



dSEAS  
dipartimento  
scienze economiche  
aziendali e statistiche  
department  
of economics  
business  
and statistics

# Working Papers

ISSN 'in fase di assegnazione', volume I, 2017

## University student *talent*: the real driver for performance?

Giovanni Boscaino · Giada Adelfio

**Abstract** Investigation about the university student performance, and its measurement, are very crucial issues for any policy maker. Since the economic crisis, jobs market requires even higher skills and competences. Literature offers a lot of papers about the university student quality and performance, in order to identify the main determinants of them. Often, results are very different, and they seems to hold just in a specific context. This paper aims to investigate the role of a latent variable that can take into account the student motivation, aptitude, and abilities, here conveniently called *talent*. A random effect Quantile Regression on a new measure of Italian student performance has been adopted, and results seem to highlight the main role of the *talent*.

**Keywords** Student performance · Indicator · Random effects Quantile Regression

**Riassunto** *Gli studi sulla performance delle carriere degli studenti sono oggetto di grande attenzione sia da parte della comunità scientifica sia da parte dei policy maker. L'attenzione negli ultimi anni sembra essere cresciuta, a pari passo con quella della valutazione dei sistemi formativi: la valutazione dell'efficacia e dell'efficienza di un servizio di formazione passa anche dalla valutazione delle performance degli studenti. Pertanto, l'esigenza di formare studenti ben preparati e in tempo è diventata una priorità per tutto il mondo accademico. La letteratura*

G. Boscaino: Dipartimento di Scienze Economiche, Aziendali e Statistiche

Università degli Studi di Palermo

viale delle Scienze ed. 13, 90128

E-mail: giovanni.boscaino@unipa.it

· G. Adelfio: Dipartimento di Scienze Economiche, Aziendali e Statistiche

Università degli Studi di Palermo

viale delle Scienze ed. 13, 90128

E-mail: giada.adelfio@unipa.it

scientifica offre diversi spunti di riflessione circa le dinamiche e le determinanti di una buona o cattiva performance studentesca ma i risultati non sembrano convergere verso un'unica direzione. Quello che emerge è che da un lato il contesto (area geografica/economica) ha un ruolo fondamentale, dall'altro la non convergenza verso un insieme di fattori determinanti (ad eccezione del Genere che risulta significativo nella maggior parte degli studi) potrebbe suggerire che si stia guardando nella direzione sbagliata. Lo studio qui riportato si inserisce in questo contesto: utilizzando le "classiche" variabili esplicative disponibili per lo studio del successo universitario, è possibile trovare indicazione del fatto che queste non sono sufficienti? In altre parole, il successo universitario può essere spiegato da caratteristiche intrinseche dello studente, che qui chiameremo genericamente "talento", piuttosto che da quelle socio-demografiche? A tal proposito, si è deciso di adottare una misura della performance universitaria alternativa ai Crediti Formativi Universitari (CFU) accumulati e ai voti conseguiti: tale misura tiene conto contemporaneamente del voto e del CFU così che, sempre in una scala da 18 a 30, l'unico valore espresso sia portatore di entrambe le informazioni. Il principio è che guardare solo ai CFU accumulati nel tempo non permette di valutare quali CFU siano stati accumulati, e che conseguire un 30 per un insegnamento di 3 CFU non ha lo stesso valore che conseguire un 30 in un insegnamento da 12 CFU. La misura proposta in Adelfio et al. (2014) tiene conto di entrambi gli aspetti. Tale nuova misura è stata oggetto di un'analisi condotta secondo un approccio di Regressione Quantilica (RQ) ad effetti casuali, dove la variabile risposta considerata è la nuova misura rilevata per ogni esame di ciascuno studente mentre quelle esplicative sono state il Genere, il Tipo di Diploma, la Residenza, il Voto di Diploma, lo Status di "In corso"- "Fuori corso" dello studente alla laurea, e il Corso di Laurea di immatricolazione. I dati hanno riguardato la coorte di immatricolati nel 2002 che si sono laureati entro 7 anni ai Corsi di Laurea (CdL) in Economia e Finanza e in Biologia dell'Università degli Studi di Palermo (Italia). L'approccio scelto è stato preferito per due motivi: da un lato la RQ consente di studiare dipendenze che non siano solo "in media" ma anche in altri punti della distribuzione; dall'altro, l'approccio per misure ripetute ha consentito di tenere in considerazione la variabilità della performance intrinseca di ogni studente. I risultati hanno evidenziato che la variabile più significativa è stata il CdL, piuttosto che quelle socio-demografiche. Secondo l'ottica descritta in precedenza, il CdL scelto dallo studente potrebbe essere considerato come una proxy delle personali inclinazioni e attitudini, della sfera di interessi, capacità, abilità e indole dello studente: il suo "talento". Probabilmente la selezione e l'orientamento degli studenti verso un Corso di Laurea dovrebbero tenere in maggiore considerazione le reali aspettative, capacità e attitudini dei soggetti: questa potrebbe essere la chiave per ridurre gli abbandoni, conseguire il titolo nei tempi regolari e con voti migliori. In futuro, il presente studio sarà ripetuto su altre coorti e Corsi di Laurea al fine di valutare la robustezza dei risultati osservati.

**Parole chiave** Performance dello studente - Indicatore - Regressione Quantilica ad effetti casuali

## 1 Introduction

Academic student performance is a crucial issue for the university policy makers, today more than ever. The global job market needs more and more competitiveness, and high skills and competences, therefore an improved quality of the graduates is required. Obviously, the definition of graduate's quality is a difficult matter. It could concern the actual student ability in solving practical problems, or the wideness and depth of his/her knowledge, or the number of years to graduation (in those countries where there are no time limits to get the degree), etc. or a combination of them. In this paper the attention is more devoted to the student performance at university, instead of his/her quality. But its measurement is really challenging too. Someone uses the final grade, others refer to the distribution of examinations marks, or pay attention to the credits earned (usually related to time, or marks, or both), or, again, to the time spent to get the degree. Or all of them. There is not a shared and acknowledge measure (and measurement methodology) of the student performance. In fact, literature offers several studies about student performance, mainly devoted to find its determinants, and, despite they are often based on the same measure, results address sometimes to different directions. Just to quote some of most recent ones, Cheesman et al. (2006) 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. Tattersall et al. (2006) 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, e.g. Learning interruption and Changes of the Study Programme (SP). The influence of the “Change of SP” on the expected time to graduation was also analysed by Adelfio and Boscaino (2015), highlighting its significant and negative effect. Birch and Miller (2006) used a Quantile Regression approach identifying in Tertiary Entrance Rank, Gender, and High School the most important determinants of high and low performance. Boscaino et al. (2007) 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. Van Bragt et al. (2011) 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, Jansen, and Yu (Horn et al.) performed an exploratory analysis on the determinants of success of second-year students. The authors asked if the factors that leads 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. (2013) 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. Grilli et al. (2013) 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, binomial mixture model): they highlighted the poor role on the Pre-enrolment test as predictor of number of credits. Adelfio et al. (2014) proposed a new measure for student performance, and Quantile Regression results showed no significant effect of the social and demographic variables, but just of the Attended SP.

In conclusion, at first sight, the literature generally suggests 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.

This paper considers the Italian context, and aims to highlight a new study perspective: what if the real determinants for the good performance of the students were their motivations or attitudes, instead of their socio-demographic characteristics? That question arises from some considerations about particular results (here reported) obtained by a Quantile Regression approach on a recent new measure of student performance (Adelfio et al., 2014), aimed to identify its possible determinants.

Therefore, the paper is organised as follow: Section 2 recalls the new measure for the student performance, introduced in Adelfio et al. (2014); Section 3 is devoted to a brief description of the Quantile Regression model, used to investigate the determinant of the performance; in Section 4 some consideration arose from the results showed in section 3 are reported, and the random effect QR adopted model is illustrated in Section 5; last two Sections (6 and 7) discuss the results of the analysis and some remarks.

## 2 The measure of student performance

Student performance at university is often simply measured by the exam marks. But, obviously, same marks got in different classes play a different role for the final grade due to the different workload of the classes. The workload is usually measured by the credits. For the sake of simplicity, and because this paper refers to the Italian university context, next example is based on the Italian university grading system. In Italy, universities use a 30-point scale system with 0-17 as non passing grades and 18-30 as passing ones, with 30 the maximum (for outstanding results, the “lode” is added to the maximum in order to praise the real deserving student). Therefore, the grade 28 got in a course with 10 credits weights more then a 28 got in a 1 credit course. Indeed, the final grade is a weighted mean of the exam grades, with credits as weights. If we want to study the student performance, we have to use the couple grade-credits. Adelfio et al. (2014) proposed a new unique measure of the student performance, still based on grades and credits, that takes values in the same quantitative grading scale used by a given Country C:

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

where  $best_C$  and  $su\!f_C$  are the marks that correspond to the best and to the minimum passing mark stated in the country C system, respectively;  $m_{ij}^w = \frac{m_{ij}Cr_j}{\sum_{j=1}^J Cr_j}$ , with  $m_{ij}$  is the grade got by the student  $i$  and  $Cr_j$  is the credit for the course  $j$ .

In such a way, each country can measure the performance of its students accounting both for marks and credits in one only measure ( $m_{ij}^w$ ), and adopting an indicator that still gets values in the same original scale for passing marks ( $m'_{ij}$ ).

Figure 1 illustrates the effect of the new indicator. Our 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. The top and bottom plots on the left report the distributions of marks conditioned by credits for the two different degree courses (E and L, respectively). The E box plots show a more variability for the mark means across the credits than L ones. While the distributions of marks for L are highly negative asymmetric for all the credits, the distributions of marks for E are very different, showing high mean grades in correspondence to the low credits and lower mean grades 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 rescaled accounting for the workload of the courses penalizing high grades got in low credit courses with respect to the low marks got for the high credits courses. The distributions of new marks are more fair and easily comparable.

The new indicator is the object of the analysis reported in this paper. First of all a Quantile Regression model is adapted to our data in order to investigate the determinants of the performance. Briefly, results give evidence to consider a new perspective of analysis and a random effect Quantile Regression model is considered.

### 3 The Quantile Regression model

The crucial aim of the paper concerns the investigation of the determinants of the student performance. We refer to the Quantile Regression (QR) approach in order to investigate the influence of some determinants over the whole shape of distribution of the proposed indicator (1).

QR (Koenker, 2005) 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 method of Least Squares results in estimates that approximate 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. This allows comparing how some percentiles of the variable of interest may be more affected by certain subject characteristics than other percentiles.

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

Therefore, the  $\phi$ th conditional quantile function can be estimated by solving:

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

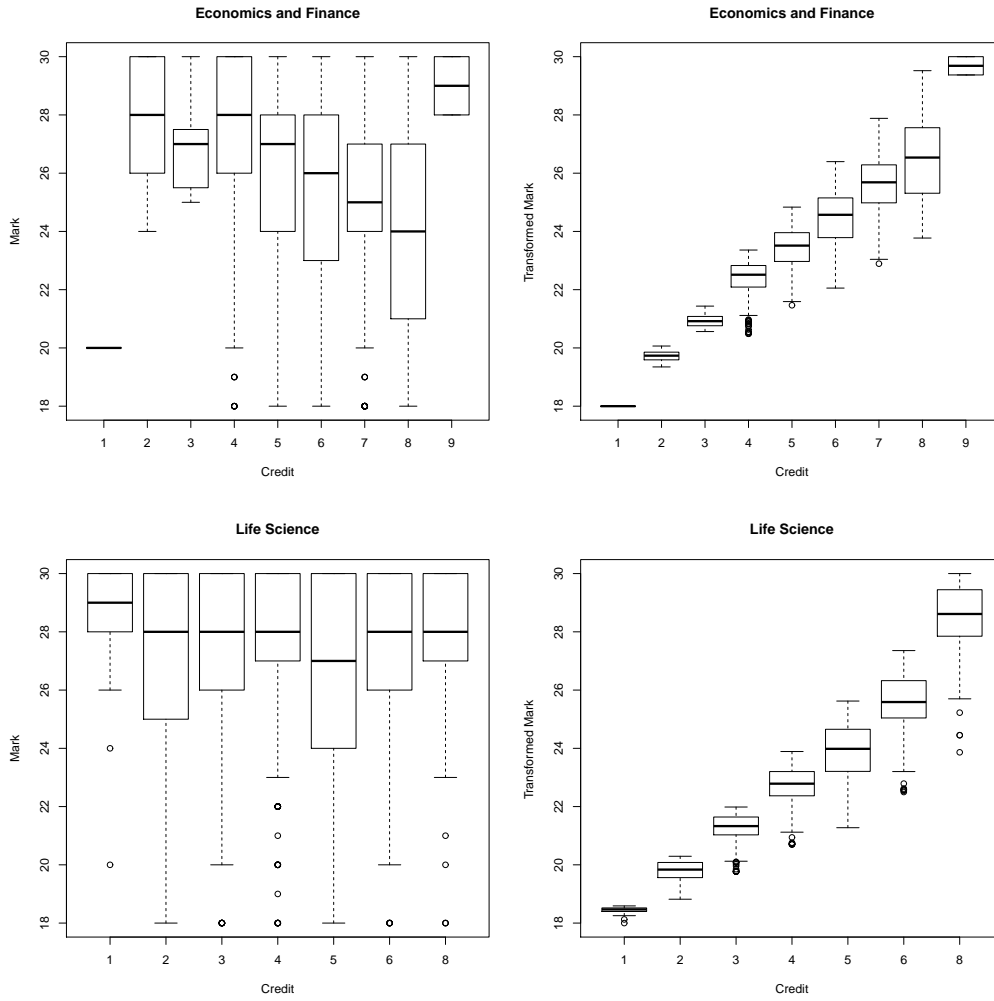


Fig. 1: Grades distributions MMM conditioned by credits, for E and L Study Programme

where the function  $\rho_\phi(\cdot)$  is the tilted absolute value function, that yields the  $\phi$ th 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  $f_\phi(\mathbf{x}) = \mathbf{x}^T \beta_\phi$  where  $\beta_\phi = (\beta_{\phi,1}, \beta_{\phi,2}, \dots, \beta_{\phi,d})^T$ , such that the conditional  $\phi$ th quantile function  $f_\phi(x)$  is a linear function of the parameters  $\beta$ , a linear quantile regression is considered.

The QR estimates are obtained by the R package `quantreg` (Koenker, 2012), that considers by default the modified version of the Barrodale and Roberts algorithm described in Koenker and d'Orey (1987) and Koenker and d'Orey (1994). We follow that approach because of the dimension of the analysed data set, since 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 Dataset introduced in Section 2 reports information about course credits and marks for each student, but it includes also other variables: Gender (Female vs Male), High School (Lyceum vs Not Lyceum), Residence (Palermo vs Not Palermo), and Diploma mark (centred at the mean). In addition, since in Italy the legal duration of the first level Study Programme (SP) is 3 years, but it is not compulsory (then, for example, a very unwilling student can get the degree 9 years after matriculation), we considered the Student Status at graduation as variable that classified as “On-Time” (OT) those students that get the degree in 3 years, and as “Out-of-time” (OOT) the others.

A preliminary linear QR model on  $Me'_i = Me_i(m'_{i,j})$  (not here reported for the sake of brevity) was performed on E and L separately, and it suggested that the only covariate with a significant effect was the Student Status at graduation. Therefore, we considered the linear QR model conditioned to the Status and results showed no significant variable with respect to the OT graduates. The performance of the students who get their degree on time do not seem to depend on the considered social and demographic features. It seems that good performer students are just good, and their positive performance may be ascribed to own motivation, inclination, method of study, etc.. Otherwise, the analysis of the OOT students (here reported for both E and L) showed some covariates are significant, only with respect to some percentiles. Conditioning to five quantiles  $\tau = (0.05, 0.25, 0.50, 0.75, 0.95)$ , the parameter estimations for the OOT students are reported in Table 1.

In short, the Intercept of the models represent the estimated conditional quantile of  $Me'_i$  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. Whereas the OLS results inform about the conditional mean of the response variable – given certain values of the predictor variables – QR aims to estimate the fixed quantiles of the response variable, using different measures of central tendency (and statistical dispersion), in order to obtain a more comprehensive analysis of the relationship between variables. Indeed, the QR analysis allows to interpret results also for the tails of the distribution (excellent and

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

| <b>E cohort</b> | OLS     | $\tau_1$ | $\tau_2$ | $\tau_3$ | $\tau_4$ | $\tau_5$ |
|-----------------|---------|----------|----------|----------|----------|----------|
| 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**   |
| <b>L cohort</b> |         |          |          |          |          |          |
| 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  $\alpha = 0.1$ )

mediocre students), instead of focusing just on the “average student”. More detailed comments about the QR results for E are in Adelfio et al. (2014), since in this section, we just highlight the comparison between E and L results. 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, for the L cohort 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. L cohort shows a higher effect on performance than the E cohort and, 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).

Following a different point of view, we can study if there is a statistical difference among the estimated values of the coefficients for E and L. We restrict the comparison to the couples of significant coefficients, conditioned to 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, Knezevic), it is possible to notice (tab. 2) that Gender has always a different effect (for E and L), conditional to the same significant percentiles. At the opposite, considering the Residence, it happens for just the 14% of the comparisons.



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

| High School | Residence | Gender | Dipl. mark |
|-------------|-----------|--------|------------|
| 0.50        | 0.14      | 1.00   | 0.25       |

#### 4 Some reflections on QR results

Results showed in previous paragraph need some reflections. First of all, among the considered covariates, just the OOT Status seems to play a role on the student performance. Secondly, with respect to the OOT students, just a little effect of their characteristics is noticed, and in particular just for few quantiles. Therefore, performance could be affected by some not-here-considered social, demographic, economic, and/or by some latent student characteristics. In particular, E and L are two different SPs with respect to course subjects (in Italy, L courses subjects could be more specific than E ones) and to job market opportunities (in Italy, E gives more wide-ranging knowledge than L, hence E graduates can have access to a larger set of jobs than L graduates). Therefore, it is plausible that L and E students are different in motivation, basic knowledge, abilities, and aptitudes. We call these aspects with one unique convenient word: *talent*. Then, is the student *talent* a determinant for his/her university performance? In order to answer to this question, a different (from the usual regression) perspective has been followed. Instead of investigating the effect of the covariates on just one single synthetic measure of the students marks distribution, via the Quantile Regression approach it is possible to analyse the covariates effect on different parts (the quantiles) of that distribution. The *talent* will be highlighted by the comparison results between E and L students.

#### 5 Quantile Regression for repeated measurements

In the light of the previous considerations, a linear mixed QR approach is considered in order to take into account the whole mark distribution for each student. In such a way, it is possible to focus on the subject specific variability.

More formally, let  $(\mathbf{x}'_{ij}, y_{ij})$ , for  $j = 1, \dots, n_i$  and  $i = 1, \dots, N$ , 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:

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

where  $G_{y_{ij}|u_i}(\cdot) \equiv F_{y_{ij}|u_i}^{-1}(\cdot)$  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 and identically and independently distributed according to some density  $f_u$ , usually  $u_i \sim N(0, \alpha)$ , characterized by a  $\tau$ -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 as 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  $\mu_{ij} = \mathbf{x}'_{ij}\beta + u_i$  is the linear predictor of the  $\tau$ th quantile, fixed and known, and  $\sigma$  is the usual scale parameter. The random effects, that induce a correlation structure among observations on the same subject, are assumed to be independent. That is a likelihood-based approach to the estimation of the QRs based on the ALD and it is better than the penalized fixed effects based approach in terms of mean squared error of the QR estimators. Alternative models with non-normally distributed residuals were developed (Seltzer and Choi, 2002).

The ALD approach has been considered as it provides an automatic choice of the optimal level of penalization and also because it represents a suitable error law for the least absolute estimator (and therefore a natural choice in QR).

## 6 Data Analysis via QR mixed model

We apply the linear QR model with a subject-specific random intercept that accounts for the within-group correlation (3) with respect to the two formerly considered cohorts of students graduated in E and L, managed together in a unique dataset. The analysis was performed using the R package `lqmm`: Linear Quantile Mixed Models (Geraci, 2014).

In Figure 2, results are reported with respect to the fixed coefficients estimates. To assess the suitability of (3), results are also commented in the light of those reported in the previous paragraphs. For each of the estimated coefficients we plot the QR estimates of the fixed parameters of (3), conditional to each quantile  $\tau$  ( $\tau = 0.05, 0.25, 0.50, 0.75, 0.95$ ), by the dashed curve with filled dots. These points 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, graduated On-time 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. 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 – in most of the quantiles – of the covariates. This could be a partial confirmation that

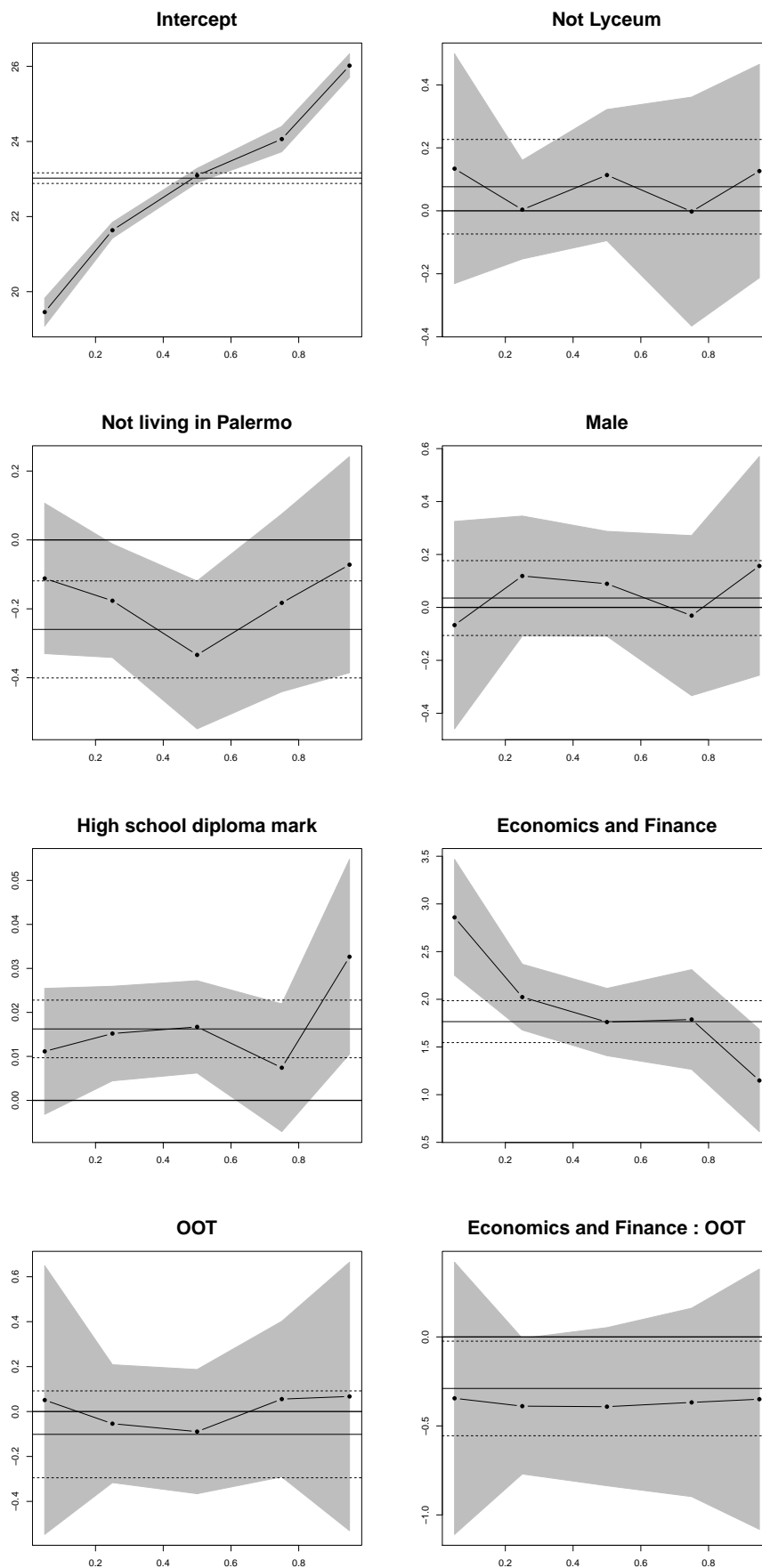


Fig. 2: Fixed coefficients estimates of random intercept QR model

Table 3:  $\tau$  dependent estimated variance of  $u_i$  random intercept

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

student performance seems to be mainly due to students own *talent* rather than social and demographic characteristics.

The random intercept aims to catch the subject specific *talent* effect on performance. In fact, the  $\tau$  dependent estimated variances of  $u_i$  random intercepts of (3) are all in (0.36, 0.62) and they reflect the heterogeneity among students due to their own unobservable *talent* (tab.3).

Results from linear mixed QR, then, seem to confirm our previous assumption, that is the only variable that has a significant effect in the performance study is the SP choice. In other words the only relevant aspect from this analysis is the choice of the SP, easily ascribable to the personal inclinations and subjective sphere of any student (the *talent*).

## 7 Final remarks

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

This paper is based on first results reported in Adelfio et al. (2014): these led to a new investigation perspective, for analysing the determinants of the students performance.

For this reason the choice of students about the SP has been considered in order to catch the effect of a latent variable for their subjective characteristics (such as aptitude, motivation, or inclination) on the overall performance.

In order to have comparable results, also among different SPs, we needed a measure like the one in (1) that accounts in a unique quantity both for the workload and marks. Results from the linear mixed QR seem to confirm our previous assumption: in fact the only variable that has a significant effect on the performance is the SP.

In other words, the chosen SP could be a proxy of the personal inclinations and subjective sphere of any student.

In the light of these, though ascribed to two big SPs of the University of Palermo, in order to improve the students performance, probably policy makers should look at the student personal abilities, aptitudes, motivation, and *talent*, 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, therefore, to help them in following their own *talent*).

As future work, more Sps will be considered to assess the robustness of the current results.

## References

- Adelfio, G. and G. Boscaino (2015). A mixed effect model for overdispersed data to analyse the effect of the degree course change on the expected time to graduation. *Journal of Applied Statistics*, 10.
- Adelfio, G., G. Boscaino, and V. Capursi (2014). A new indicator for higher education student performance. *Higher Education* 68(5), 653–668.
- Attanasio, M., G. Boscaino, V. Capursi, and A. Plaia (2013). May the students career performance helpful in predicting an increase in universities income. pp. 9–16. series in studies in classification, data analysis, and knowledge organization, Switzerland Springer International Publishing: Statistical models for data analysis.
- Birch, E. R. and P. W. Miller (2006). Student outcomes at university in australia a quantile regression approach. *Australian Economic Press* 45(1), 1–17.
- Boscaino, G., V. Capursi, and F. Giambona (2007). The careers' performance of a university students' cohort. Technical report, DSSM Working paper, n. 2007.1.
- Cheesman, J. S., N. Simpson, and G. Wint (2006). *Determinants of student performance at university Reflections from the Caribbean*.
- Geraci, M. (2014). Linear quantile mixed models the lqmm package for laplace quantile regression. *Journal of Statistical Software* 57(13), 1–29.
- Geraci, M. and M. Bottai (2006). Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics* 8, 140–154.
- Grilli, L., C. Rampichini, and R. Varriale (2013). Predicting students academic performance a challenging issue in statistical modelling. CLEUP: Cladag 2013 Book of abstracts.
- Horn, P., 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).
- Knezevic, A. *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. (2012). quantreg quantile regression. *R package version 4.9*.
- Koenker, R. and G. Bassett (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Koenker, R. and V. d’Orey (1987). Computing regression quantiles. *Applied Statistics* 36, 383–393.
- Koenker, R. and V. d’Orey (1994). Remark as r92. a remark on algorithm as. *Applied Statistics* 229, 410–414.
- Seltzer, M. and K. Choi (2002). Model checking and sensitivity analysis for multilevel models, in n. duan and s. In N. Duan and S. Reise (Eds.), *Multilevel modeling Methodological advances, issues, and applications*. Hillsdale, NJ Lawrence Erlbaum.
- Tattersall, C., W. Waterink, P. Hoppener, and A. Koper, R. (2006). case study in the measurement of educational efficiency in open and distance learning. *Distance Education* 27, 391–404.
- Van Bragt, C. A. C., A. W. E. A. Bakx, T. C. M. Bergen, and M. A. Croon (2011). Looking for students personal characteristics predicting study outcome. *Higher Education* 61, 59–75.