Student background determinants of reading achievement in Italy. A quantile regression analysis

Francesca Giambona
Mariano Porcu

Pre-print version

Cite this article as:

Francesca Giambona, Mariano Porcu, Student background determinants of reading achievement in Italy. A quantile regression analysis, *International Journal of Educational Development*, Volume 44, 2015, Pages 95-107, https://doi.org/10.1016/j.ijedudev.2015.07.005. (https://www.sciencedirect.com/science/article/pii/S0738059315000814)

Elsevier Editorial System(tm) for International Journal of Educational Development
Manuscript Draft

Manuscript Number: EDEV-D-14-00378R2

Title: Student Background Determinants of Reading Achievement in Italy. A Quantile Regression Analysis

Article Type: Full Length Article

Keywords: PISA 2009, quantile regression, students' achievement

Corresponding Author: Dr. Francesca Giambona,

Corresponding Author's Institution:

First Author: Francesca Giambona

Order of Authors: Francesca Giambona; Mariano Porcu

Abstract: In recent years determinants of students' achievement has received much attention. Empirical studies have found that students' characteristics, family background, school attended, and regional residence are major factors affecting student performance. In this paper, we analyse the 2009 OECD-PISA (spell PISA) survey to examine individual
background characteristics in uencing the reading achievement of Italian 15 years-old students using the Quantile Regression (QR) approach. The QR approach allows researchers to analyse changes in size and direction of
predictor estimates on student performance across the entire distribution of reading achievement scores. Results indicate significant effects of predictors on reading achievement operating differently across quantiles, suggesting different pathways to achievement for low and high performing readers. In particular, some family background predictors (parental education, computer
availability at home, and availability of a desk for homework at home), the school program attended and, the region of student residence play important but differing role for low and high performing readers. For example, parental education shows a positive effect on student reading, academic (general) programs perform better than vocational or technical, and Northern regions perform better than Center-Southern ones, with differentiated effects along the distribution of students' reading scores. These findings should be carefully considered by policymakers when outlining strategies to enhance student
performance at all levels along the reading continuum of low and high scores.

*Title Page (WITH author details)

Student Background Determinants of Reading Achievement in Italy. A Quantile Regression Analysis

Francesca Giambona and Mariano Porcu
Università degli Studi di Cagliari

Correspondence concerning this article should be addressed to
Francesca Giambona
Dipartimento di Scienze Sociali e delle Istituzioni,
Università degli Studi di Cagliari,
Viale S. Ignazio 78, 09123 Cagliari, ITALY.
E-mail: francesca.giambona@unica.it

Abstract

In recent years determinants of students' achievement has received much attention. Empirical studies have found that students' characteristics, family background, **school attended, and regional residence** are major factors affecting student performance. In this paper, we **analyse** the 2009 OECD-PISA (spell PISA) survey to examine **individual background characteristics influencing the reading achievement of Italian 15 years-old students using the Quantile Regression (QR) approach. The QR approach allows researchers to analyse changes in size and direction of predictor estimates on student performance across the entire distribution of reading achievement scores. Results indicate significant effects of predictors on reading achievement operating differently across quantiles, suggesting different pathways to achievement for low and high performing readers. In particular, some family background predictors (parental education, computer availability at home, and availability of a desk for homework at home), the school program attended and, the region of student residence play important but differing role for low and high performing readers. For example, parental education shows a positive effect on student reading, academic (general) programs perform better than vocational or technical, and Northern regions perform better than Center-Southern ones, with differentiated effects along the distribution of students' reading scores. These findings should be carefully considered by policymakers when outlining strategies to enhance student performance at all levels along the reading continuum of low and high scores.**

*Manuscript (WITHOUT author details)
Click here to view linked References

Running head: QR ANALYSIS OF STUDENTS' READING SCORES

Student Background Determinants of Reading Achievement in Italy. A Quantile Regression Analysis

Abstract

In recent years determinants of students' achievement has received much attention. Empirical studies have found that students' characteristics, family background, **school attended, and regional residence** are major factors affecting student performance. In this paper, we **analyse** the 2009 OECD-PISA (spell PISA) survey to examine **individual background characteristics influencing the reading achievement of Italian 15 years-old students using the Quantile Regression (QR) approach. The QR approach allows researchers to analyse changes in size and direction of predictor estimates on student performance across the entire distribution of reading achievement scores. Results indicate significant effects of predictors on reading achievement operating differently across quantiles, suggesting different pathways to achievement for low and high performing readers. In particular, some family background predictors (parental education, computer availability at home, and availability of a desk for homework at home), the school program attended and, the region of student residence play important but differing role for low and high performing readers. For example, parental education shows a positive effect on student reading, academic (general) programs perform better than vocational or technical, and Northern regions perform better than Center-Southern ones, with differentiated effects along the distribution of students' reading scores. These findings should be carefully considered by policymakers when outlining strategies to enhance student performance at all levels along the reading continuum of low and high scores.**

Student Background Determinants of Reading Achievement in Italy. A Quantile Regression Analysis

Introduction

Education **plays a key role in** individuals' life chances and **promotes** the economic development of countries by enhancing productivity, social development, and reducing social inequality. Higher education is associated with markedly higher earnings, lower unemployment, higher labour force participation and lower criminality; where a high performing educational system is taken to be fundamental in achieving national economic competitiveness (OECD, 2012 ; Hanushek & Luque, 2003). Due to the technological progress, rising skill demands have made qualifications at the upper secondary level of education (general, technical or vocational) the minimum credential for successful labor market entry (Rangvid, 2003).

The goal of education has shifted its emphasis from the collection and memorization of information only, to the inclusion of a broader concept of knowledge. The meaning of "knowing" has shifted from being able to remember information, to being able to find and use it (Simon, 2000). The ability to access, understand and reflect on all kinds of information is essential if individuals are able to participate fully in our knowledge-based society. More specifically, reading literacy is considered an essential skill for future literacy and it is about understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop knowledge needed to participate in society. Reading literacy includes a wide range of cognitive competencies, from basic decoding, to knowledge of words, grammar and larger linguistic and textual structures and features, to knowledge about the

world (OECD, 2010, 2011).

Reading achievement is not only a bedrock for achievement in other subject areas within the educational system, but also a prerequisite for successful participation in most areas of adult life (Cunningham & Stanovich, 1998 ; Smith, Mikulecky, Kibby, Dreher, & Dole, 2000). Similarly, reading skills are essential to the academic achievement of middle- and high- school students (Holloway, 1999). Furthermore, reading literacy provides access to modern social institutions and has an impact on cognition, or thinking processes as it also shapes the way in which we think (Kern & Friedman, 2008 ; Olson, 1977 ; Pretorius, 2000). Assessing the reading literacy of students, therefore, focuses on reading literacy skills that include finding, selecting, interpreting and evaluating information from the full range of texts associated with real life situations that reach beyond the classroom (OECD, 2011). But, what makes a student or a school system successful in terms of literacy or competence? Many studies have analyzed the determinants of students' achievement using the standard regression methods based on the Ordinary Least Squares (OLS) to estimate the effect of predictor variables on the students' achievement measured, usually, as a test score **in mathematic, sciences or reading**. Since OLS estimators show the effect of predictor variables at one point of the distribution of the dependent variable (the conditional mean) the information gathered by OLS regression is limited to this specific point of the distribution. In terms of achievement this can lead to incomplete findings when the effects of predictors vary at points along the distribution, i.e. at different quantiles.

This analysis identifies **individual background determinants of reading success** using the last PISA 2009 survey through a Quantile Regression (QR) approach. QR allows us to describe the effect of predictor variables along the entire students'

achievement scores distribution, when educators are interested in assessing the uniformity of changes by predictor influencing the entire range of reading skills. In particular, the investigator aims to identify which variables affect the lower part of the distribution (below the median value) and on the opposite the upper part (above the median). This is done by calculating coefficient estimates at various quantiles of the conditional distributions. The article is divided as it follows: Section 2 contains a review on typical findings in international research including evidence on social background factors predicting achievement especially in Italy. Section 3 describes the QR method, while Section 4 details the data, which is analyzed in Section 5. Section 6 reports conclusions and implications.

Review

Education affects the individuals' life as it shapes **their** capabilities, values, aspirations and desires. It allows individuals to think, feel and act in different ways, enables new ways of organizing and supporting social action that depend on numeracy and literacy, technologies of communication and abstract thinking skills and, at the same time, societies use educational access and attainment as a primary mechanism to sort and select future generations into different social and economic roles (Lewin, 2007). These educational experiences in achievement have implications for social policies in more advanced economies, where active social policies focus on integrating people into the labor market through education that ensures disadvantaged regions and students are not left behind in the quest for success (OECD, 2012). The key role of education in the social and economic policies highlights the need to monitor student achievement.

Since the 1990s, major international student achievement surveys have sought to quantify student performances in different fields of knowledge by comparing different educational systems worldwide. These include: (i) the International Adult Literacy Survey (IALS) carried out in three editions (1994, 1996 and 1998) by the Organization for

Economic Cooperation and Development (OECD) and Statistics Canada; (ii) the Trends in Maths and Science Study (TIMSS) (1995, 1999, 2003, 2007, 2008 and 2011) and the Progress in International Reading Literacy Study (PIRLS) (2001, 2006 and 2011), both conducted by the International Association for the Evaluation of Educational Achievement (IEA).

Since the year 2000, the OECD carried out the Program for International Student Assessment (PISA). It is administered every three years to provide comparisons of students' achievement among the participating countries; it has completed in 2012 its fourth round. To date, PISA 2009 offers **together with the IEA-PIRLS survey** the most comprehensive and rigorous international measurement of student reading skills.

PISA collects information on all three areas of competencies (mathematics, reading and science) in terms of test scores, with a focus on one of the three competencies every three years (in PISA 2009 the focus is on reading literacy for which sub-scores have been provided), unlike the IEA surveys that collect information in reading literacy (PIRLS) and mathematics and science literacy (TIMSS) separately. Both PIRLS and PISA are sample surveys, but their sampling design is quite different.The PIRLS survey is administered to a sample of students formed by one or two whole classes in each school selected, while the PISA test is administered to a group of students who attend the school sampled without taking into account the class group. The choice made by the PIRLS survey allows to focus more attention on the class and, therefore, on the relationship between the taught and learned curriculum. The PISA survey, focuses more attention on the effective level of achievement gained by the students regardless the content of teaching received, as students are randomly selected for each school selected; thereby the classroom effect is got over. Furthermore, although vocational students are not exposed in depth to the kind of formal knowledge taught in academic programs and to a lesser extent in technical and vocational tracks, PISA considers student knowledge in relation to students ability to

reflect on their knowledge and experience and to apply them to real world issues. Both surveys cover a wide range of domains pertaining the assessment of the student reading literacy. To keep to a minimum the assessment burden on each student and in order to avoid that the scaling of achievement would be influenced by the "booklet effect" each student is asked to cope with only part of the assessment following a systematic booklet assembly and rotation procedure. Furthermore, TIMSS assesses mathematics and science achievement at fourth and eighth grade levels, the target population for PIRLS are students enrolled at the fourth grade, while PISA targets 15-year-old students, thus allowing comparisons of competency levels useful in adult life and labor market participation both within and across national education system.

Recently, Hanushek et Woessman (2010) review the economic literature on international differences in educational achievement reporting the main findings of contributions that have analyzed TIMSS and PISA surveys over decades. Because PISA survey cover the most educational systems, its results are reported frequently across a wide range of educational topics. For example, Brunello et Rocco (2013) use aggregate PISA data for 19 countries over the period 2000-2009 to study whether a higher share of immigrant pupils affects the school performance of natives; Bulut, Delen, et Kaya (2012), analyzing PISA 2009 data for Turkish students through a structural equation modeling, focus on the relationship between reading scores and the use of technology for reading; Fonseca, Valente, et Conboy (2011) compare Portuguese students' performance in PISA 2006 scientific literacy with those of some others OECD countries; Martins et Veiga (2010) using PISA 2003 data evaluate socioeconomic-related inequalities in students' math achievement in 15 EU countries, investigate their main causes and analyse differences between countries; Beese et Liang (2010) use the PISA 2006 data to investigate how school resources indicators (such as teacher qualifications, school facilities, and school type) as well as student level variables (such as socioeconomic status and family resources) affect

the literacy in science in United States, Canada and Finland; Bybee, McCrae, et Laurie (2009) highlight the importance of PISA 2006 about information for the science education community; Suggate (2009) analyses the relationship between reading achievement and the early reading instruction controlling for social and economic differences using PISA 2006.

Worldwide, a broad literature in the field of education and economics is aimed at investigating the determinants of student achievement. These studies typically search for empirical support on educational policies to improve students' achievement and learning (Sahlberg, 2007). As schooling is a major means through which young people accumulate human capital, it is important to understand the relative importance of achievement determinants - student background, household possessions, school quality, community affluence, etc. This is especially important since social inequalities can reproduce themselves through educational inequalities (Lauer, 2003). For example, when in some developing nations or regions school characteristics are found to be more important than the home in promoting academic achievement, policies to promote learning can be applied differently than when in some advanced countries or regions, the home is found to have a greater impact on learning than schools and teachers (Heyneman & Loxley, 1983 ; Hanushek, 1995 ; Baker, Goesling, & LeTendre, 2002).

Since the publication of the Coleman Report (Coleman & John Hopkins University, 1966) family background has been found to play a prominent role in improving student performance. The authors of this contribution conducted a comprehensive study on background characteristics of schools and students able to affect the outcomes of education. Following their main findings other Authors have proved the relationship between students' achievement and family background (Agasisti & Vittadini, 2012 ; Bratti, Checchi, & Filippin, 2007 ; Hanushek & Luque, 2003 ; Lauer, 2003 ; Korupp, Ganzeboom, & Van Der Lippe, 2002), highlighting a considerable persistence

across generations in educational achievement (Fertig & Schmidt, 2002 ; Currie & Duncan, 1999 ; Miller, Mulvey, & Martin, 1999). A major goal of the theoretical paradigms aimed to reform school systems for more than a decade has been to narrow achievement gaps between students from low-income families and their more advantaged peers. The many research and policy studies cited above have explored possible explanations for achievement gaps and the ways to narrow them, concluding that various dimensions of socio-economic status (SES) (household income and wealth, parents educational attainment, family structure, home cultural possessions, and so on) account for some of the main facets of the achievement gaps (Duncan & Magnuson, 2005). Students from disadvantaged families tend to have fewer opportunities at home to foster competencies, not to be encouraged to find interest or see value in learning, parents who not promote autonomous learning, or develop social relationships that support the achievement. This is especially true for reading achievement as family background variables are more powerful predictors of reading than math or science since parents are able to encourage children to read from an early age, but are not so good at helping with other competencies (math or science). Hence, family background create a disadvantage for families who lack some of the resources needed to support students' achievement. Heckman (2011) suggests that the solution to this problem is not to blame parents for their childrens skill development, but rather to provide disadvantaged families with the resources to prevent these gaps. Considering students' characteristics some of them as gender, the age of student and the ICT (Information and Communication Technology) availability at home and at school have been discussed in literature. Gender plays an important role in students' reading choices. A reading achievement

gap in favor of female students over male is significant in all 139 studies in which such comparisons were made (Leitz, 2006); female students consistently read more than male students from primary education to higher education (Gambell & Hunter, 2000) as female students devote more time to narrative fiction than male students (Hughes-Hassell & Rodge, 2000 ; Coles & Hall, 2002). Regarding the age of students, some studies have found that if students were enrolled a year before they achieve better than other students, and if students were not admitted to the next grade during their past career its achievement worsen (Agasisti & Vittadini, 2012). Finally, the effect of ICT on students' achievement has received attention during the last decade and the evidence is not conclusive (Bulut et al., 2012 ; Judge, 2005 ; Fuchs & Wößmann, 2005 ; Angrist & Lavy, 2002). In particular it is interesting to explore whether there is any difference in the effect of having access to ICT at home or at school, indeed some studies have proved that it is not the same (Gamboa & Garcia-Suaza, 2011).

In recent years, research on Italian school achievement has shown a growing interest in analysing the determinants of Italian students' performances. Information about students' achievement allow to monitor the effect of the Italian educational policies and the allocation of public budgets. The first and most extensive contributions have been collected in a book from Bratti et al. (2007) with the aim to provide a wide set of explanations of achievement differentials among Italian students. This research, carried out at student level, uses PISA 2003 data. Briefly, the main results of recent contributions highlight that factors affecting students' achievement are principally related to (i) the socio-economic status of family (higher the status is, the better the achievement is), (ii) the region of residence (students in Northern regions perform much better than those in Center-Southern area), and (iii) the type of school attended (academic and technical

schools perform better than vocational) (Quintano, Castellano, & Longobardi, 2012 ; Agasisti & Vittadini, 2012).

Italy is a particularly interesting case of study since national and international surveys on students' achievement emphasize the serious gap that on average separate Italy and the rest of the OECD countries (OECD, 2012). Despite some Italian regions perform well above the OECD average, there is a very wide gap between different geographical areas within the country (INVALSI, 2011). Geographical differences together with a separated tracking system (academic from vocational and technical school programs) are commonly identified as the most important factors accountable for the observed Italian inequalities. Previous studies on the Italian students achievement have demonstrated that there are relevant differences across the different areas of the country, with students in the Central part of Italy performing worse than those in the North and better than those in the South area (Quintano et al., 2012 ; Agasisti & Vittadini, 2012), and overall the Northern regions perform better than the others. The high explanatory power associated to Italian regions induced a further reflection on the role played by the geographical factors, and especially by the different socio-economic characteristics of the regions themselves. Indeed, Italy is one of the OECD countries with the higher level of regional socio-economic disparities. These differences in the regional socio-economic conditions seem to have an effect on the students' achievement. Indeed, differences in educational results will lead to increasing differences in economic development, which in turn will reproduce gaps in educational performance. Furthermore, Northern regions have a better functioning scholastic system (and in general they have better public services) compared to the Southern ones (Agasisti & Cordero-Ferrera,

2013).

Another feature of the Italian educational system is its stratification across a general (academic) training track (the so called "Liceo") and a non-academic training track (technical and vocational). The type of school attended comes out to be a very important factor as students in academic programs perform better than those in technical or vocational (INVALSI, 2011).

The advantage of academic high schools does not necessarily reflect a causal effect, since it might account for self-selection of students in this type of schools (Bratti et al., 2007). The main problem is that the choice of school program is clearly influenced by family background, with children of better educated parents showing a higher probability to be enrolled in the academic oriented high-school track (Brunello & Checchi, 2007 ; Checchi & Flabbi, 2007 ; Flabbi, 2001). Finally, another key aspect of learning environments in a school context is the disciplinary climate in class which has emerged as one of the most important factors related to student achievement (Hattie, 2009 ; Scheerens, 2005 ; Wang, Heartel, & Walberg, 1993).

Method

It is generally accepted that standard linear regression techniques measure the relationship between a set of explanatory variables and the average value of a quantitative response is modeled with the conditional mean function $E(y|x)$. Consequently, the focus is on average when the relationship between dependent and predictor variable is considered. However, researchers might be interested in considering the relationship at other points of the conditional distribution of the response y. Quantile Regression (QR) allows

for inspection of multiple points when considering the relationship between explanatory and response variables at the specific quantile q of the y distribution.

QR models have been proposed by Koenker et Basset (1978) as a "robust" technique alternative to OLS regression when the error terms are not normally distributed. The technique has been applied in many empirical fields and with different types of data over the years (Chamberlain, 1994 ; Eide & Showalter, 1998 ; Fitzenberger, 1999 ; Levin, 2001).

Contrary to the usual OLS mean regression model, the QR model is the most available and often used alternative in regression analysis.

QR is widely regarded as a robust estimation technique which is substantially less sensitive to outliers than the OLS (Gould & Rogers, 1994). The advantage of the QR is the possibility to estimate the whole of quantiles of the conditional distribution of the response variable. In other words, the estimation of the conditional mean is replaced with the estimate of 99 conditional quantiles, that is QR allows to consider predictor effects in any chosen quantile and does not rely on any parametric specification of the conditional distributions.

Apart from estimating variables effect at different parts of the distribution, QR has several advantages compared to OLS: it gives less weight to outliers in the response variable than OLS; the estimation is a more robust method since it allows the marginal effects of explanatory variables to differ across the quantiles of the response variable. When error-terms are not normal, quantile regression estimators may be more efficient than OLS estimators. Finally, the semi-parametric nature of QR tempers the restrictions on the parameters to be held constant across the entire distribution of the dependent variable.

In modeling students' math achievement with earlier analised 2003 OECD-PISA data, some of the explanatory variables that will be analysed in

the following have been already considered (Martins & Veiga, 2010). In that research, the standard linear regression models fitted (in a Multilevel setting) for Italy display a significant effect of covariates on students' achievement such as parents' education, number of books available at home and possessions of goods. Nevertheless, that significant effect has been assessed only at the mean level of the response variable. In that research we will show how the effect of the above mentioned covariates changes over student reading scores along different point of the response distribution. The basic QR model specifies the conditional quantile as a linear function of explanatory variables. For the τ-th quantile, a common way to write the model is:

$$Q_y(\tau|X = x) = x_i'\beta(\tau)$$

with

$$\hat{\beta}(\tau) = \arg\min_{\beta\in\Re^+} \sum \rho_\tau(y_i - x_i'\beta)$$

The τ-th regression quantile $(0 < \tau < 1)$ of y is the solution to the minimization of the sum of absolute deviations residuals:

$$\min_{\beta^\tau} \sum_{\sigma_k<0} \tau|y_k - X_k\beta^\tau| + \sum_{\sigma_k>0} (1-\tau)|y_k - X_k\beta^\tau| \tag{1}$$

where τ determines the conditional quantile of interest and all positive residuals receive a weight of τ, while the negative ones receive a weight of $(\tau - 1)$. Hence, any one of the components of the QR coefficients $\beta(\tau)$ provides an estimate of the marginal effect of the associated explanatory variable on the response one for the τ-th quantile, controlling for the remaining variables.

The most important feature of this method is that the marginal effects of the predictors, given by $\beta(\tau)$, may differ over quantiles. If the effect of each variable is homogenous across the conditional distributions, we would expect the slope of the

estimated coefficients at the quantiles to remain constant. In fact, QR allows to evaluate the different behavior of predictors at different intensity of the response variable (Koenker & Hallock, 2001).

QR coefficients are typically computed by expressing equation (1) as a linear program where a simplex method (Koenker & d'Orey, 1987 ; Ng, 1996) or interior-point algorithm (Portnoy & Koenker, 1997 ; Koenker & Ng, 2005) can be applied. Other available approaches to estimate the quantile regression coefficients are the Majorise-Minimise (MM) algorithm of Hunter et Lange (2000), the reversible jump Markov Chain Monte Carlo algorithm of Yu (2002), and the smoothing algorithm of Chen (2007). QR software is readily available in most modern statistical platforms (STATA, EViews, SAS, R, etc.). This analysis employed the STATA Program. **The complex sampling design of the PISA survey employs student weight using the Balanced Repeated Replicates (BRR) Method to estimate standard errors.** In analyzing PISA data, the ordinary methods to calculate standard errors are biased for two reasons. First, there is intra-cluster correlations among schools. To correct for the effect of intra-cluster correlations, PISA survey design provides a series of weights for Balanced Repeated Replicates (BRR), which is like bootstrapping except that the re-samples are pre-defined. For that reason, we use BRR weights provided in the dataset to remove differences in the selection probabilities of students by removing clustering by schools that could bias standard error estimates (i.e. BRR weights are needed to **take into account for clustering and stratification, thus to** compute unbiased-standard error estimates). Second, the PISA dataset provides not only a unique measure of achievement but five Plausible Values (PV) estimated by considering a plausible range of distributions across the test score. Consequently, the standard error calculations have to take into account the sampling variance in the estimate of the full range of the response variable. PV estimates are a recent innovation in item response theory and are increasingly used in surveys on

student achievement to estimate population parameters, such as mean performance or population regression coefficients, which may be superior to OLS conditional means (Monseur & Adams, 2009). Accordingly, we consider the average coefficients deriving from using all the five PV in reading for each student. Furthermore we use the "final student weight" to obtain estimates representative of the population. Finally, to take into account for sampling design in order to get unbiased standard errors we use the 80 replicate weights available in the PISA 2009 student file (for a detailed description see OECD (2012)). We use STATA software as the command "pv" (Macdonald, 2011) allows to estimate statistics when there are multiple estimates of the dependent variable, by calculating the statistics for each estimate of the response variable and then averaging.

Data

Following the detailed assessment of each PISA's three main subjects – reading, mathematics and science – in 2000, 2003 and 2006, the 2009 PISA survey marks the beginning of a new round with a return to a focus on reading (nevertheless, students's knowledge in mathematics and science has been also assessed).

Reading literacy is defined as it follows: "[...] understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society [...]" (OECD, 2012). PISA reading assessment focuses in the following sub-areas i) individual engagement in reading; ii) learning time; iii) students' views on their test language lessons; iv) access to and use of libraries; v) students' strategies in reading and understanding texts.

In the following analysis we will consider the overall reading achievement assessed for each student. However, in the exploratory analysis performed on each of sub-test individually, we found no evidence of significant differential

effects from region, school type or other covariates. Therefore, the total reading score served as the dependable variable for the analysis.

The PISA target population is that of students aged between 15.25 and 16.25 years at the time of the survey and who have completed a minimum of 6 years of formal education regardless of the type of institution they are enrolled in. The choice of 15-16 years old students marks, for many countries, the transition from a basic education to a more advanced education or professional training. Detailed information on sampling design and procedures are available in a series of thematic and technical reports at PISA-OECD website.

Students participating in PISA program fill-in a very comprehensive questionnaire and information on school characteristics. The comprehensiveness of the PISA dataset allows us to consider for the analysis a number of variables that are not usually available in most other studies. The 2009 PISA survey has been carried out in 65 countries. For the present analysis only the Italian data set was examined (database release version December 2011).

The total sample size for Italy consisted of 30,905 students and 1,097 schools. It is worth noting that in order to perform regional comparisons, an oversampling scheme was adopted for Italy. The main results of the Italian PISA 2009 survey were published in 2011. **Considering the ISCED 1997 classification (that in use at the time of the survey) we found that (94.7%) were enrolled in a ISCED 3A/B school program that allows access to tertiary education, and those enrolled in a ISCED 2C program that allows direct access to the labor market upon graduation, or terminates school upon completion at age 13-14 account for 4.8% and 0.4% respectively.**

We consider as response variable the student's reading scores and, as explanatory variables, following the main findings of the literature: i) some students' characteristics, ii)

some students' family background (related to the economic, social and cultural status and home educational resources), iii) the familiarity with ICT (Information and Communication Technology), iv) two control variables to take into account for regional differences and for the school program attended.

Concerning student characteristics, we consider the gender (`SEX`), whether failed a year in school (`RIPISCED`), and whether student has been enrolled a year before (`ANTICIP`). For purposes of this study, **family background** variables include variables related to the highest level of parents occupation (`HISEI`); highest educational level of parents in years (`PARED`); an indicator of family wealth (`WEALTH`); a measure of family cultural possessions (`CULTPOSS`); an indicator of availability at home of educational resources (`HEDRES`); and the number of books at home (`HOMEBOOK`) categorized into a six-level. These variables represent multidimensional facets of family background on the students' achievement. **The effect of ICT on reading achievement has been considered using two variables related to ICT at home (`ICTHOME` and `ICTRES`), and one variable related to ICT at school (`ICTSCH`).**

The variables `WEALTH`, `CULTPOSS`, `HEDRES`, `ICTHOME`, `ICTRES` and `ICTSCH` are indexes available in the PISA database defined with a model-based scaling procedure belonging to the family of Item Response Theory (IRT) and described in details in OECD (2012).

Table 1 approximately here

Moreover, to take into account for geographical differences we include a regional dummy, `REGION` (one for each Italian region). In order to include the regional dummies we match the PISA dataset with the one provided by the INVALSI, the Italian National Evaluation Committee that is a a OECD partner in administering PISA surveys in Italy.

Finally, we control for the school program attended (`PROGRAM`). In Italy, the main subdivision found in the upper secondary school system consist of (i) Liceo **(44% of the sample units)** (ii) Technical **(31% of the sample)**, and (iii) Vocational **(31% of the**

sample). Liceo **is a school** oriented towards the classics and sciences that aims to train students for higher education programs (at ISCED levels 4 or 5). Technical programs are oriented toward practical subjects, such as business administration, computer science, chemistry, nautical disciplines, and aeronautics. Vocational programs specifically focus on practical subjects. Both technical and vocational enable students to seek employment upon graduation. The disciplinary climate at school (`DISCLIMA`), available as an indicator variable (built up by IRT modelling) in the PISA dataset, has been also considered in the analysis.

Table 1 defines and provides descriptive statistics for the variables used in the study.

Analysis

In this Section, QR analysis is used to estimate the effects on student reading of predictor variables described in Section 4 to verify whether changes occur along the reading score distribution. Differences across quantiles of the conditional distribution of reading indicate heterogeneous effects of explanatory variables. Therefore, we estimate a linear regression function at different quantiles ($\tau = 0.1, \ldots, 0.9$) and we examine whether there is homogeneity in the effect of each explanatory variable on students' reading score **comparing QR results with the OLS estimates.**

Table 2 presents the results obtained with OLS and QR from the 10-th to 90-th percentile of the distribution of the students' reading score. **The first column reports the selected covariates, in columns 2-7 the β coefficients for the three main quantiles 0.1, 0.5, 0.9 have been reported with their level of significance, finally the last two columns report the OLS coefficients and the related level of significance. The β coefficients in OLS measure the students gap on**

average, while the quantile regression coefficients measure the gap in the lowest quantiles (lower-performing students), in the median quantile (median performing) and in the upper quantiles (higher-performing students) distributions of students' achievement scores. We shall note that if the β coefficients of the quantile regression are significant and different from the OLS β regression then the use of quantile regression is more efficient than the regression on the average. In Table 2 for a number of covariates the effect on reading achievement changes on average and on different points of the distribution, proving that quantile β coefficients provide a richer description of data, allowing us to consider the impact of a covariate on the entire distribution of reading achievement and not merely on its conditional mean.

Table 2 approximately here

Comparing the results of QR with of OLS, we can note that student and family background have a statistically significant impact on the reading achievement **in both models and in relatively the same proportions**.

Female students perform better than males: quantile coefficients indicate that the `SEX` effect decreases as estimates move from the 0.1 to 0.9 quantiles (changing from $\hat{\beta}_{\tau=0.1} = -26.21$ to $\hat{\beta}_{\tau=0.9} = -17.41$) indicating a more considerable negative effect for boys that are lower reading performers than for the higher performers. The effect of the variable related to whether the student was enrolled one year before (at 5 instead of 6 years of age – `ANTICIP`) is more important for the lower-performing students than for the higher ones ($\hat{\beta}_{\tau=0.1} = -9.84$ and $\hat{\beta}_{\tau=0.9} = -4.81$). If the student has never failed a year at school, then reading achievement is higher than for those who failed and the effect of (`RIPISCED`) does not change across the quantiles. **These above results are in line with the main empirical evidences in the literature; although QR coefficients do suggest an added result that the gender gap decreases at the upper quantiles**

of the distribution. Especially noteworthy is the increasing value for male students at the upper tail of the response distribution (the 75-th percentile).

The coefficients related to the family background are significant and highlight some difference across reading score distribution. The coefficients of the highest occupational status of parents (`HISEI`) and the highest educational level of parents in years of education according to ISCED classification (`PARED`) are positive. This suggests that reading performances is better for students with graduate parents and higher occupational status. This result confirms the substantial intergenerational dependence of educational achievement (Checchi, Fiorio, & Leonardi, 2008, 2007 ; Becker & Tomes, 1986). Nonetheless, the effect of `PARED` is different at the two extreme ends of the distribution ($\hat{\beta}_{\tau=0.1} = 19.21$; $\hat{\beta}_{\tau=0.9} = 11.62$), that is the quantile regression results indicate that the effect exhibits a decreasing trend from the 0.1 to 0.9 quantiles.

The coefficients related to cultural possessions (`CULTPOSS`), home educational resources (`HEDRES`) and books in the home (`HOMBOOK`) have positive effects on reading score, whilst the index of family wealth (`WEALTH`) shows a negative sign. These results highlight the importance of home inputs directly related to learning rather than possession of global indicators of family wealth including cellular phones, televisions, computers, cars, and so on.

This finding suggests that the availability of educational resources at home (a desk and a quiet place to study, a computer that students can use for schoolwork, educational software, books to help with students' school work, technical reference books and a dictionary) are more important for low performing readers than for students with adequate resources found in the home. It appears, low performers with adequate resources study more effectively than when these resources. On the contrary, the `HOMEBOOK` variable suggests that the availability of books in the home is more important for high performing readers than for lower scoring ones ($\hat{\beta}_{\tau=0.1} = 7.01$; $\hat{\beta}_{\tau=0.9} = 8.80$). **Likely,**

lower-performing students need basic educational resources, whilst higher-performing students benefit mostly of other educational support as a number of books at home.

These empirical findings prove poor readers are more sensitive to family background

The effect of computer resources depends on the place where students have ICT access: home or school. The ICT availability at school does not affect reading achievement, whilst the ICT at home have a positive effect on achievement, both in terms of availability and learning resources, and with differential effects on lower and higher performers. In fact, estimation results for variables related to the ICT availability at school (`ICTSCH`) show a non significant effect on reading achievement (likely because students can find additional support for educational activities at home), whilst the ICT availability at home (`ICTHOME`) and the ICT resources for learning (`ICTRES`) have a positive effect. However, the ICT availability at home has a stronger effect for the lower-performing students ($\hat{\beta}_{\tau=0.1} = 4.27$; $\hat{\beta}_{\tau=0.9} = 0.74$), whilst the ICT home resources for learning has a stronger effect for the higher-performing students ($\hat{\beta}_{\tau=0.1} = 3.45$; $\hat{\beta}_{\tau=0.9} = 4.78$).

Regarding school variable, better disciplinary climate (`DISCLIMA`) is positively correlated with reading performance and the effect is greater for lower-performing students than for the higher performers. The type of school attended (`PROGRAM`) has a strong impact on reading performance since students from **academic programs (or** "Liceo") perform better than technical-vocational students. Quantile estimates show that the higher performing students enrolled in vocational programs achieve better than the lower ($\hat{\beta}_{\tau=0.1} = -97.75$; $\hat{\beta}_{\tau=0.9} = -84.55$ (see Table 2). **Although students in academic tracks perform better than all other curriculum programs, still the high performing students in vocational programs have a narrower range of**

difference from high academic track performers. This finding is likely due to the narrower range found between high and low performing students in vocational track given selection into this track of students who do not qualify for more academically demanding courses. Table 3 shows the predicted values for the student reading scores across different school programs. Any higher performing student, the school program has less influence than for low performers with the academic program reporting that the difference respect to the Liceo is decreasing moving from 63.55 to 61.50 for Technical programs, and from 127.14 to 110.89 for Vocational.

The geographical variable (`REGION`) shows the presence of noteworthy territorial effects in reading performances (see Table 2 and Figure **3**). Students in Northern regions have better reading scores, indeed the average scores differ strongly among the Northern and the Southern regions and these considerable differences originate a wide North-South literacy-divide (Quintano et al., 2012). Moving from the South to the North (and assuming as a baseline the region where the reading score is the lowest, i.e. Calabria), the effect of `REGION` differs significantly along the distribution.

Table **3** approximately here

In addition, Figures **1-3** in Appendix show the estimated coefficients for each covariate plotted as a function of the quantiles (x-axis). In the case of quantile regression results the estimated coefficients change by varying the quantiles of the response distribution; this allows us to provide a more specific interpretation of the influence of the explanatory variables on the reading scores; in particular, for those predictors for which the value of estimated coefficients changes significantly with the change of quantiles (the line representing the coefficients in Figures 1-3 is not parallel to the x-axis, but has a negative/positive slope). The vertical distance of a solid dot from the horizontal axis can be interpreted as the effect of a one-unit change of one explanatory variable on the reading

scores for the chosen τ-th quantile, controlling for the covariates. The results of the OLS regression and the median regression do not differ very much.

The gray surface represent the confidence intervals, a horizontal continuous line at 0 and a horizontal dashed line is in correspondence of the estimated OLS coefficient value. Hence, a confidence band corresponding to the estimate coefficient of a specific predictor that does not contain the horizontal x-axis represents a quantile regression coefficient that is statistically significant at the 5% level for that particular τ. In this way, we can better evaluate the results obtained fitting QR. Figure **1** shows that the positive effect of `PARED` on score for lower-performing students is stronger than for higher-performing ones; on the contrary for the `HISEI` variable. Furthermore it shows shows the positive impact of `HEDRES` on reading score for lower-performing students, which is stronger than for the higher-performers. Figure 2 shows that a better disciplinary climate (DISCLIMA) is more important for lower-performing students than for the higher ones. Futhermore, quantile estimates show that the higher performing students enrolled in Vocational programs achieve better than the lower. In Figure 3 quantile estimates for regional variable have been plotted; the effect of REGION differs significantly along the reading score distribution showing a more balanced achievement in almost all Northern regions and a very unbalanced achievement path in almost all Southern regions.

Figure **1** in Appendix approximately here

Figure **2** in Appendix approximately here

Figure **3** in Appendix approximately here

The well known Italian North-South divide characterizes many economic and social disparities (occupation, quality of life, criminality, and so on) also plays a role on reading students' achievement pointing out the Italian North-South literacy-divide. Historically, Northern regions (Bolzano, Emilia

Romagna, Liguria, Lombardy, Friuli, Piedmont, Aosta Valley, Veneto, Trento), perform better than those in Center-South. The regional effect is strongest for higher-performing students in Southern regions as for example Campania, Apulia, Sardinia, Sicily and Umbria, where the effect increases over quantiles. Furthermore the inverse U-shaped form of the distribution in Northern regions highlights the effect, which is approximatively the same for good and poor readers (that is a different slope before and after the median value, increasing and decreasing, respectively). Consequently, QR reveals whether the reading achievement pattern is balanced between lower and higher-performing students within regions. QR results suggest that in Italy there are *different regional achievement patters*: i) balanced achievement as in almost all Northern regions and a few Southern regions (Abruzzo and Basilicata) where the same effect exists between the two ends of the reading score range; and ii) a second pattern consisting of unbalanced achievement patterns found in almost Southern regions and few Northern regions (Friuli and Liguria), where the effect increases or decreases similarly across high and low performers.

In order to better evaluate the regional differences on the reading score distribution Table **4** reports the predicted values of the scores for three quantiles 0.1, 0.5, and 0.9. In comparison to the baseline region (Calabria) it is interesting to note that better performing regions as Bolzano, Tuscany or Emilia-Romagna have higher scores at the median quantile although in the two extreme quantiles (0.1 and 0.9) the differences are the same for Calabria, Lazio, Lombardy, while in Piedmont the distance from the baseline in the 90-th quantile is smaller than in the 10-th. Finally, for some Southern regions, such as Sardinia, Sicily and Apulia, the influence increases so that the regional effect is stronger for higher-performing students **that is higher-performing students in these region area are able to increase the reading performance despite the region where**

they live.

Table **4** approximately here

In Italy, the **North-South territorial dualism** also affects student achievement. Historically, the quality of education in Italy has been geographically determined: two areas with different opportunities and levels of literacy. This analysis highlights different regional achievement patterns. Northern regions perform better than Southern regions. By means of QR, a further distinction shows up as achievement patterns that are either inverted U-shaped or symmetric between lower and higher performing students (in the North), or a pattern that is asymmetric across the two reading performance groups (in the South). For example, by comparing Friuli and Sicily the regional effect is opposite: decreasing for Friuli and increasing for Sicily. These large regional gaps need for policy intervention directly, in terms of educational reforms, public budget allocation and/or indirectly acting on variables that affect the different part of achievement distribution. To reduce differences among high and low reading performers it is necessary to address specific regional policies to reduce first the socio-economic gap. In each region we compared the QR predicted values with those of OLS. **Table 5** reports the ratio between the predicted value of achievement for $\tau = 0.1$ and $\tau = 0.9$ to OLS and the Relative Variation About OLS (RVAO) index. Some Northern regions show a very balanced achievement path (Veneto, Trento, Bolzano and Liguria), as well as some of the South (Apulia and Basilicata). Whilst regions as Friuli or Emilia-Romagna, with good mean levels of achievement, show unbalanced distributions.

Table **5** approximately here

Conclusions and Implications

This paper provides an analysis of Italian students' reading achievement determinants by using the last PISA 2009 survey. **Applying quantile regressions**

(QR) the paper assessed the impact of selected covariates at more than one level of the reading achievement distribution, and found effects for student scores at the high and low end of the spectrum. For comparison, the empirical results from ordinary least squares (OLS) regression were also estimated. Both OLS and QR models have shown that individual and family background variables, school program, and region of residence affect reading achievement. However, QR adds detail not available from OLS because it analyses the effect of explanatory variables across different points along the distribution. Although OLS coefficient estimates are in line with those of QR, the former do not capture relevant information found at different points of the reading achievement distribution. Consequently, this paper confirms that the QR approach is more robust because it measures the effect of predictors at various quantiles, thus capturing relevant differences from the median measure of the achievement distribution. This complex pattern can be seen in gender comparison where female low performers scores noticeably higher than males at the lower end of reading performances. Generally, it is know that girls do better than boys at reading, but knowing that this effect is more pronounced among low performing girls is new information that would normally not be observed in OLS findings

Student social background plays a key role in influencing reading achievement. To proxy the family background six covariates available from the 2009 PISA data were used: parental occupational, education, wealth, cultural possessions, educational resources, and books in home. On average, all these covariates had positive effects on reading although the QR results showed the highest family background effect on reading for lower-performing students. This is true especially in Southern regions where the socio-economic context is

particularly disadvantaged and the family background must overcome the unfavorable conditions. The index of family wealth shows a negative effect on reading achievement; on the other hand the importance of possession of goods directly related with learning (books and other educational resources) has a positive effect that is quite differentiated. While the possession of educational resources (a desk and a quiet place to study, a computer available for schoolwork, technical references books and a dictionary) is very important for lower-performing students, the number of books at home is more important for higher-performing ones. Estimates for variables related to the ICT availability at home, along with ICT home resources used in learning both showed statistically positive effects with ICT resource availability at home having a stronger effect for the lower-performing students. ICT home resources had a positive effect on reading mostly for higher-performing students. ICT resources at schools did not have any effects for either good and poor readers. The type of school attended was an important factor since students from academic tracks performed better than the others most likely due to selection based on merit. However, quantile estimates showed that higher-performing students (as opposed to lower performing) enrolled in vocational programs display a smaller gap with their counterpart performing academic students. Lastly, the regional variable has shown the existence of a wide gap between Italian geographical areas. Students in the Northern area perform better than many Center-Southern areas, and in line with the rest of Europe, while students in the South perform worse and fairly below the European average. QR results suggest that the effect of regional dummies varies across the distribution of reading achievement scores.

The above results suggest that the relationships between achievement

and its predictors differ across the conditional distribution of reading performance. This information would not have been forthcoming using standard OLS techniques as it was using QR, which offers more insight into the impact of gender, school type, region, and social background resource on learning to read. In this respect, QR is a useful tool to help policymakers improve the educational system, by targeting programs directly to lower-performing or higher-performing students or both (i.e., when the effects are homogenous in the entire achievement distribution). Such information could be helpful in tuning the institutional policies toward the less performing students as QR should be an important research tool to better understand home, school and regional effects on achievement. In recent years the central government and local education authorities have put money into programs to strengthen ICT availability for students. According to these results, it seems that the effect of ICT home resources for learning on reading is well differentiated for low and high performing students. Information provided by QR helps to define the best strategy to pursue specific outcomes, that is paying attention on ICT resources that improve the achievement of lower readers (i.e., ICT home resources affect reading performances and not ICT availability at school. At the same time, **the effect of school related covariates is also noteworthy differentiated considering the different quantiles of the reading achievement distribution.**

With QR it is possible to consider the lower or the upper quantiles and uncover at each quantile the most appropriate policy to reach the chosen target; i.e., the different quantile estimates help on selecting the most effective policy at a given quantile and that cannot be uncovered otherwise, suggesting what happens in the tails of the achievement distribution. For some predictors information provided by quantile coefficients are merely cognitive (as gender

for example), but for other predictors this could be a guide for policy makers (i.e. to reduce the regional gap, or for the school program gap).

If geographical differences and the separation between academic and technical/vocational education are important factors predicting inequality patterns, then policies to enhance fairness have been unsurprisingly focused on these two problems. Although a number of reforms have attempted to remove the impact of tracking, *de facto* these attempts were followed by a number of counter-reforms supporting a clear separation of school programs as noted by Benfratello et Turati (2013). In the context of research in education, the South of Italy is the weakest part of the country based economic and social indicators. This disadvantage gets translated into the the school system as international surveys on student competencies have long detected large gap between regions. Inequities in the Italian education system penalize the lower-performing students who live in disadvantaged areas. Finally, as QR has confirmed the existence of differential effects of predictors across the reading performance spectrum, **a further step of this kind of analysis should be** to perform a Multilevel QR to take into account for the hierarchical structure of data **and to focus, also, on the heterogeneity observed among Italian schools**. A Multilevel QR model would allow us to control also for individual school variability and characteristics. This model, at the time of writing, is still in the implementation phase with **the available statistical packages** and, consequently, it requires a deeper analysis. Recently, Geraci et Bottai (2013) have provided a methodological contribution through an extension of Multilevel Linear Models to more complex dependence structures in the data, which are modeled by including multiple random effects in the linear conditional quantile functions.

Références

Agasisti, T., & Cordero-Ferrera, J. M. (2013). Educational disparities across regions: A multilevel analysis for italy and spain. *Journal of Policy Modeling*, *35*(1), 1079–1102.

Agasisti, T., & Vittadini, G. (2012). Regional Economic Disparities as Determinants of Students' Achievement in Italy. *Research in Applied Economics*, *4*(1), 33–54.

Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *Economic Journal*, *112*(482), 735–765.

Baker, D. P., Goesling, B., & LeTendre, G. K. (2002). Socioeconomic Status, School Quality, and National Economic Development: A Cross-National Analysis of the Heynemanan Loxley Effect on Mathematics and Science Achievement. *Comparative Education Review*, *46*, 291–312.

Becker, G. S., & Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labout Economics*, *4*(3), S1–S39.

Beese, J., & Liang, X. (2010). Do resources matter? PISA science achievement comparisons between students in the United States, Canda and Finland. *Improving Schools*, *13*, 266–279.

Benfratello, L., & Turati, G. (2013). *Tracking in the Tracks. Understanding Inequality Patterns in the Italian Public Schooling System.* CHILD Working Papers. No. 19.

Bratti, M., Checchi, D., & Filippin, A. (2007). Geographical differences in Italian students' mathematical competencies: evidence from PISA 2003. *Giornale degli Economisti e Annali di Economia*, *66*(3), 299–333.

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? new international evidence. *Economic Policy*, *52*(1), 781–861.

Brunello, G., & Rocco, L. (2013). The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores. *Economics of Education*

Review, *32*, 234–246.

Bulut, O., Delen, E., & Kaya, F. (2012). An SEM Model Based on PISA 2009 in Turkey: How Does the Use of Technology and Self-regulation Activities Predict Reading Scores? *Procedia - Social and Behavioral Sciences*, *64*, 564–573.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An Assessment of Scientific Literacy. *Journal of Research in Science Teaching*, *46*(8), 865–883.

Chamberlain, G. (1994). Quantile regression, censoring, and the structure of wages. In C. A. Sims (Ed.), *Advances in econometrics. sixth world congress. volume 1* (pp. 171–210). Cambridge : Cambridge University Press.

Checchi, D., Fiorio, C. V., & Leonardi, M. (2007, August). *Intergenerational Transmission of Educational Attainment in Germany: The Last Five Decades.* IZA Discussion Papers. No. 2985.

Checchi, D., Fiorio, C. V., & Leonardi, M. (2008). *Intergenerational Persistence in Educational Attainment in Italy.* IZA Discussion Papers. No. 3622.

Checchi, D., & Flabbi, L. (2007). *Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks.* IZA Discussion Papers. No. 2876.

Chen, C. (2007). A Finite Smoothing Algorithm for Quantile Regression. *Journal of Computational and Graphical Statistics*, *16*(1), 136–164.

Coleman, J. S., & John Hopkins University. (1966). *Equality of Educational Opportunity.* National Center for Educational Statistics.

Coles, M., & Hall, C. (2002). Gendered reading: learning from children's reading choices. *Journal of Research in Reading*, *25*(1), 96–108.

Cunningham, A. E., & Stanovich, K. E. (1998). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, *33*(1), 934–945.

Currie, J., & Duncan, T. (1999). The intergenerational transmission of "intelligence": Down the slippery slopes of the bell curve. *Industrial Relations*, *38*(1), 297–330.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children*, *15*(1), 35–54.

Eide, E., & Showalter, M. H. (1998). The effect of school quality on student performance: A quantile regression approach. *Economic Letters*, *58*(3), 345–350.

Fertig, M., & Schmidt, C. M. (2002). *The Role of Background Factors for Reading Literacy: Straight National Scores in the pisa 2000 Study.* IZA Discussion Papers. No. 545.

Fitzenberger, B. (1999). *Wages and Employment Across Skill Groups.* Berlin-Heidelberg : Physica-Verlag.

Flabbi, L. (2001). Secondary schooling decisions in italy. *Rivista di Politica Economica*, *91*(7/8), 85–114.

Fonseca, J., Valente, M. O., & Conboy, J. (2011). Student characteristics and PISA science performance: Portugal in cross-national comparison. *Procedia - Social and Behavioral Sciences*, *12*(322–329).

Fuchs, T., & Wößmann, L. (2005). *Computer and student learning: Bivariate and multivariate evidence on the availability and use of computers at home and at school.* CIfo Working paper No. 8.

Gambell, T., & Hunter, D. (2000). Surveying gender differences in Canadian school literacy. *Journal of Curriculum Studies*, *32*(5), 689–719.

Gamboa, L. F., & Garcia-Suaza, A. F. (2011). *Access to computer and academic achievement. Where is it best: At home or at school?* Center for Studies on Inequality and Development.

Geraci, M., & Bottai, M. (2013). Linear quantile mixed models. *Stat Comput*. (doi: 10.1007/s11222-013-9381-9)

Gould, W., & Rogers, W. H. (1994). Quantile regression as an alternative to robust regression. In *Proceedings of the Statistical Computing Section* (pp. 171–210). Alexandria, VA : American Statistical Association.

Hanushek, E. A. (1995). Interpreting Recent Research on Schooling in Developing Countries. *World Bank Research Observer*, *10*(2), 227–246.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Journal of Education Review*, *22*, 481–502.

Hanushek, E. A., & Woessman, L. (2010). *The Economics of International Differences in Educational Achievement.* IZA Discussion Papers. No. 4925.

Hattie, J. (2009). *Visible learning - a synthesis of over 800 meta-analyses relating to achievement.* London : Routledge.

Heckman, J. J. (2011). The American family in black and white: A post-racial strategy for improving skills to promote equality. *Daedalus, the Journal of the American Academy of Arts and Sciences*, *140*(1), 70–89.

Heyneman, S. P., & Loxley, W. A. (1983). The Effect of Primary-School Quality on Academic Achievement across Twenty-nine High- and Low-Income Countries. *The American Journal of Sociology*, *88*(6), 1162–1194.

Holloway, H. (1999). Improving the reading skills of adolescents. *Educational Leadership*, *57*(2), 80–82.

Hughes-Hassell, S., & Rodge, P. (2000). The leisure reading habits of urban adolescents. *Journal of Adolescent and Adult Literacy*, *51*(1), 22–33.

Hunter, D. R., & Lange, K. (2000). Quantile Regression via an MM Algorithm. *Journal of Computational and Graphical Statistics*, *9*(1), 60–77.

INVALSI. (2011). *Rapporto Nazionale PISA 2009* (Rapport technique). Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e Formazione : Auteur.

Judge, L. (2005). The impact of computer technology on academic achievement of young

african american children. *Journal of Research in Childhood Education*, *20*(2), 91–101.

Kern, L., Margaret, & Friedman, S., Howard. (2008). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology*, *30*(4), 419–430.

Koenker, R., & Basset, G. J. (1978). Regression Quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & d'Orey, V. (1987). Computing Regression Quantiles. *Applied Statistics*, *36*, 383–393.

Koenker, R., & Hallock, K. F. (2001). Quantile Regression. *Journal of Economic Perspectives*, *15*(4), 143–156.

Koenker, R., & Ng, P. (2005). Inequality Constrained Quantile Regression. *The Indian Journal of Statistics*, *67*(2), 418–440.

Korupp, S. E., Ganzeboom, H. B., & Van Der Lippe, T. (2002). Do Mothers Matter? A Comparison of Models of the Influence of Mothers' and Fathers Educational and Occupational Status on Children's Educational Attainment. *Quality & Quantity*, *36*, 17–42.

Lauer, C. (2003). Family background, cohort and education: a French-German comparison based on a multivariate ordered probit model of educational attainment. *Labour Economics*, *10*, 231–251.

Leitz, P. (2006). A meta-analysis of gender differences in reading achievement at the secondary school level. *Studies in Educational Evaluation*, *25*(1), 317–344.

Levin, J. (2001). Where the reductions count: A quantile regression analysis of effects of class size on scholastic achivement. *Empirical Economics*, *26*, 221–246.

Lewin, K. M. (2007). *Improving Access, Equity and Transitions in Education: Creating a Research Agenda.* Centre for International Education – University of Sussex.

Macdonald, K. (2011). *Pv: Stata module to perform estimation with plausible values.*

Consulté sur http://EconPapers.repec.org/RePEc:boc:bocode:s456951

Martins, L., & Veiga, P. (2010). Do inequalities in parents' education play an important role in PISA students' mathematics achievement test score disparities? *Economics of Education Review*, *29*, 1016–1033.

Miller, P., Mulvey, C., & Martin, N. (1999). Family characteristics and the returns to schooling: Evidence on gender differences from a sample of australian twins. *Economica*, *64*(1), 119–136.

Monseur, C., & Adams, R. (2009). Plausible Values: How to Deal with Their Limitations. *Journal of Applied Measurement*, *10*(3), 320–334.

Ng, P. (1996). An algorithm for quantile smoothing splines. *Computational Statistics and Data Analysis*, *22*, 99–118.

OECD. (2010). *Pisa 2009 Results: What Students Know and Can Do - Student Performance in Reading, Mathematics and Science* (Rapport technique). OECD Publishing : Auteur.

OECD. (2011). *Pisa: Do students today read for pleasure?* (Rapport technique). OECD Publishing : Auteur.

OECD. (2012). *PISA 2009 Technical Report* (Rapport technique). OECD Publishing : Auteur.

Olson, D. (1977). The language of instruction: The literate bias of schooling. In R. S. Anderson & W. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, NJ : Lawrence Erlbaum Associates.

Portnoy, S., & Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators. *Statistical Science*, *12*(4), 279–296.

Pretorius, E. (2000). What they can't read will hurt them: reading and academic achievement. *Innovation*, *20*(1), 33–41.

Quintano, C., Castellano, R., & Longobardi, S. (2012). The literacy divide: territorial differences in the Italian education system. In *Advanced Statistical Methods for the Analysis of Large Datasets.* Berlin-Heidelberg : Springer-Verlag.

Rangvid, B. S. (2003). *Do Schools Matter? The Influence of School Inputs on Student Performance and Outcomes.* (Mimeo)

Sahlberg, P. (2007). Education policies for raising student learning: the Finnish approach. *Journal of Education Policy*, *22*(2), 147–171.

Scheerens, J. (2005). *Review of school and instructional effectiveness research.* Paper commissioned for the EFA Global Monitoring Report 2005.

Simon, A., H. (2000). Observations on the Sciences of Science Learning. *Journal of Applied Developmental Psychology*, *21*(1), 115–121.

Smith, M. C., Mikulecky, L., Kibby, M. W., Dreher, M. J., & Dole, J. A. (2000). What will be the demands of literacy in the workplace in the next millennium? *Reading Research Quarterly*, *35*(3), 378–383.

Suggate, S. P. (2009). School entry age and reading achievement in the 2006 Program for International Student Assessment (PISA). *International Journal of Educational Research*, *48*, 151–161.

Wang, M. C., Heartel, G. D., & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, *63*(3), 249–294.

Yu, K. (2002). Reversible jump MCMC Approach Quantile Regression. *Computational Statistics and Data Analysis*, *40*(2), 303–315.

Table 1:: Descriptive statistics categorical variables

Label	Description	n	Cum. n	%	Cum. %
PARED	*At least one parent graduated*				
	1=no	10,297	10,297	0,334	0,333
	2=yes	20,608	30,905	0,667	1,000
HOMBOOK	*How many books at home*				
	1=0-10	3,106	3,106	0,101	0,101
	2=11-25	5,400	8,506	0,176	0,278
	3=26-100	9,557	18,063	0,312	0,590
	4=101-200	5,904	23,967	0,193	0,783
	5=201-500	4,092	28,059	0,134	0,917
	6=> 500	2,537	30,596	0,083	1,000
ANTICIP	*If the student has anticipated the entrance to school*				
	1=yes	3,459	3,459	0,116	0,116
	2=no	26,260	29,719	0,884	1,000
RIPISCED	*If the student has ever been flunked*				
	1=no	27,385	27,385	0,959	0,959
	2=yes	1,173	28,558	0,041	1,000
SEX	*Gender*				
	1=female	15,209	15,209	0,492	0,492
	2=male	15,696	30,905	0,508	1,000
PROGRAM	*School Program*				
	1=Liceo	13,677	13,677	0,442	0,442
	2=Technical	9,544	23,221	0,309	0,751
	3=Vocational	6,063	29,284	0,196	0,947
	4=Other	1,621	30,905	0,052	1,000
REGION	*Region of residence of student*				
	–Bolzano (North)	2,144		0,069	
	–Emilia-Romagna (N)	1,494		0,048	
	–Friuli (N)	1,576		0,051	
	–Liguria (N)	1,427		0,046	
	–Lombardy (N)	1,512		0,049	
	–Piedmont (N)	1,518		0,049	
	–Trento (N)	1,449		0,047	
	–Aosta Valley (N)	879		0,028	
	–Veneto (N)	1,577		0,051	
	–Abruzzo (Centre)	1,450		0,047	
	–Lazio (C)	1,462		0,047	
	–Marche (C)	1,512		0,049	
	–Molise (C)	1,209		0,039	
	–Tuscany (C)	1,444		0,047	
	–Umbria (C)	1,562		0,050	
	–Calabria (South)	1,483		0,048	
	–Basilicata (S)	1,530		0,049	
	–Campania (S)	1,431		0,046	
	–Apulia (S)	1,497		0,048	
	–Sardinia (S)	1,416		0,046	
	–Sicily (S)	1,333		0,043	

Continued on next page

Table 1 Descriptive statistics (continuous variables) – *Continued from previous page*

Label	Description	mean	st.dev	min	max
PV	*Plausible values in reading*	486.050	93.020	41.320	751.170
CULTPOSS	*The index is based on the students' responses to whether they had the following items at home: classic literature, books of poetry, and works of art. Higher values indicate better cultural possessions at home*	-0.037	0.830	-1.611	0.816
DISCLIMA	*The index is derived from the students' reports on how often the following happened in their lessons: i) students don't listen to what the teacher says; ii) there is noise; iii) the teacher has to wait a long time for the students to quiet down; iv) students cannot work well; and v) students don't start working for a long time after the lesson begins. Items are inverted for scaling: higher values indicate better disciplinary climate*	0.031	0.054	-2.809	1.838
HEDRES	*The index is based on the items measuring the existence of educational resources at home including a desk and a quiet place to study, a computer that students can use for schoolwork, educational software, books to help with students' school work, technical reference books and a dictionary*	0.070	0.911	-4.250	1.005
ICTHOME	*The index was derived from students' reports on whether any of the following are available for them to use at home: i) a desktop computer ii) a portable laptop or notebook; iii) an Internet connection; iv) a video games console v) a cell phone; vi) MP3/MP4 or iPod or similar; vii) a printer; and viii) a USB stick. Higher values indicate greater ICT availability at home*	0.063	0.893	-4.193	1.415
ICTRES	*The index was derived from students' reports on whether they have an educational software and/or a link to the Internet at home and the number of computers at home. Higher values indicate more ICT resources for learning*	-0.127	0.978	-3.035	1.541
ICTSCH	*The index was derived from students' reports on whether any of the following are available for them to use at school: i) a a desktop computer; ii) a portable laptop or notebook iii) an Internet connection; iv) a printer; v) a USB stick. Higher values indicate more ICT resources for learning at school*	-0.471	1.113	-2.791	1.800
HISEI	*It corresponds to the higher Social Economic Index SEI score of either parents or to the only available parent's SEI score. It uses values from 16 to 90, low values representing lower socio-economic status*	46.860	16.393	16.000	90.000
WEALTH	*The index is based on students' responses on whether they had the following at home: a room of their own; a link to the Internet; a dishwasher; a DVD player; 3 other country specific items; and their responses on the number of cell phones, TVs, computers, cars and the rooms with a bath or shower. Higher values represent higher wealth*	0.079	0.765	-5.125	2.738

Table 2: Estimated coefficients for OLS and QR models for $\tau = 0.1$, $\tau = 0.5$, $\tau = 0.9$

	$\beta_{\tau=0.1}$	p-value	$\beta_{\tau=0.5}$	p-value	$\beta_{\tau=0.9}$	p-value	β_{OLS}	p-value
SEX	-26.21	0.000	-20.99	0.000	-17.41	0.000	-21.41	0.000
ANTICIP	-9.84	0.016	-7.43	0.037	-4.81	0.125	-7.74	0.000
RIPISCED	-45.92	0.000	-47.59	0.000	-46.74	0.000	-46.35	0.000
HISEI	0.62	0.000	0.57	0.000	0.53	0.000	0.60	0.000
PARED	19.21	0.000	13.89	0.000	11.62	0.000	15.23	0.000
HEDRES	5.71	0.009	3.59	0.026	2.98	0.081	4.26	0.000
WEALTH	-9.43	0.000	-8.97	0.000	-10.52	0.000	-9.90	0.000
CULTPOSS	4.91	0.024	4.56	0.004	4.02	0.034	4.78	0.000
HOMBOOK	7.01	0.000	8.36	0.000	8.80	0.000	8.06	0.000
ICTRES	3.45	0.102	4.30	0.004	4.78	0.041	3.99	0.000
ICTHOME	4.27	0.016	1.50	0.201	0.74	0.620	2.21	0.019
ICTSCH	-1.15	0.348	-1.20	0.197	-0.19	0.887	-0.82	0.332
REGION								
–Bolzano (North)	66.17	0.000	75.24	0.000	61.89	0.000	70.06	0.000
–Emilia R. (N)	60.81	0.000	69.70	0.000	60.40	0.000	64.26	0.000
–Friuli F.G. (N)	73.03	0.000	73.79	0.000	62.55	0.000	71.14	0.000
–Liguria (N)	43.06	0.000	52.63	0.000	49.13	0.000	49.21	0.000
–Lombardy (N)	77.99	0.000	82.73	0.000	73.88	0.000	78.90	0.000
–Piedmont (N)	61.34	0.000	66.66	0.000	54.09	0.000	61.60	0.000
–Trent (N)	74.47	0.000	77.69	0.000	63.88	0.000	73.37	0.000
–Aosta Valley (N)	73.51	0.000	80.47	0.000	67.19	0.000	76.52	0.000
–Veneto (N)	71.05	0.000	78.14	0.000	67.28	0.000	73.66	0.000
–Abruzzo (Centre)	32.97	0.000	38.53	0.000	29.99	0.000	34.36	0.000
–Lazio (C)	29.93	0.000	33.75	0.000	24.99	0.000	29.32	0.000
–Marche (C)	58.21	0.000	65.66	0.000	54.12	0.000	60.27	0.000
–Molise (C)	19.60	0.007	25.13	0.000	16.85	0.009	21.38	0.000
–Tuscany (C)	45.76	0.000	52.77	0.000	44.15	0.000	48.21	0.000
–Umbria (C)	39.01	0.000	52.60	0.000	45.91	0.000	47.43	0.000
–Basilicata (South)	23.06	0.005	26.05	0.000	22.74	0.006	24.9	0.000
–Campania (S)	4.71	0.609	10.21	0.175	6.973	0.277	8.126	0.227
–Apulia (S)	37.04	0.000	46.11	0.000	42.42	0.000	42.5	0.000
–Sardinia (S)	17.54	0.012	20.66	0.000	18.87	0.005	18.96	0.000
–Sicily (S)	6.334	0.603	14.76	0.031	14.45	0.133	11.95	0.107
DISCLIMA	8.66	0.000	8.19	0.000	7.27	0.000	8.05	0.000
PROGRAM								
-Technical	-42.72	0.000	-44.68	0.000	-42.79	0.000	-43.73	0.000
-Vocational	-97.75	0.000	-92.51	0.000	-84.55	0.000	-91.39	0.000
-Other	-126.03	0.000	-119.50	0.000	-103.47	0.000	-117.6	0.000
CONSTANT	427.99	0.000	503.75	0.000	582.73	0.000	502.6	0.000

Baseline is the category coded with "1" in Table 1

Table 3: Predicted reading scores by school program and differences with baseline school program

				Δ with Program Baseline		
PROGRAM	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
Liceo	465.72	550.38	627.13	0.00	0.00	0.00
Technical	402.17	485.39	565.63	-63.55	-64.99	-61.50
Vocational	338.58	429.62	516.24	-127.14	-120.76	-110.89
Other	318.37	410.82	502.60	-147.35	-139.56	-124.53

Table 4: Predicted reading scores by region and differences with baseline region

				Δ with Region Baseline		
REGION	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
Abruzzo (C)	408.73	494.03	572.48	28.95	35.31	27.10
Apulia (S)	408.60	495.90	580.00	28.82	37.18	34.61
Aosta V. (N)	437.42	525.03	601.29	57.64	66.30	55.91
Basilicata (S)	398.82	480.91	564.96	19.03	22.19	19.58
Bolzano (N)	424.26	513.26	591.13	44.48	54.54	45.75
Calabria (S)	379.78	458.72	545.38	0.00	0.00	0.00
Campania (S)	381.75	466.19	549.68	1.97	7.47	4.29
Emilia-R. (N)	430.08	519.22	597.36	50.30	60.50	51.98
Friuli (N)	445.27	525.85	603.63	65.48	67.13	58.25
Lazio (C)	408.67	492.93	571.31	28.88	34.21	25.93
Liguria (N)	420.35	509.75	594.54	40.56	51.03	49.15
Lombardy (N)	451.15	535.81	615.19	71.36	77.09	69.81
Marche (C)	427.31	514.22	590.37	47.52	55.50	44.98
Molise (C)	393.78	478.24	557.52	13.97	19.51	12.14
Piedmont (N)	432.83	518.40	594.38	53.05	59.68	49.00
Sardinia (S)	402.71	485.08	569.80	22.92	26.35	24.42
Sicily (S)	388.40	474.98	560.70	8.61	16.26	15.32
Tuscany(C)	424.73	511.16	590.20	44.94	52.44	44.81
Trento (N)	436.55	520.83	597.47	56.76	62.11	52.08
Umbria (C)	408.35	502.51	583.91	28.56	43.79	38.52
Veneto (N)	433.56	520.81	599.09	53.78	62.09	53.71

Table 5: Relative variation about OLS index – RVAO – by region

			RVAO	
REGION	$QR_{0.1}/OLS$	$QR_{0.9}/OLS$	$Q_{0.1}$	$Q_{0.9}$
Abruzzo	0.8001	1.1207	-20%	12%
Aosta V.	0.8925	1.2268	-11%	23%
Apulia	0.8477	1.2032	-15%	20%
Basilicata	0.8366	1.1852	-16%	19%
Bolzano	0.8232	1.1470	-18%	15%
Calabria	0.7112	1.0213	-29%	2%
Campania	0.7715	1.1109	-23%	11%
Emilia-R.	0.8865	1.2313	-11%	23%
Friuli	0.9388	1.2727	- 6%	27%
Lazio	0.8036	1.1235	-20%	12%
Liguria	0.8101	1.1458	-19%	15%
Lombardy	0.9038	1.2325	-10%	23%
Marche	0.8160	1.1275	-18%	13%
Molise	0.7596	1.0754	-24%	8%
Piedmont	0.8804	1.2090	-12%	21%
Sardinia	0.7882	1.1152	-21%	12%
Sicily	0.8430	1.2170	-16%	22%
Tuscany	0.9115	1.2666	- 9%	27%
Trento	0.8467	1.1588	-15%	16%
Umbria	0.7772	1.1114	-22%	11%
Veneto	0.8526	1.1781	-15%	18%

Appendix: Figures

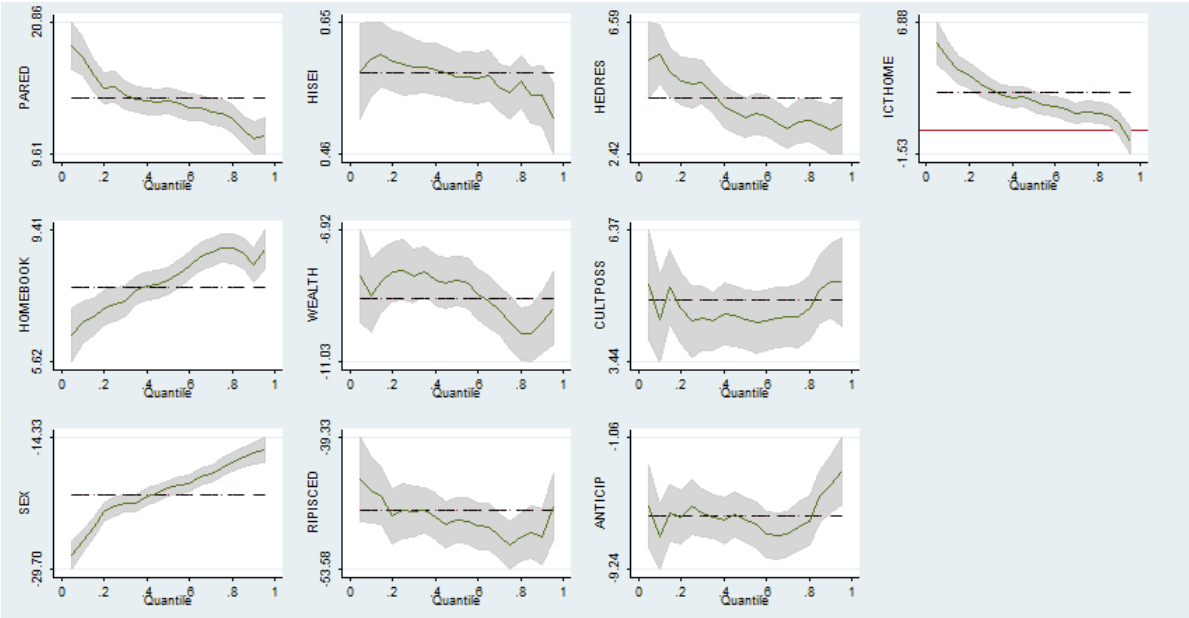

Figure 1. Student and family background covariates: effect on different quantiles

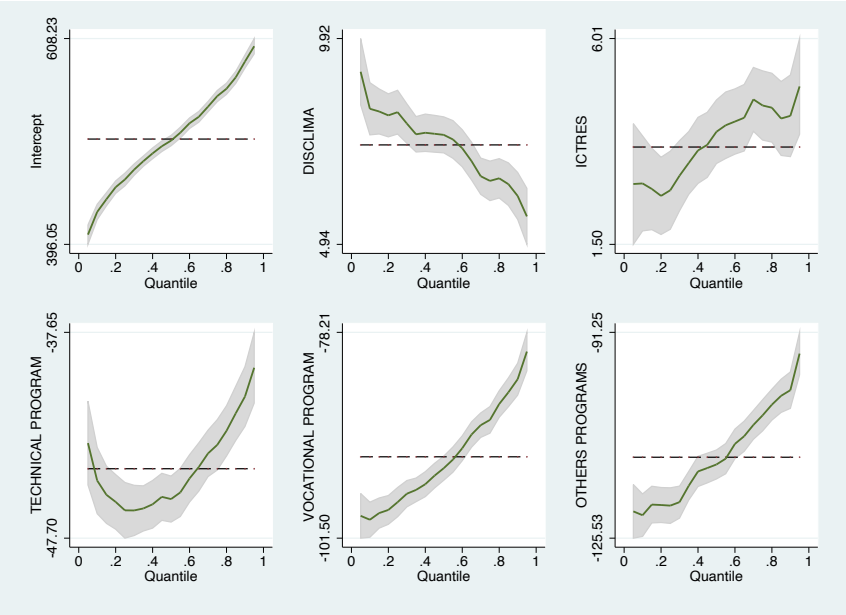

Figure 2. School related covariates: effect on different quantiles

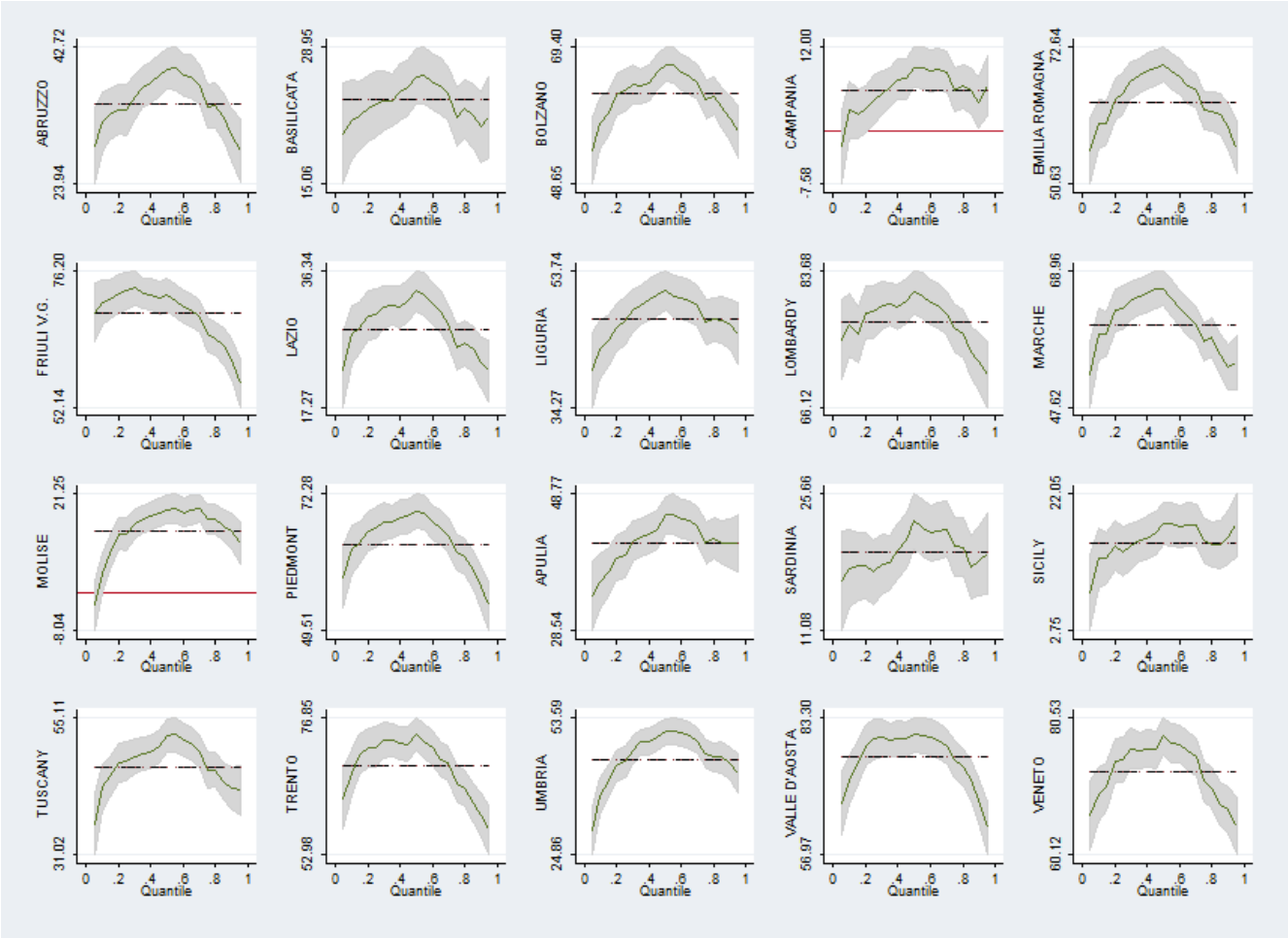

Figure 3. Region covariate: effect on different quantiles