Student Performance in Mathematics in Portugal: a multilevel quantile regression model

Susana Faria
CMAT - Centre of Mathematics, Department of Mathematics and Applications, University of Minho, Portugal;
Maria C.S. Portela
CEGE - Centro de Estudos em Gestão e Economia da Católica Porto, Business School, Portugal.

Abstract

This paper is based on Portuguese data from PISA-2009, and it focuses on the measurement of student achievement in mathematics and on the determinants of this achievement both at the student and at the school levels. Data on about 3900 Portuguese students and 194 schools who participated in PISA-2009 were used to accomplish our objectives. We used a multilevel quantile regression model to analyse the determinants of students' success, where the potential determinants are student and school variables. Our study provides evidence that a stable relation over quantiles of achievement is expected for some variables (e.g. gender, repetition, or socio economic background), while other variables show varying impacts depending on the students' location on the rank of achievement in math (e.g. immigrant status of students, or some study strategies like control strategies). In spite of schools having a significant impact on students' achievement, we found that most school-level variables (except location) were not significant in explaining the school effect.

Introduction

The Programme for International Student Assessment (PISA) is an international study launched by the Organization for Economic Co-operation and Development (OECD) in 1997. It aims at evaluating education systems worldwide by assessing 15-year-olds' competencies in three key subjects (reading, mathematics and science) every three years. Studies based on PISA data to explain student performance are very common in literature. Examples of country specific analysis using PISA results can be found in Agasisti and Cordero-Ferrera (2013), where Italian and Spanish students were analysed in PISA 2006 through multilevel models, in Alacaci and Erbas (2010), where Turkish students were analysed, also through multilevel models and using the same PISA dataset, or in Mancebón et al. (2012) who analysed Spanish students' science scores in PISA2006 through multilevel models. Analysing PISA-2003 mathematics test scores, Martins and Veiga (2010) show that in Portugal 37% of the variation in students' achievement is explained by school differences, whereas a value around 60% is found for Austria, Belgium Netherlands and Germany, while for Finland it is only 5%.

This work analysed the performance in PISA-2009 Math test scores of Portuguese students through a multilevel quantile regression model. The objective of the analysis was twofold. On the one hand we wished to understand the drivers of students' success in math and whether the impact of these drivers differed for students located in different positions of the ranking of test scores; and on the other hand, we wished to understand the magnitude of school effects, and the extend to which these effects could be explained by some school level variables.
Methodology

Recently, Geraci and Bottai (2014) have introduced a new method for quantile regression with mixed effects called linear quantile mixed models. They propose a conditional quantile regression model for continuous responses where random effects are added to the model taking into account the dependence between units when an hierarchical data structure is present.

We have adopted the procedure proposed by Geraci and Bottai (2014) to perform a multilevel quantile regression model.

Assume that we have data from J schools, each with a different number of students $n_j$. Consider data in the form $(X_{ij}, y_{ij})$ for $i = 1, \cdots, n_j$ and $j = 1, \cdots, J$, $N = \sum_{j=1}^{J} n_j$; where $X_{ij}$ is a vector of the of student level variables (student i attending school j) and $y_{ij}$ are student PISA scores in mathematics. We considered the random intercept quantile regression model

$$y_{ij} = \beta_{0j}^\tau + X_{ij}^T \beta^\tau + \varepsilon_{ij}$$

$$\beta_{0j}^\tau = \gamma_{00}^\tau + u_{0j},$$

where $\beta^\tau$ is a vector of unknown fixed effects, $\beta_{0j}^\tau$ is the intercept representing the average achievement for the j school (this intercept varies at the school level), $\gamma_{00}^\tau$ is the average achievement of the school means, and $u_{0j}$ is the random effect associated with school j. The dependence among the students within the j-th school is induced by the random effect $u_{0j}$ which is shared by all students within the same school. We assume that $y_{ij}$ conditionally on $u_{0j}$ are independently distributed according to an Asymmetric Laplace distribution with location and scale parameters given $\mu_{ij} = X_{ij}^T \beta^\tau + u_{0j}$ and $\sigma^\tau$. The skew parameter $\tau$ is set a priori and defines the quantile level to be estimated. Also, we assume that the random effects $u_{0j}, j = 1, \cdots, J$ are mutually independent and identically distributed according to some density $f(u_{0j} | \varphi_u^\tau)$ where $\varphi_u^\tau$ is a scale parameter. We assume the $\varepsilon_{ij}$ are independent and $u_{0j}$ and $\varepsilon_{ij}$ are independent of one another.

School level variables ($W_j$) can be introduced in the model for explaining school effects. In that sense the equation above becomes

$$\beta_{0j}^\tau = \gamma_{00}^\tau + W_j^T + u_{0j}$$

In applying the multilevel quantile model we first applied model (1), estimating school effects without taking into account school characteristics. Then we re-estimated the model with school level variables (using (2)). The aim of this second step estimation was to understand which school level variables revealed significant in explaining school effects.

Conclusions

The main results of our analysis point to the existence of some stable and important effects of gender, socio-economic background, and repetition, and important and growing (with the rank position of the student) effects of the grade attended (advantage for students attending the secondary education), and the students’ expectations. Interestingly study strategies appeared significant in explaining achievements, with most of them showing stable impacts across different students' rank in math, except for control strategies that proved more beneficial for students with the lowest math scores. Two variables showed counter-intuitive signs (the existence of tutoring classes outside the school and the help in homework by parents). The negative impact of these variables is likely to be related with endogeneity problems, since the
negative sign identifies a negative relationship between the variables and not a cause-effect relationship.

From our analysis it became apparent that schools still played an important role in explaining students’ success in math even after student-level variables being accounted for. The highest impact of schools happened for students in the 50% quantile, with schools explaining about 9% of the variability in test scores. The lowest impact of schools happened in the extreme quantiles meaning that for very good or very ‘bad’ students the school attended does not seem to make as much difference as for students in the middle of the distribution. We also show in this paper that Portuguese schools can have very different effects on student’s achievement, with schools showing math scores above expected in all quantiles considered, and others showing math scores below expected.

Acknowledgements

The authors acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT) through project PTDC/GES/68213/2006. This research was financed by Portuguese Funds through FCT – “Fundação para a Ciência e a Tecnologia”, within the Project UID/MAT/00013/2013.

References


