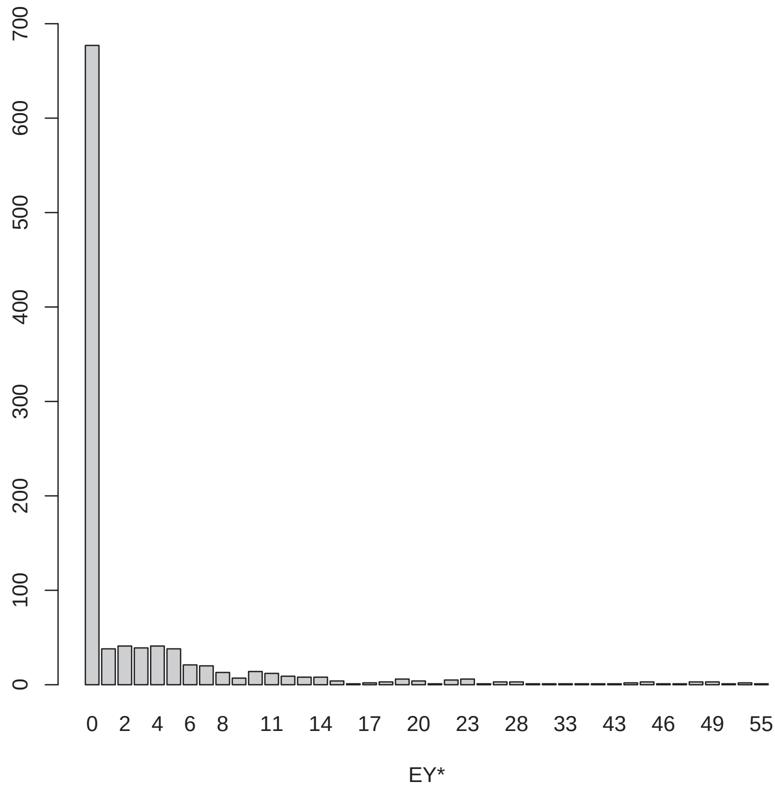

Figure 3. Marginal distribution of EY_0^*.

causes the overdispersion. In regression models for count data, it is commonly achieved by the quasi-Poisson and negative-binomial GLM.

A particular kind of overdispersion can be observed when the number of zeros is more than it is consistent with a Poisson (or negative-binomial) distribution. Several statistical models have been proposed for count data with an excess of zeros, including the zero-inflated Poisson (ZIP) regression model [16]. The usual ZIP model consists of two components (i) a binary logistic-regression model for membership in the latent class of individuals for whom the response variable is 'necessarily' (highly probable) 0, and (ii) a Poisson-regression model for the latent class of individuals for whom the response may be 0 or a positive count.

In using the ZIP model in subject-specific studies, as in our analysis, it is necessary to account for the individual variability, due to the correlation between subject-specific observations. In a general dependency random effect model, the subject variability is caught by a random parameter. Usually, the simple random intercept is added to the model and its variance informs about the individual specific source of variability, or unobserved heterogeneity, that is not adequately controlled by the covariates in the model.

Rodrigues-Motta *et al.* [22] model the mixture and Poisson parameters hierarchically, each as a function of two random effects, representing the subject specific and environmental sources of variability. In the literature, Welsh *et al.* [27], Hall [12], Wang *et al.* [26], Kuhnert *et al.* [15], Min and Agresti [18] considered models for zero-inflated count data, with random effects parameter (denoted here by α). It represents a random subject-specific variable that accounts for intra-cluster correlation and dependence of clustered observations, both in the logistic regression model of the mixture parameter and in the log-linear model of the Poisson parameter.

In this work, we follow the Rodrigues-Motta *et al.* [22] approach just for computational convenience, considering a random intercept model, with ZIP distribution for a heterogeneous variance structure.

More formally, we assume that the response vector $\boldsymbol{Y}$ contains data from k independent clusters (k students), so that $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k)$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, $i = 1, \ldots, k$, (where n_i is the number of AY observed for the ith student) follows a zero-inflated Poisson distribution, with conditionally independent observations. Hence, we assume that, conditional on a random effect parameter α_i,

$$Y_{ij} \sim \begin{cases} 0 & \text{with prob } p_{ij}, \\ \text{Poisson}(\lambda_{ij}) & \text{with prob } 1 - p_{ij}, \end{cases}$$

where $j = 1, \ldots, n_i$.

According to this approach, we model $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{in_i})$ and $\boldsymbol{p}_i = (p_{i1}, \ldots, p_{in_i})$ with log-linear and logistic regression models:

$$\log(\boldsymbol{\lambda}_i) = \boldsymbol{B}_i\boldsymbol{\beta} + \alpha_i,$$
$$\text{logit}(\boldsymbol{p}_i) = \boldsymbol{G}_i\gamma,$$

where $\boldsymbol{B}$ and $\boldsymbol{G}$ are design matrices, $\alpha_1, \ldots \alpha_k$ independent standard Normal random variables, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ the parameter vectors.

Therefore, the likelihood function for the ZIP regression mixed model is

$$l(\boldsymbol{y}) = \sum_{i=1}^{k} \log \int \left[\prod_{j=1}^{n_i} P(Y_{ij} = y_{ij}|\alpha_i) \right] \phi(\alpha_i)\, \mathrm{d}\alpha_i \tag{6}$$

with

$$P(Y_{ij} = y_{ij}|\alpha_i) = \left[p_{ij} + (1 - p_{ij})\mathrm{e}^{-\lambda_{ij}}\right]^{u_{ij}} \left[\frac{(1 - p_{ij})\mathrm{e}^{-\lambda_{ij}}\lambda_{ij}^{y_{ij}}}{y_{ij}!} \right]^{1-u_{ij}},$$

where $\phi(\cdot)$ denotes the standard Normal probability density function, and $u_{ij} = 1$ if $Y_{ij} = 0$ and $u_{ij} = 0$ otherwise. The need of iterative estimation approach is related to the complexity of the maximization of the log likelihood (6), complicated by the integration with respect to $\boldsymbol{\alpha}$.

4.2 *Model estimation*

In the light of the above considerations, a mixed model with ZIP distribution is chosen, estimated by Markov Chain Monte Carlo (MCMC) algorithm [13, 17]. In fact, although on the one hand the MCMC algorithm can be slow, on the other hand analytical results for non-Gaussian generalized linear mixed models are generally not available, and the usual restricted maximum likelihood (REML) may not work well. Moreover, REML uses large sample theory to derive approximate confidence intervals that may have very poor coverage, especially for variance components, while MCMC is an approximation, which accuracy increases the longer the analysis is run for, being exact at the limit. Finally, MCMC measures of confidence are exact, up to Monte Carlo error [22].

Computational results are obtained by using the `MCMCglmm` package [11] of software R, performing 50,000 iterations in order to get quite exact estimates. Although the MCMC approach is technically challenging, we have chosen the `MCMCglmm` package because it seems to be more stable than other packages.

5. Model estimation and discussion

In order to answer the previous question (Section 4) about the effect of the DCc on student performance (measured by EY_0^*), this section is devoted to some considerations about the model computation and its outcomes.

To test the ZIP model suitability, with respect to the Poisson one, we performed the usual Vuong test [25]: $V = 12.27$ ($p = 0.00$) suggests that the ZIP model is more consistent with our data than the ordinary Poisson.

Considering the DCc ($0 =$ do not change, $1 =$ change) and the AY as covariates, we investigate the pertinence of three different models: the ZIP model, the ZIP model with interaction between DCc and AY, and the case of the Poisson model (just for further comparison). In particular,

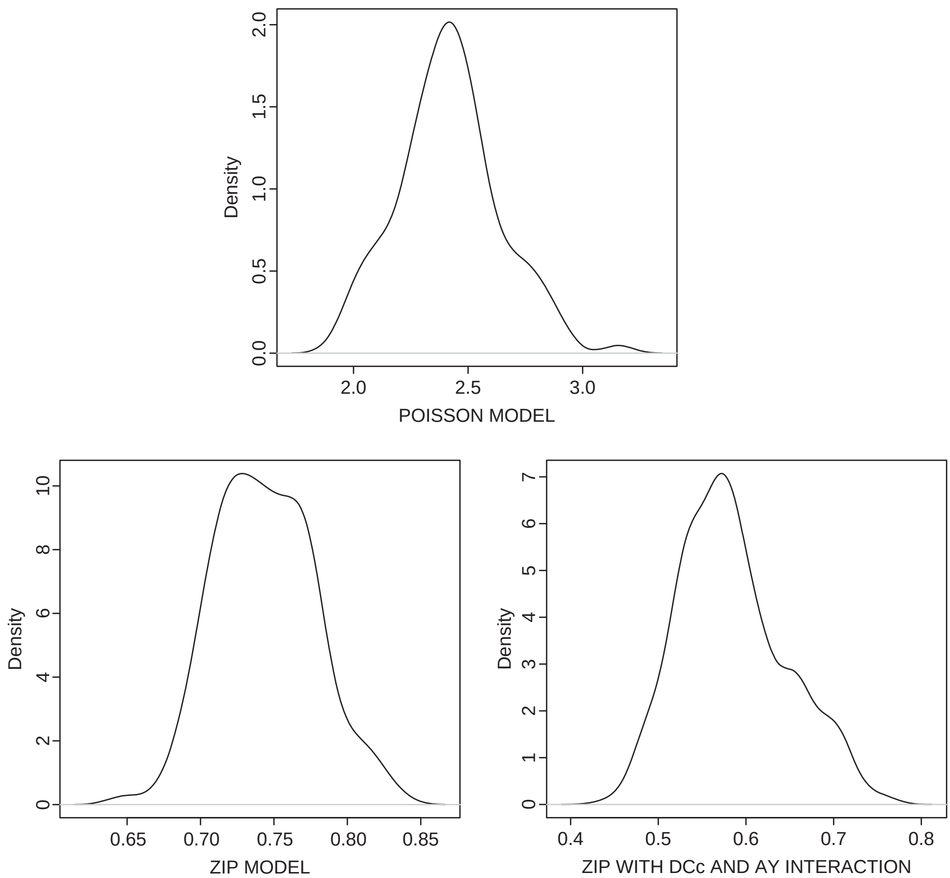

Figure 4. Residual variance components distributions for Poisson, ZIP, and ZIP with interaction models.

Table 3. Comparison between ZIP and ZIP with interaction models.

	Model	
Criterion	ZIP	ZIP with interactions
DIC	2564.294	2739.13
pRE 95% interval	[0.83, 0.88]	[0.75, 0.85]

Table 4. Model output for ZIP model, with random intercept, and without interactions.

	post.mean	l–95% CI	u-95% CI
Intercept Logit	− 2.41	− 2.46	− 2.37
Intercept	2.48	1.98	2.92
DC Change	− 0.27	− 0.30	− 0.25
AY			
Year 2	0.87	0.85	0.89
Year 3	0.63	0.62	0.64
Year 4	0.88	0.86	0.89
Year 5	0.52	0.52	0.54
Year 6	1.03	1.03	1.04
Year 7	0.85	0.84	0.86

we have considered AY as a factor, in order to better investigate the variation in performance, conditioned to each academic year.

The residual variance components of the three models are reported in Figure 4. The Poisson model (top plot) shows a much higher residual variance than the ZIP models (bottom plots). That confirms our previous remarks about the appropriateness of a ZIP model. The two ZIP models are compared by the deviance information criterion (DIC) [8] and by the proportion of the total variance explained by the random effect (pRE). The latter has been obtained dividing the individual variation by the sum of the other variance components (the error term and the ZIP part). Results (Table 3) show evidence in favour of the ZIP model without interactions.

In Table 4, the output of the estimated chosen model (with random intercept, ZIP distribution, and without interactions) is reported. The baseline is the first-year student that did not change the DC. All the covariates are significant, and our attention is obviously caught by the DCc effect. Its coefficient estimate shows that changing DC, on average, reduces the EY_0^* by about nine months. In other words, it seems that (on average) those who change speed up. The AY coefficients estimates highlight the long-time persistence of the student in IUS: the expected years to graduation seem to increase as time goes by, in particular after the third year (legal duration). For example, second-year students are expected to graduate two years and five months later than expected at the end of the first AY. That aspect highlights the real need of implementing suitable policies aimed to support OOT in taking their degree.

In this paper, we stress that in the IUS the duration of students careers is very long, well over the legal duration (three years). Considering the expected years to graduation as a measure for the students performance, we have noticed that in our cohort there are several students who change their DC and an excess of expected years to graduation that tends to infinity.

Since we want to take into account the correlation of the repeated measures for each student and the distribution of EY^*, a mixed-effect model with ZIP distribution is chosen to study the effect of DC change on the expected years to graduation. Results show that ZIP performs better than Poisson and ZIP with interaction models, and that the effect of the DCc is significant and negative. Obviously, we could have taken into account many other covariates, such as socio-demographic ones, but in the light of previous results on similar data [3], we would not expect different conclusions.

Since this is a ‘first step’ analysis, as a further development, it could be interesting to perform the model for other DCs, and consider correlated individual effects to account for correlation among students (e.g. in a multilevel model with different faculties or different universities). Other improvements could take into account a mixture model with a zero-inflated component, or a model that considers the DCc as a ‘changing point’ in a time series analysis [20].

Acknowledgements

The authors thank the Editors and the anonymous referees for their valuable comments which improved the quality and the clarity of the manuscript. The authors also gratefully acknowledge the support and useful comments of Vincenza Capursi, University of Palermo. The article is the result of the productive collaboration among the authors. In particular, Sections 1 and 4 can be ascribed to Giada Adelfio; Sections 2 and 3 to Giovanni Boscaino.

Disclosure statement

The dynamic version of the Bubble Chart illuatrated on Section 3 is available on http://www.unipa.it/gioboscaino/chart.html

Funding

This work was supported by University of Palermo, Italy [grant number 2012-ATE-0454], "Misure e modelli per la valutazione del Sistema Universitario".

References

[1] G. Adelfio and G. Boscaino, *The student talent in a random effect quantile regression model for university performance*, in *Proceedings of the 28th International Workshop on Statistical Modeling*, V.M.R. Muggeo, V. Capursi, G. Boscaino, and G. Lovison, eds., Vol. 2, 2013, pp. 479–483. ISBN 978-88-96251-49-2.

[2] G. Adelfio, G. Boscaino, and V. Capursi, *A new indicator for higher education student performance*, Higher Educ. 68(5) (2014), pp. 653–668.

[3] M. Attanasio, G. Boscaino, V. Capursi, and A. Plaia, *May the students' career performance helpful in predicting an increase in universities income?* in *Statistical Models for Data Analysis*. Series in Studies in Classification, Data Analysis, and Knowledge Organization, P. Giudici, S. Ingrassia, and M. Vichi, eds., Springer International Publishing, Switzerland, 2013. Available at http://link.springer.com/book/10.1007%2F978-3-319-00032-9.

[4] E.R. Birch and P.W. Miller, Student outcomes at university in Australia: A quantile regression approach. Aust. Econ. Pap. 45, 2006, pp. 1–17, doi:10.1111/j.1467-8454.2006.00274.x.

[5] G. Boscaino, V. Capursi, and F. Giambona, The careers' performance of a University students' cohort. DSSM Working paper, n. 2007.1, 2007.

[6] J.S. Cheesman, N. Simpson, and G. Wint, Determinants of student performance at university: Reflections from the Caribbean, 2006. Available at http://www.mona.uwi.edu/opair/research/student-performance-paperrevised.pdf.

[7] CNVSU, *XI Rapporto sullo stato del sistema universitario*, 2012. Available at http://www.cnvsu.it/_library/downloadfile.asp?id = 11778.

[8] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis: Texts in Statistical Science*, 2nd ed., CRC Press, Boca Raton, FL, 2004. ISBN 1-58488-388-X.

[9] M. Gesmann and D. de Castillo, Interface between R and the Google Chart Tools. *Package Version 0.4.3*, 2013.

[10] L. Grilli, C. Rampichini, and R. Varriale, Predicting students academic performance: A challenging issue in statistical modelling, in *Cladag 2013 Book of abstracts*, Monerva, Morlini and Palumbo, eds., CLEUP, Padova, 2013, pp. 249–254.

[11] J.D. Hadfield, *MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package*. J. Stat. Softw. 33(2) (2010), pp. 1–22. Available at http://www.jstatsoft.org/v33/i02/.

[12] D.B. Hall, *Zero-inflated Poisson and binomial regression with random effects: A case study*, Biometrics 56 (2000), pp. 1030–1039.

[13] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika 57 (1970), pp. 97–109.

[14] P. Horn, A. Jansen, and D Yu, *Factors explaining the academic success of second-year economics students: An exploratory analysis*, South African Journal of Economics 79(2) (2011).

[15] P.M. Kuhnert, T.G. Martin, K. Mengersen, and H.P. Possingham, *Assessing the impacts of grazing levels on bird density in woodland habitat: A Bayesian approach using expert opinion*, Environmetrics 16(7) (2005), pp. 717–747.

[16] D. Lambert, *Zero-inflated Poisson regression, with an application to defects in manufacturing*, Technometrics 34(1) (1992), pp. 1–14.

[17] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chem. Phys. 21 (1953), pp. 1087–1092.

[18] Y. Min and A. Agresti, *Random effect models for repeated measures of zero-inflated count data*. Statist. Model. 5, 1–19, 2005.

[19] MIUR, Osservatorio studenti-didattica, 2014, Available at http://anagrafe.miur.it/php5/cerca-laureati.php.

[20] V.M.R. Muggeo and G. Adelfio, *Efficient change point detection for genomic sequences of continuous measurements*, Bioinformatics 27(2011), pp. 161–166.ISSN 1367-4803 doi:10.1093/bioinformatics/btq647.

[21] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013, http://www.R-project.org/.

[22] M. Rodrigues-Motta, D. Gianola, and B. Heringstad, *A mixed effects model for overdispersed zero inflated Poisson data with an application in animal breeding*, J. Data Sci. 8 (2010), pp. 379–396.

[23] C. Tattersall, W. Waterink, P. Höppener, and R. Koper, *A case study in the measurement of educational efficiency in open and distance learning*, Distance Educ. 27 (2006), pp. 391–404.

[24] C.A.C. Van Bragt, A.W.E.A. Bakx, T.C.M. Bergen, and M.A. Croon, *Looking for students' personal characteristics predicting study outcome*, Higher Educ. 61 (2011), pp. 59–75.

[25] Q.H. Vuong, *Likelihood ratio tests for model selection and non-nested hypotheses*, Econometrica 57(2) (1989), pp. 307–333.

[26] K. Wang, K.K.W. Yau, and A.H. Lee, *A zero-inflated Poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays*, Comput. Methods Programs Biomed. 68 (2002), pp. 195–203.

[27] A.H. Welsh, R.B. Cunningham, C.F. Donnelly, and D.B. Lindenmayer, *Modelling the abundance of rare species: Statistical models for counts with extra zeros*, Ecol. Model. 88(1) (1996), pp. 297–308.