Journal of Environmental Statistics
September 2018, Volume 8, Issue 4. http://www.jenvstat.org

Hints of latent drivers investigating university student performance

Giovanni Boscaino and **Giada Adelfio**

Abstract

Job market, nowadays, asks for higher and higher skills and competences. Therefore, also the measurement and assessment of the university students performance are crucial issues for policy makers. Although the scientific literature provides several papers investigating the main determinants of university student performance, often results are very different, and they seem to hold just in very specific contexts. This paper aims to contribute to the international literature, focusing on the role of student specific characteristics, supporting the idea that unobservable variables (such as motivation, aptitudes or abilities) should be more investigated.

Keywords: student performance, indicator, random effects quantile regression.

1. Introduction

Academic student performance represents, nowadays more than ever, a very important topic of interest for university policy makers. Indeed, the global job market needs more and more competitiveness, and high skills and competences, and then an improved quality of the graduates. Obviously, the definition of the graduate 'quality' is a quite complex issue. For instance, it may concern the real student ability in solving practical problems, or the wideness and depth of his/her knowledge, etc. or a combination of them. In this paper, the attention is devoted to the student performance at university, as a proxy of the student quality, although its measurement represents a challenge for researchers. Indeed, there is not a shared measure (and methodology) for the student performance. Some authors refer to the grade, others to the distribution of marks, or pay attention to the credits earned (usually related to time, marks, or both), or, again, to the time spent to get the degree (where there are not time limits to get the degree).

Just to quote some of the most recent studies, Gillette, Rudolph, Rockich-Winston, Blough, Sizemore, Hao, Booth, Broedel-Zaugg, Peterson, Anderson, Riley, Train, Stanton, and Anderson (2017) predict student performance using a multivariate approach highlighting the role of the admission test and the cumulative GPA gained during a three years course. Logan, Lundberg, Roth, and Walsh (2017) focus their attention on student individual motivation and cognitive ability effect on performance, in an on-line distance education course, and results show a positive association with the academic performance. Family expenditures and student personal information are significant for Daud, Aljohani, Abbasi, Lytras, Abbas, and Alowibdi (2017) in predicting student success. Cheesman, Simpson, and Wint (2006) apply 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. Tattersall, Waterink, Hoppener, and Koper (2006) measure the educational efficiency in terms of comparison between inputs and outputs. The output-input ratio is analysed including several aspects of the students path to graduation, e.g. Learning interruption and Changes of the Study Programme (SP). The influence of "Change of SP" on the expected time to graduation is also analysed in Adelfio and Boscaino (2016), highlighting its significant and negative effect. Birch and Miller (2006) use a Quantile Regression approach identifying in Tertiary Entrance Rank, Gender, and High School the most important determinants of high and low performance. Boscaino, Capursi, and Giambona (2007) focus 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. Van Bragt, Bakx, Bergen, and Croon (2011) follow a more social psychological root making a deeper analysis of performance: they also perform an ad-hoc survey to investigate the impact of the Big Five personality characteristics, Personal learning orientations, and study Approach on performance, using a logistic model. Results show a positive effect of Conscientiousness and a negative one of Ambivalence and Lack of regulation. Horn, Jansen, and Yu (20011) perform an exploratory analysis on the factors that lead the success at the end of the first year as drivers for the performance at the second year. They highlight that Lectures and Tutorial attendance are significant factors, and the most important determinant was the performance during the first year. Attanasio, Boscaino, Capursi, and Plaia (2013) highlight 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. Grilli, Rampichini, and Varriale (2013) use the pre-enrolment assessment test outcomes and some personal student characteristics, in order to predict earned credits at the end of the first year. Using different models (hurdle and binomial mixture model), they highlight the poor role on the pre-enrolment test.

Summing up, at first sight scientific literature seems to generally suggest that the impact of the determinants varies (in terms of extent and direction) according to the context (economic, social, political, demographic, etc.) and results should hold just in that context, because every university, city, country has its own settings and political features.

The main question still holds: on which aspects should the policy makers pay attention in order to reduce the drop-outs and degree delays? Our study aims to contribute to find an answer. In particular, we focused on the Italian University System and specifically on University of Palermo, as first study step. Available student information for the University of Palermo regard just few aspects, and there are not any psychological measures or tests. To access at the most of the Degree Courses, students have to pass an attitudinal test (not psychological but just a test on mathematics, literature, foreign language, and general knowledge). Therefore, this paper investigates the role of students characteristics affecting the academic performance, accounting for two study perspectives: both looking at the performance of students measured by a transformed version of marks, and at their subject specific characteristics, that in turn reflect the students variability. In particular, a recently new measure for the performance is considered and Quantile Regression (QR) approach is used (Section 2). In Section 3 data are introduced and results of the analysis following the two perspectives are provided. Conclusive remarks are reported in Section 4.

2. Method

2.1. The measure of student performance

Scientific literature offers several measures for student performance at university. Adelfio, Boscaino, and Capursi (2014) propose a measure that takes into account for both credits and marks, in order to better relate a mark to the complexity of the course (summarized by the credit), since a 1 credit course is supposed to be 'easier' than a 12 credits one. Therefore, Adelfio *et al.* (2014) introduce the following measure:

$$m'_{ij} = \frac{best_C - suf_C}{max_j(m^w_{ij}) - min_j(m^w_{ij})} \times (m^w_{ij} - min_j(m^w_{ij})) + suf_C \tag{1}$$

where $best_C$ and suf_C are the marks that correspond to the best and to the minimum passing mark stated in the country C system, respectively; $m^w_{ij} = \frac{m_{ij}Cr_j}{\sum_{j=1}^{J} Cr_j}$, where m_{ij} is the mark earned by the student i and Cr_j is the credit for the course j.

It is important to note that the measure in 1 takes values in the same quantitative marks scale used in the Country C. In addition, the measure in Eq. (1) transforms marks distribution (conditioned by credit) in a more symmetric one and allows easier comparisons. Figure 1 shows the effect of the new indicator, starting from the data analysed in section 3 and referred to the students enrolled at the First Level Degree in Economy and Finance (E) and in Life Sciences (L), respectively, of the University of Palermo (Italy) in 2002, and graduated from 3 to 7 years after. The top left and bottom left plots report the conditional marks distributions by credits for the two degree courses (E and L, respectively). The E boxplots show more marks variability, across the credits, than the L one. While the marks distribution for L are highly negative asymmetric for all the credits, the marks distribution for E has higher mean marks in correspondence of the lower credits and lower mean marks for the higher credits (with the exception of the courses with 9 credits). The top right and bottom right plots show the effect of the new indicator (1): marks are re-scaled accounting for the workload of the courses penalizing higher marks for lower credit courses with respect to the lower marks for the higher credits courses. The distributions of new marks are fairer and more easily comparable.

As mentioned before, the new measure is used in this paper for investigating two different perspectives: 1) finding the determinants of student performance, summarized by the Median of m'_{ij}, that is $Me(m'_{ij})$ and 2) analysing the m'_{ij} distribution for each student. The first perspective aims to identify the effect of covariates on the Median performance; the second perspective aims to take into account the student variability of each m'_{ij} distribution. In particular, we assume that the variability of the students m'_{ij} distributions reflects the variability among students, related to all the personal, individual, latent students characteristics, like motivation, aptitude, etc.

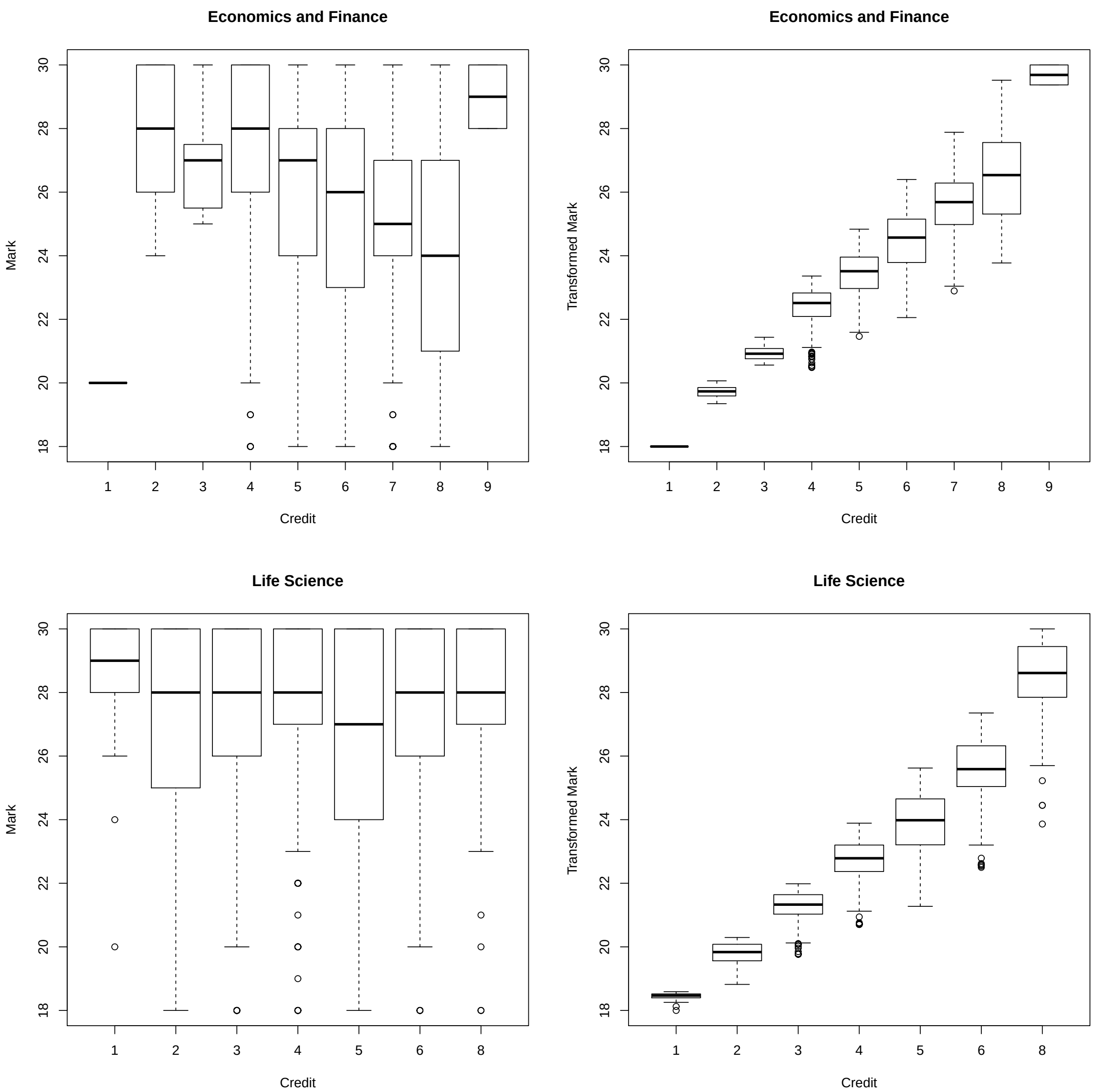

Figure 1: Grades distributions of the marks and transformed marks, according to the measure (1), conditioned by credits, for the Economics and Finance and Life Science Study Programmes

2.2. The Quantile Regression model

We refer to the Quantile Regression (QR) (Koenker 2005) approach in order to investigate the influence of some determinants over the whole shape of the distribution of the proposed indicator in eq. (1). Indeed, QR allows to study the effect of covariates on the whole response distribution, that is, in our context, to highlight the role of some factors on a subgroups of students (e.g. just for the best students or the worst).

QR deals with the estimation of conditional quantile functions, for models in which quantiles of the conditional distribution of the response variable are expressed as functions of the observed covariates. Whereas the usual linear regression model approximates the conditional mean of the response variable, QR aims at estimating either the conditional median or other quantiles of the response variable; QR also provides more robust estimates than the usual OLS based regression. Unlike the ordinary linear regression, the QR parameters measure the change in specified quantiles of the response variable produced by one unit change in the predictor variables. In other words, there are percentiles of the response distribution that may be more affected by some specific covariates, than others.

In particular, from a more formal point of view, let $\{(\mathbf{x}_i, y_i),\ i = 1, \ldots, n\}$ be a sample of size n from some unknown population, where $\mathbf{x_i} \in \mathbf{R^d}$. The conditional ϕth quantile function $f_\phi(x)$ is defined such that $P(Y \leq f_\phi(X)|X = x) = \phi$, for $0 < \phi < 1$.

Therefore, the ϕth conditional quantile function can be estimated by solving:

$$\min_{f_\phi \in \mathbf{R}} \sum_{i=1}^{n} \rho_\phi \left(y_i - f_\phi(\mathbf{x}_i)\right) \tag{2}$$

where the function $\rho_\phi(\cdot)$ is the tilted absolute value function, that yields the ϕth sample quantile as its solution and is defined by $\rho_\phi(r) = \phi r$ if $r > 0$, and $-(1-\phi)r$ otherwise (Koenker and Bassett 1978).

Setting $f_\phi(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}_\phi$ where $\boldsymbol{\beta}_\phi = (\beta_{\phi,1}, \beta_{\phi,2}, \ldots, \beta_{\phi,d})^T$, such that the conditional ϕth quantile function $f_\phi(x)$ is a linear function of the parameters $\boldsymbol{\beta}$, a linear quantile regression is considered. Observe that, contrary to a common misconception, the estimation of coefficients for each quantile regression is based on the weighted data of the whole sample, not just the portion of the sample at that fixed quantile Hao and Naiman (2007).

In this paper, QR estimates are obtained by the modified version of the Barrodale and Roberts algorithm described in Koenker and d'Orey (1987) and Koenker and d'Orey (1994). This approach is quite efficient for problems up to several thousand observations, computing also confidence intervals for the estimated parameters, based on inversion of a rank test (Koenker 1994). The here described QR model is used in this paper in order to follow the perspective 1) of Section 2, whereas the perspective 2) is investigated following a linear mixed QR approach, that accounts for the subject specific variability.

More formally, let $(\mathbf{x}'_{ij}, y_{ij})$, for $j = 1, ..., n_i$ and $i = 1, ..., N$, be repeated measurements data, where $\mathbf{x}'_{ij}$ are row p-vectors of a known design matrix and y_{ij} is the j-th measurement of a continuous random variable on the i-th subject.

According to the considered approach, the linear mixed quantile functions of the response y_{ij} is:

$$G_{y_{ij}|u_i}(\tau|\mathbf{x}_{ij}, u_i) = \mathbf{x'}_{ij}\boldsymbol{\beta} + u_i,\ j = 1, \ldots, n_i,\ i = 1, \ldots, N \tag{3}$$

where $G_{y_{ij}|u_i}(\cdot) \equiv F^{-1}_{y_{ij}|u_i}(\cdot)$ is the inverse of the cumulative distribution function of the response conditional on a location-shift random effect u_i Geraci and Bottai (2006). For this model, the location-shift effects are assumed random identically and independently distributed according to some density f_u, usually $u_i \sim N(0, \alpha)$, characterized by a τ-dependent dispersion parameter $(\alpha(\tau))$. Moving away from the penalized approach provided by Koenker (2004),Geraci and Bottai (2006) assume that y_{ij}, conditionally on u_i, are independently distributed according to an Asymmetric Laplace Distribution (ALD):

$$f(y_{ij}|\beta, u_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_\tau\left(\frac{y_{ij}-\mu_{ij}}{\sigma}\right)\right\}$$

where $\mu_{ij} = \mathbf{x'}_{ij}\boldsymbol{\beta} + u_i$ is the linear predictor of the τth quantile, fixed and known, and σ is the usual scale parameter. The random effects, that induce a correlation structure among observations on the same subject, are assumed to be independent. This is a likelihood-based approach for estimating the QR based on the ALD; this approach is better then the penalized fixed effects one in terms of mean squared error of the QR estimators. Alternative models with non-normally distributed residuals are developed in Seltzer and Choi (2002). The ALD approach is here considered since it provides an automatic choice of the optimal level of penalization and also because it represents a suitable error law for the least absolutes estimator (and therefore a natural choice in QR).

3. Data and Results

As first analysis step, data concerns the cohorts of 131 and of 98 students enrolled at the First Level Degree in Economy and Finance (E) and in Life Sciences (L) respectively, of the University of Palermo (Italy) in 2002, and graduated from 3 to 7 years after. In Italy, indeed, there is a legal duration (three years) for Bachelor's degree but it is not compulsory: therefore, a student can take the degree beyond three years from enrolment These two Study Programmes are different: E topics mostly regard Economics, Law, Mathematics and Statistics; L topics are more devoted to Chemistry, Physics, and activities in medical laboratories. Their related job markets are different too: E gives access to a wider one (with different contract and salary perspectives) then L. Then, it is reasonable assuming that they usually draw different kind of students, with different job ambitions, wishes, abilities, and basic skills (MIUR 2017). In addition, starting from our previous local knowledge, we assume that the E course also draws those students who are still uncertain about their preferences (on average, in Italy, they basically choose E for the wide job opportunities that it offers), while the L course has so specific topics that it would draw more motivated (or aware) students.

The analysed dataset reports information for each student about both Courses, Credits, Marks, Gender (Female vs Male), High School (Lyceum vs Not Lyceum), Residence (Palermo vs Not Palermo), High School Diploma Mark, and Status at Graduation (students have been classified as "On-Time" - OT - if they got the degree in 3 years, and as "Out-Of-Time" - OOT - the others).

With respect to the perspective 1), several preliminary linear QR models on the $Me'_i = Me_i(m'_{ij})$ are performed (not here reported for the sake of brevity). Results show that the only covariate with a significant effect is the Student Status at graduation (that is OT vs OOT); however, the conditional linear QR model for the Status provides no significant variable for the

OT graduates: in other words, the performance of the students who get their degree on time does not seem to depend on the considered social and demographic characteristics. Probably, it seems that good performers are just good students, and their positive performance may be ascribed to their own motivation, inclination, method of study, etc. Otherwise, in Table 1 we report the QR parameter estimations for the OOT students conditioning to the five quantiles $\tau = (0.05, 0.25, 0.50, 0.75, 0.95)$, for both E and L: results shows that some covariates are significant only with respect to some few percentiles.

Table 1: OLS and QR estimates for Me'_i for OOT graduates, in E and L cohorts.

E cohort	OLS	τ_1	τ_2	τ_3	τ_4	τ_5
Intercept	25.28**	24.71**	24.95**	25.31**	25.56**	25.94**
High School	0.17	0.34*	0.27	0.07	0.06	0.13
Residence	-0.27**	-0.47**	-0.40**	-0.33*	-0.08	-0.10
Gender	-0.20	-0.61**	-0.39*	-0.12	-0.06	0.27*
Dipl. Mark	0.02**	0.02*	0.01	0.02**	0.02**	0.02**
L cohort						
Intercept	23.60**	22.66**	23.15**	23.80**	23.80**	24.23**
High School	-0.01	0.47**	0.032	-0.34	-0.10	-0.20
Residence	-0.35**	-0.31**	-0.28	-0.49*	-0.23	0.17
Gender	0.02	0.47**	0.11	-0.08	0.00	-0.18
Dipl. Mark	0.01	0.01**	0.01	0.00	0.00	0.02**

(** for $\alpha = 0.05$, * for $\alpha = 0.1$)

In short, the Intercept of the models represents the estimated conditional quantile of the Me'_i distribution of students that are Female, with Lyceum diploma, living in Palermo, and with mean-centred Diploma Mark. As expected, the higher the quantile the higher the performance. The other values refer to the distribution of the estimated coefficients for different quantiles. Focusing on L, the Intercept is steeper than the E cohort from the 5-th to 50-th percentile, but it gets always lower values. With respect to the covariates, both for the E and the L cohorts all the variables are significant just around the 5-th percentile, that is the lowest performance students group, and the Diploma Mark around 95-th percentile, the highest performance students group. In addition, a different role is played by the Gender: if for the E cohort Males perform worse than Females, for the L cohort we notice the reverse (but just conditioned to the significant 5-th percentiles).

In order to study the statistical differences between the estimated values of the covariates coefficients for E and L, the comparison is just based on the pairs of significant coefficients, given the percentiles. In Table 2, the proportions of confidence intervals that do not overlap are reported for each covariate. As it is always true that if the confidence intervals do not overlap, then the statistics will be statistically different (Knezevic 2008), it is possible to notice (tab. 2) that Gender has always a different effect (for E and L), conditional to the same significant percentiles. Otherwise, considering the Residence, it happens for just the 14% of the comparisons.

Therefore, among the considered covariates, just the OOT Status seems to play a role on the

Table 2: Proportions of E and L non-zero-including confidence intervals that do not overlap, by covariate.

High School	Residence	Gender	Diploma Mark
0.50	0.14	1.00	0.25

student performance and, with respect to the OOT students, the effect of student characteristics is low and significant for few quantiles. Hence, student performance could be affected by some other not here considered social, demographic, economic (e.g. no information on income is here available), and/or by some specific unobservable student characteristics. Considering the differences between E and L described at the beginning of this section, it is plausible to think that L and E students are actually different in motivation, basic knowledge, abilities, or aptitudes. We call these aspects with one unique convenient word: *talent*. Then, could be the student *talent* a determinant for his/her university performance? The answer is not straightforward: both because of the impossibility of observing the *talent*, and of measuring it. Psycho-aptitude tests should be a solution, but they could be very difficult to implement, time consuming and expensive. Thereby, in order to answer to the question, we try to characterize the *talent* by the all unknown subject specific characteristics that could influence the new 'weighted' marks distribution of each student: in such a way, marks distribution variability could be considered as a measure of each student talent. Following the perspective 2) (Section 2), student marks variability is formally accounted by a QR mixed approach for repeated measurements, adding a student specific random intercept (Section 2.2).

Our results are obtained using the R package lqmm: Linear Quantile Mixed Models Geraci (2014), where a subject-specific random intercept accounts for the within-group correlation (see eq. (3)). In Figure 2, results are reported with respect to the fixed coefficients estimates. To assess the suitability of the model in eq. (3), results are also commented in the light of those reported in the previous part of this section. In Figure 2 there are the plots of the QR estimates of the fixed parameters of (3), for each of the estimated coefficients (the dashed curve with filled dots), conditional to the quantiles τ ($\tau = 0.05, 0.25, 0.50, 0.75, 0.95$). Dots may be interpreted as the impact of a unit-change of each covariate on the response variable, fixed the others. The grey area represents the 95% pointwise confidence band. The solid horizontal line, together with its 95% confidence intervals (horizontal dashed lines), refers to the estimate for a Linear Model with random intercept. The Intercept panel refers to the expected m'_{ij} (vertical axis) conditional to each quantiles (horizontal axis) for students that are Female, living in Palermo, with a Lyceum diploma, OT in Life Science, with a mean High school diploma mark equals to 90. We also added an interaction term between the student Status and the SP, to analyse how the two variables interact in affecting the response. The Intercept is steeper than the QR models without random effect (tab. 1): considering the m'_{ij} distribution rather than their median allows to appreciate the variation of the expected performance when we move between two consecutive quantiles. With the exception of the SP panel, other panels show no significant effect of the covariates, in the most of the quantiles. This partially confirms that student performance is mainly caused by the students specific characteristics, which in turn mainly influence the choice of the student program, rather than their social and demographic ones.

Moreover, the random intercept aims to catch the subject specific *talent* effect on the per-

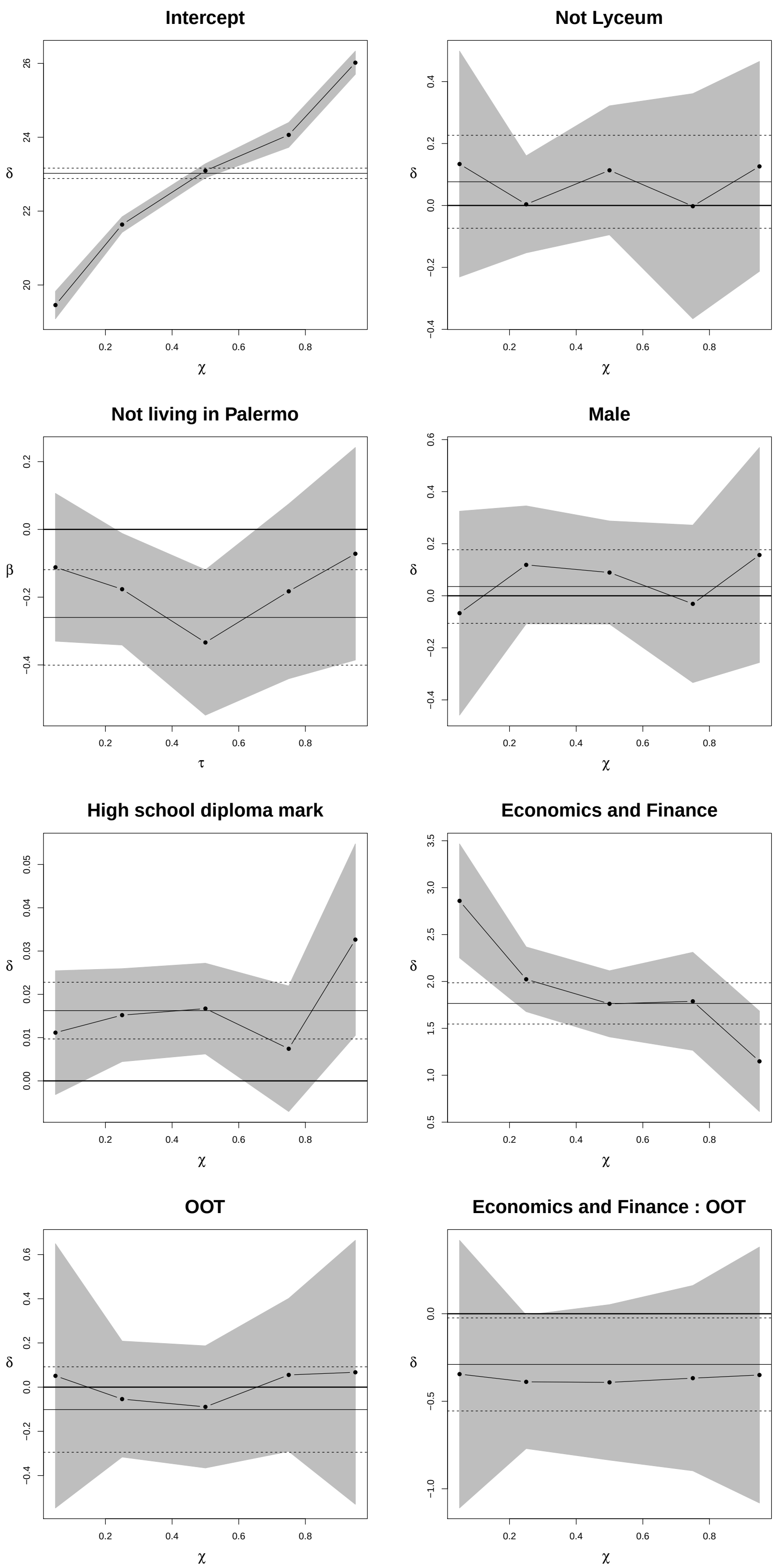

Figure 2: Estimated quantile regression coefficients of the model with random intercept (Eq. 3) and 95% confidence intervals (shaded area).

Table 3: τ estimated conditional variances of the u_i random intercepts

τ	0.05	0.25	0.50	0.75	0.95
$\hat{\alpha}(\tau)$	0.367	0.391	0.398	0.574	0.613

formance. In fact, the τ dependent estimated variances of u_i random intercepts of eq. (3), reported in Table 3, are all in the interval $(0.36, 0.62)$ and they reflect the heterogeneity among students due to their specific characteristics or *talent* (Diggle, Liang, and Zeger 2002).

4. Final remarks

The necessity of investigating the effect of possible social-demographic variables to the overall performance of university students is widely recognized, although the literature offers results that seem to be dependent on the analysed context, both in space and time.

This paper is based on the first results reported in Adelfio *et al.* (2014): these lead to a new, though starting, investigation perspective in analysing the determinants of the students performance. In particular, the student choice of the Study Programme is considered as a final consequence of several subjective student characteristics (such as aptitude, motivation, inclination, ...).

In order to have comparable results, also among different Study Programmes, we use a measure like the one in eq. (1) that accounts in a unique quantity both for the workload (credits) and marks. The provided analyses highlight the role of the SP variable, though for two big courses of the University of Palermo. Indeed, the chosen SP could be a reflect the personal inclinations and subjective sphere of student. In the light of this, in order to improve the students performance policy makers should look at the student personal abilities, aptitudes, motivations, instead of just at the usual social and demographic characteristics. Therefore, following this perspective, student orientation and tutoring activities, and proper selective University entry-tests could be useful tools to address students to the appropriate Study Programme and to their success (helping them in following their own *talent*).

As future work, also several Study Programmes will be considered, to assess the robustness of these current results obtained from our starting hypothesis of research.

References

Adelfio G, Boscaino G (2016). "Degree course change and student performance: a mixed-effect model approach." *Journal of Applied Statistics,*, **43**, 3–15.

Adelfio G, Boscaino G, Capursi V (2014). "A new indicator for higher education student Performance." *Higher Education*, **68**(5), 653–668.

Attanasio M, Boscaino G, Capursi V, Plaia A (2013). "May the students career performance helpful in predicting an increase in universities income." In SI P Giudici, M Vichi (eds.), *Series in studies in classification, data analysis, and knowledge organization, Switzerland Springer International Publishing*, pp. 9–16. Statistical models for data analysis.

Birch ER, Miller PW (2006). "Student outcomes at University in Australia a quantile regression approach." *Australian Economic Press*, **45**(1), 1–17.

Boscaino G, Capursi V, Giambona F (2007). "The careers' performance of a University students' cohort." *Technical report*, DSSM Working paper, n. 2007.1.

Cheesman JS, Simpson N, Wint G (2006). *Determinants of student performance at university Reflections from the Caribbean.*

Daud A, Aljohani N, Abbasi R, Lytras M, Abbas F, Alowibdi J (2017). "Predicting Student Performance using Advanced Learning Analytics." *WWW '17 Companion Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 415–42.

Diggle P, Liang K, Zeger S (2002). *Analysis of longitudinal data.* Oxford Statistical Science Series.

Geraci M (2014). "Linear Quantile Mixed Models The lqmm Package for Laplace Quantile Regression." *Journal of Statistical Software*, **57**(13), 1–29.

Geraci M, Bottai M (2006). "Quantile regression for longitudinal data using the asymmetric Laplace distribution." *Biostatistics*, **8**, 140–154.

Gillette C, Rudolph M, Rockich-Winston N, Blough E, Sizemore J, Hao J, Booth C, Broedel-Zaugg K, Peterson M, Anderson S, Riley B, Train B, Stanton R, Anderson HJ (2017). "Predictors of student performance on the Pharmacy Curriculum Outcomes Assessment at a new school of pharmacy using admissions and demographic data." *Currents in Pharmacy Teaching and Learning*, **9**(9), 84–89.

Grilli L, Rampichini C, Varriale R (2013). "Predicting students academic performance a challenging issue in statistical modelling." In M Monerva, Palumbo (eds.), *Cladag 2013 Book of abstracts.* CLEUP.

Hao L, Naiman D (2007). *Quantile Regression. series: Quantitative Applications in the Social Sciences.* Sage Publications.

Horn P, Jansen A, Yu D (20011). "Factors explaining the academic success of second-year economics students an exploratory analysis." *South African Journal of Economics*, **79**(2).

Knezevic A (2008). *Overlapping Confidence Intervals and Statistical Significance. StatNews n.73.* Cornell Statistical Consulting Unit. Cornell University.

Koenker R (1994). "Confidence intervals for regression quantiles." In M Huskova (ed.), *Asymptotic Statistics Proceedings of the 5th Prague Symposium on Asymptotic Statistics* ., pp. 349–359. Heidleberg Physica-Verlag.

Koenker R (2004). "Quantile Regression for Longitudinal Data." *Journal of Multivariate Analysis*, **91**, 74–89.

Koenker R (2005). *Quantile Regression*. Cambridge University Press.

Koenker R, Bassett G (1978). "Regression quantiles." *Econometrica*, **46**, 33–50.

Koenker R, d'Orey V (1987). "Computing regression quantiles." *Applied Statistics*, **36**, 383–393.

Koenker R, d'Orey V (1994). "Remark AS R92. A Remark on Algorithm AS." *Applied Statistics*, **229**, 410–414.

Logan J, Lundberg O, Roth L, Walsh K (2017). "The effect of individual motivation and cognitive ability on student performance outcomes in a distance education environment." *Journal of Learning in Higher Education*, **13**(1), 83–92.

MIUR (2017). *Gli immatricolati nell'a.a. 2016/2017 il passaggio dalla scuola all'università dei diplomati nel 2016*. Focus.

Seltzer M, Choi K (2002). "Model checking and sensitivity analysis for multilevel models, In N. Duan and S." In N Duan, S Reise (eds.), *Multilevel modeling Methodological advances, issues, and applications*. Hillsdale, NJ Lawrence Erlbaum.

Tattersall C, Waterink W, Hoppener P, Koper R A (2006). "case study in the measurement of educational efficiency in open and distance learning." *Distance Education*, **27**, 391–404.

Van Bragt CAC, Bakx AWEA, Bergen TCM, Croon MA (2011). "Looking for students personal characteristics predicting study outcome." *Higher Education*, **61**, 59–75.

Affiliation:

Giovanni Boscaino and Giada Adelfio
Dipartimento di Scienze Economiche, Aziendali e Statistiche
University of Palermo
Palermo, Italy giovanni.boscaino@unipa.it, giada.adelfio@unipa.it

Journal of Environmental Statistics http://www.jenvstat.org
Volume 8, Issue 4 Submitted: 2017-12-19
September 2018 Accepted: 2018-09-07