METHODOLOGY

Beyond the mean estimate: a quantile regression analysis of inequalities in educational outcomes using INVALSI survey data

Antonella Costanzo[*] and Marta Desimoni

*Correspondence:
antonella.costanzo@invalsi.it
National Institute for the
Evaluation of Education
System, Via Ippolito Nievo,
35, 00153 Rome, Italy

Abstract

The number of studies addressing issues of inequality in educational outcomes using cognitive achievement tests and variables from large-scale assessment data has increased. Here the value of using a quantile regression approach is compared with a classical regression analysis approach to study the relationships between educational outcomes and likely predictor variables. Italian primary school data from INVALSI large-scale assessments were analyzed using both quantile and standard regression approaches. Mathematics and reading scores were regressed on students' characteristics and geographical variables selected for their theoretical and policy relevance. The results demonstrated that, in Italy, the role of gender and immigrant status varied across the entire conditional distribution of students' performance. Analogous results emerged pertaining to the difference in students' performance across Italian geographic areas. These findings suggest that quantile regression analysis is a useful tool to explore the determinants and mechanisms of inequality in educational outcomes. A proper interpretation of quantile estimates may enable teachers to identify effective learning activities and help policymakers to develop tailored programs that increase equity in education.

Keywords: Large-scale assessment, Educational predictors, Primary school, Quantile regression

Background

Learning outcomes are considered positive indicators towards future economic social and cultural opportunities of a number of countries (Woessmann 2004). Therefore, over the last decades, studies facing inequality issues in educational outcomes using cognitive achievement tests and variables from large-scale assessment data have increased. From a methodological point of view, the traditional approach used to explore the relationship between explicative variables and students' performance is based on average effects within a classical linear regression setup (OECD 2012). Undoubtedly, estimates on average will yield straightforward interpretations but will represent only a part of the information concerning the complex and nuanced nature of the relation between predictors and conditional performance distribution. Essentially, the main concern for policy purposes might be not only to assess if the relevant variables carry an impact or not, but

also to investigate if and how they are associated with greater or lower variation in educational outcomes.

Consistently, some studies have been recently enhanced by using models which extend the viewpoint on the whole conditional distribution of performance representing the different levels of students' attainment (Fryer and Levitt 2010; Davino et al. 2013), i.e. the quantile regression model (Koenker and Basset 1978). In the present study, the differential impact of variables related to inequalities in educational outcomes will be assessed through the quantile regression (QR) approach using data from the Italian Annual Survey on Educational Achievement (SNV) carried out by the National Institute for the Evaluation of Education System (INVALSI). Focusing on Italian primary school, this paper explores the added value of the QR approach as a policy research tool to expand knowledge concerning educational predictors of pupils' performance within the INVALSI large-scale assessment setting.

Quantile regression model: the essentials

Quantile regression (Koenker and Basset 1978) may be viewed as an extension of least squares estimation of conditional mean models to estimate an ensemble of models for several conditional quantile functions, taking into account the effects a set of covariates plays on a response variable.

While the classical linear regression model specifies the change in the conditional mean of the dependent variable associated with a change in the covariates, the quantile regression model specifies changes in the conditional quantiles. Therefore, as multiple quantiles can be modeled, it would be possible to achieve a more complete understanding of how the response distribution is affected by predictors by obtaining information about changes in location, spread and shape (Koenker 2005; Davino et al. 2013). In analogy with the classical linear regression framework, a linear regression model for the θ-th conditional quantile of y_i can be expressed as

$$\mathbf{Q}_{y_i(\theta)|\mathbf{x}_i} = \mathbf{x}_i^T \beta_\theta \tag{1}$$

where y is a scalar dependent variable, $\mathbf{x}_i^T$ is the $k \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is the coefficient vector, θ is the conditional quantile of interest and it is assumed that

$$\mathbf{Q}_\theta(u_{i,\theta}|x_{i,\theta}) = 0 \tag{2}$$

where $u_{i,\theta}$ is the residual term of the regression model at the θ-th quantile.

From Eq. 1, it results that, compared with classical linear regression methods, based on minimizing sums of squares residuals, quantile regression methods are based on minimizing asymmetrically weighted absolute residuals:

$$\min_{\boldsymbol{\beta}} \sum_{y_i \geq \mathbf{x}_i^T \boldsymbol{\beta}} \theta |y_i - \mathbf{x}_i^T \boldsymbol{\beta}| + \sum_{y_i < \mathbf{x}_i^T \boldsymbol{\beta}} (1 - \theta)|y_i - \mathbf{x}_i^T \boldsymbol{\beta}| \tag{3}$$

By setting $\theta = 0.5$, Eq. 3 provides the median solution, while the use of any θ between 0 and 1 allows to study the dependence structure at any location of the response conditional distribution.

As Hao and Naiman ([2007](#)) pointed out, the estimation of coefficients for each quantile regression is based on the weighted data of the whole sample, not just the portion of the sample at that quantile. Further details about the algorithms for computing the quantile regression coefficients can be found in Koenker ([2005](#)).

The estimated $\hat{\boldsymbol{\beta}}_\theta$ in QR linear models have the same interpretation as those of other linear models, i.e.

$$\hat{\beta}_\theta = \frac{\mathrm{d}Q_\theta(y_i|\mathbf{x})}{\mathrm{d}x} \tag{4}$$

meaning that each $\hat{\beta}_\theta$ coefficient can be interpreted as the rate of change in the θ-th quantile of the dependent variable distribution per one unit change in the value of the corresponding regressor, holding constant the others.

However, important differences between LS and QR models refer to the monotone equivariance and robustness to distributional assumptions in conditional quantiles versus the lack of these properties in the conditional mean setup.

Background

In educational research, exploring if and how individual characteristics and contextual factors relate to learning outcomes is considered of great interest in order to deal with inequality issues. For example, gender differences and the impact of students' socio-economic conditions on learning achievement have been largely explored by international comparative studies, such as those carried out by the International Association of the Evaluation of Educational Achievement (IEA), the Organization for the Economic Cooperation and Development (OECD), and national large-scale assessments, e.g. the National Assessment of Educational Progress (NAEP). Also, numerous scholars have also investigated the gender gap and inequality of opportunity in education through secondary analysis and meta-analysis (Baye and Monseur [2016](#); Hansen and Gustafsson [2016](#)) using large-scale assessment data. Moreover, the relationship between educational outcomes and other predictors, e.g. children preschool attendance and psychological factors, such as attitudes, students' self-engagement and self-belief, has been largely explored in large-scale assessment studies aiming to provide information regarding factors enhancing students' achievement, e.g. OECD—Programme for International Student Assessment (PISA), IEA—Progress in International Reading Literacy Study (PIRLS) and IEA—Trends in International Mathematics and Science Study (TIMSS).

Traditional approaches used to explore the impact of selected covariates on students' performance are based on conditional expectations, i.e. within a classical linear regression framework. The classical regression model allows examination of important questions, such as: "does a certain variable influence educational achievements, on average?" or "are there differences in students' average performance conditional on a set of individual characteristics?". However, to focus only on mean differences and effects might overlook some important information in relation to the heterogeneous effect of covariates along the performance distribution. To better understand the determinants and mechanism of inequality it would be important to assess if and how educational predictors relate differently to the conditional distribution of performances, that is, according to the students' proficiency level. A number of scholars have already highlighted these

features with regard to gender gap issues in educational achievement. For example, Feingold (1995) pointed out that analyzing gender disparities by examining differences at the population mean is certainly useful but can lead to misleading conclusions when distribution of performance differs between males and females. Halpern et al. (2007) in their review on gender differences suggest that the magnitude of gender gap in educational achievement might depend on which portion of the ability distribution is investigated.

In a very recent study, through a meta-analysis carried out on databases from IEA and OECD PISA surveys, Baye and Monseur (2016) compare the effect size of gender gap differences computed at the extreme tails of the achievement distribution with the effect size for the mean scores in reading literacy, mathematics and science. Results indicate that the size of differences between males and females in learning achievement varies according to the proficiency level, i.e. between low and high achievers. In their analysis, the authors discuss the risk, from a policy perspective, of generalizing results based on central tendency statistics about gender disparities to the whole distribution of performances. Overall, a deeper understanding of the relationship between personal and socio-economic factors and students' achievement can be provided by a quantile regression approach which allows a more complete understanding of how the response distribution is affected by predictors by obtaining information about changes in location, spread and shape (Hao and Naiman 2007). Evidence related to the advantages of exploring gender impact across the whole distribution of performances through the QR approach can be found in some studies on gender differences in reading and mathematics achievements (Penner and Paret 2008; Robinson and Lubiensky 2011; Contini et al. 2016). In particular, using United States (US) nationally representative data from kindergarten to fifth grade, Penner and Paret (2008) found that at the beginning of the educational process, among high achievers boys outperform girls in mathematics, whilst this pattern is reversed at the bottom of the students' performance distribution. However, the gender gap in favor of females disappears by the end of third grade; in fact, boys outperform girls along the entire distribution of mathematical performance.

Consistently, Robinson and Lubiensky (2011) results on nationally representative US samples indicate that the gap favoring boys in mathematics emerges among high achievers at the end of kindergarten. The authors highlight that the gender gap persists in primary school and it becomes more effective as the pupils' years of schooling increase. Subsequently, in middle school, the gender gap becomes less pronounced throughout the distribution of performances and the largest reduction can be assessed at the tails of the conditional distribution. As regards students' reading abilities, Robinson and Lubiensky (2011) found that in primary school the gender gap increases as the years of schooling increase: in particular, boys lose more ground than girls and this phenomenon tends to become wider among low achievers rather than high achievers.

Moving to the Italian educational context, Contini et al. (2016) analyzed cross-sectional data from the INVALSI large-scale survey. The authors found that in Italy differences between males and females regarding mathematics achievement increase with students' age, from the second up to the eight level of schooling. Also, the evidence of disparities where males outperform females is concentrated among high achievers. As for reading achievement, by using data from OECD-PISA 2009 survey on 15 year-old

Italian students, Giambona and Porcu (2015) found that the gender gap is more pronounced at the lower tail of the conditional distribution of performances.

Overall, the results obtained through the use of the quantile regression approach describe distributional differences between males and females in addition to the well-known mean differences. The findings reveal a larger gender gap at high quantiles of the mathematics score distribution with boys outperforming girls, and a larger gender gap at the lower tail of the reading score distribution with females outperforming males compared to the mean estimates. A quantile regression perspective has been also exploited in some studies aiming to investigate the role of psychological and socio-demographical variables affecting educational outcomes.

For example, in their recent analysis of OECD-PISA 2012 data on 15-year-old Turkish students, Gursakal et al. (2016) illustrate that some variables referred to students anxiety, the degree of familiarization with information and communication technology, family background and school climate are significantly associated with mathematics achievement and the impact of these factors is different across the quantiles of the conditional PISA test scores distribution. Focusing on the impact of students' family background, the authors found that home educational resources significantly affect the entire math achievement distribution, except at the highest quantiles (95th percentile) corresponding to top performing students. Furthermore, the authors highlight that the association between family wealth and mathematics achievement becomes more effective from students occupying the lowest quantiles (i.e. 25th percentile) to those in the highest quantiles (i.e. 95th percentile) of the performance distribution.

Regarding the Italian context, the previously mentioned Giambona and Porcu (2015) analysis of OECD-PISA 2009 survey data on 15-year-old Italian students shed light on whether and how relevant predictors, i.e. individual and family background variables, the school program and the school geographic location are significantly associated with students' performance through the estimation of a quantile model exploring the effects of the covariates at different levels of reading achievement. Interestingly, it emerges that family background variables, i.e. parental occupation and basic home educational resources are related to the lower tail of the reading skills distribution more than to the upper tail. On the contrary, the number of books at home seems more positively associated with the higher quantiles rather than with the lower quantiles of the conditional performance distribution. It might be argued that the presence of cultural resources at home tends to boost the performance of high-ability students and to have a small impact on low-ability students. As for school programs (e.g. academic, technical and vocational), it emerges that high performers always exhibit narrower differences in the conditional performance distribution than low performers. Different effects at different proficiency levels of reading achievement also emerge when school geographical location is considered. In particular, quantile regression estimates suggest different patterns of students' achievement across Italian regions.

To summarize, although the quantile regression approach has been less frequently adopted in educational studies than in other fields (e.g. in economic and social sciences), promising results are emerging from recent researches. It is also worthy of notice that most of the existing studies following the QR approach refer to secondary school students' data (e.g. Gursakal et al. 2016; Giambona and Porcu 2015). Instead, less is known

about the differential impact of learning-related factors on children's performance in primary school, excepting some studies about gender gap in early grades (e.g. Penner and Paret 2008; Robinson and Lubiensky 2011; Contini et al. 2016). Notwithstanding, as some explicative variables might narrow or spread its effects on students' performance along their entire educational career, having a complete picture of the role of predictors on educational outcomes since primary school would be particularly important to foster the existing knowledge about the association of predictors and learning achievements.

Using the INVALSI survey data, this paper illustrates the added value of quantile regression to assess the relation between individual characteristics, geographical variables and pupils' performance in a nationally representative large-scale assessment setting. The advantages of QR perspective compared to LS regression will be addressed under two main aspects: (i) the possibility to approximate the whole distribution of the response variable conditional on the values of the selected predictors and (ii) the value of quantile regression in providing a more detailed picture of the relationships between covariates and educational outcomes.

Methods

Participants and procedure

This study is a secondary analysis carried out on data from the INVALSI national large-scale assessment program. INVALSI yearly carries out standardized tests to assess students' achievement in mathematics and reading (i.e. reading comprehension and grammatical knowledge), and to evaluate the overall quality of the educational offering of schools and vocational training institutes.

The INVALSI survey currently involves the universe of pupils attending the 2nd and the 5th grades primary school, 8th grade lower secondary school, and the 10th grade upper secondary school (about 2,850,000 students and 15,000 schools). The INVALSI tests are administered by the schools' teachers, who report students' answers on electronic sheets and forward the relative data to INVALSI.

In a representative sample of randomly selected classes (National Sample, NS), tests are administered in the presence of an external examiner, who monitors students during the test administration and transmit data to INVALSI (Falorsi 2007). External monitoring is thought to improve data reliability by reducing biases due to cheating phenomena during the test administration. Indeed, cheating behaviors—which may be undertaken by both students (e.g. by copying and cooperating with other students) and teachers (e.g. by suggesting the correct answers)—are considered an important issue of concern in standardized testing. In fact, they can lead to biased results, such as overestimation of achievement levels for cheating classes (Ferrer-Esteban 2013).

As well as the tests, questionnaires are also administered to students from 5th grade and 10th grade in order to collect data on socio-demographic variables. Further information regarding students (e.g., family background) together with classes/school characteristics (e.g. number of students enrolled, time schedule) are also provided to INVALSI by the schools' secretarial offices. Data on the school geographical location are also collected.

This study focuses on primary education by analyzing datasets of second- and fifth-graders referring to the school year 2014–2015. Consistently with INVALSI annual

report results (INVALSI 2015a, b), a secondary analysis is carried out on NS data, thus considering only students from classes with external monitoring in testing procedures. For each school grade, two datasets are analyzed: a dataset with students' characteristics and performance on the mathematics test and a dataset with students' characteristics and performance on the reading test. For the second-graders, after removing cases with missing values referring to the selected variables, the mathematics dataset consists of 15,132 pupils (7726 males and 7406 females) from 548 schools and the reading assessment dataset consists of 15,483 students (7390 males and 7093, females) from 543 schools. For the fifth-graders, the final sample for mathematics data consists of 19,109 pupils (9749 males and 9360 females) from 614 schools and the final sample for reading data consists of 18,388 pupils (9347 males and 9041 females) from 605 schools. All the databases are available on the INVALSI Data Repository: http://invalsi-serviziostatistico. cineca.it/ upon request. Methodological information is available in the technical reports of the National Annual Survey on the official website: http://invalsi-areaprove.cineca.it.

Measures

Reading and mathematics tests

In the school year 2014–2015, INVALSI assessed students' achievements in two main content domains: mathematics and reading. Regarding the second grade students, the reading test consists of a passage of narrative text, followed by eighteen items (17 multiple-choice questions and 1 open-ended short-term question), and of two exercises concerning lexical and semantic knowledge (1 complex multiple-choice item and 1 matching task).

As for the fifth grade students, the reading test consists of a passage of narrative text with nineteen items (13 multiple choice items; 3 complex multiple-choice items and 3 open-ended short-term questions); a passage of expositive text with twelve items (9 multiple choice items, 1 complex multiple-choice item, 2 open-ended short-term questions) and ten items about grammar knowledge (5 multiple choice items; 1 complex multiple-choice item; 4 open-ended short-term questions).

The mathematics test for the second grade students comprises twenty-three items covering three sub-domains: Numbers (14 items), Space and Figures (7 items) and Data and Previsions (2 items); the fifth grade mathematics test comprises thirty items covering Numbers (8 items); Space and Figure (8 items), Data and Previsions (6 items) and Functions and Relationships (8 items) sub-domains. Items vary by format: in second grade the mathematics test is composed of 9 multiple choice items, 3 complex multiple-choice items and 18 open-ended short-term questions; in fifth grade, the mathematics test is composed of 13 multiple choice items, 13 complex multiple-choice items and 22 open-ended short-term questions and 1 open-ended long-answer question.

The different items composing all the INVALSI tests are dichotomously scored, e.g. the items are scored as correct/incorrect. For each test, all the items are thought to be reflective indices of the same overall construct, which is described in the INVALSI test theoretical framework, and it is hypothesized and empirically verified that at least essential unidimensionality holds for the INVALSI test (INVALSI 2015b). Consistently, for each test, the observed score (that is, the percentage of correct answers) and the weighted likelihood estimates (WLE) of individual parameters of the Rasch model (Rasch 1960, 1980) are

reported in the corresponding database. In the present paper, the WLE estimates of students' mathematics and reading achievement have been considered as outcome variables.

Students' characteristics and geographical variables

Referring to the literature relative to correlates of reading and math achievements (Baye and Monseur 2016; Gursakal et al. 2016; Giambona and Porcu 2015), as well as the information about students' achievement estimates two group of variables have been considered from the INVALSI dataset.

The first group of variables includes those that might be associated with inequalities in educational outcomes, namely student gender, immigrant status (categories: native students and non native students, the latter category including students born outside Italy and whose parents were also born in another country and students born in Italy but whose parent(s) were born in another country), socio-economic background, and the geographical location of the school.

Regarding the information about family background for pupils attending the fifth grade, the individual Economic, Social and Cultural Status, known as ESCS index (OECD 2007) is exploited. This index is computed following the same procedure adopted by the OECD in the Program of International Students Achievement (PISA). The ESCS is based on the following variables: the International Socio-Economic Index of Occupational Status (ISEI); the highest level of education of the student's parents, converted into years of schooling, and a composite index of family wealth including information about the students' family home educational resources. Further details about methodological aspects can be found in Campodifiori et al. (2008). Since information about the ESCS index is not available in reference to the school's second grade, the highest level of education of students' parents (pared) according to the ISCED classification and converted into years of schooling, is considered as a proxy of pupils' socio-economic background. This choice is in line with the widespread approach adopted by numerous researches (OECD 2015).

With respect to the geographical location of the school, it is important to note that all of the most significant international surveys on educational achievement (including the national assessment INVALSI) and further researches (Checchi and Peragine 2010; Montanaro 2008) find a relevant gap in performance across the Italian regions, with students in the Southern Italy being far behind those in the north in all subjects assessed (reading, mathematics and science). Therefore, to address from a descriptive perspective the inequality of opportunity in education going beyond the well known Italian north–south division, in this study the school's geographical location (north-Italy, centre-Italy and south-Italy) is also considered.

The second group of variables includes control variables, which do not represent the focus of the study but are typically related to students' performance: kindergarten attendance (yes, no) and enrolment in primary school (early, regular, late enrollment). To better understand what the latter variable indicates it is important to explain that in the Italian educational system, primary school involves pupils from 6 to 10 years old. However, children's families have the opportunity under certain conditions- to choose an early enrolment to education for their children. This means that pupils can start primary school at 5 years old instead of 6 years old. By the same token, a late enrolment refers to students repeating the year or entering late in primary school.

Statistical analysis

Students' performance in reading and mathematics are regressed on the selected independent variables using the ordinary least squares regression (LS) and the QR approach for five conditional quantiles, $\theta = (0.1, 0.25, 0.5, 0.75, 0.9)$, that are assumed to represent, in this setting, the different levels of students' attainment.

Even if, from a theoretical point of view, it would be possible to estimate an infinite number of quantiles, focusing on a limited number permits attainment of a sufficient overview about the distributional impact of educational predictors on students' learning outcome. Indeed, this choice permits management of the trade-off between the advantage of getting additional data provided by the QR model and the informational efficiency related to how obtained results can be effectively exploited in empirical applications. Inferences about the estimated coefficients are based on non-clustered standard errors both for LS and QR estimates.

To provide more enlightened statistical statements about the resulting estimates, the homogeneity assumption of regression slopes across quantiles is formally assessed through the equivalence test (Koenker 2005), where the null hypothesis is that the distinct parameter estimates are the same at the different conditional quantiles. The test of equality of slopes is considered as a separate test meaning that the null hypothesis is equivalent to testing if each predictor in the specified model has a constant effect across different quantiles. More details on this approach can be found in Basset and Koenker (1982a, b), and Gutenbrunner et al. (1993). The typical implementation of such a test consists of an F test. All the computations have been realized with R software using the library quantreg developed by Koenker and Basset (1978).
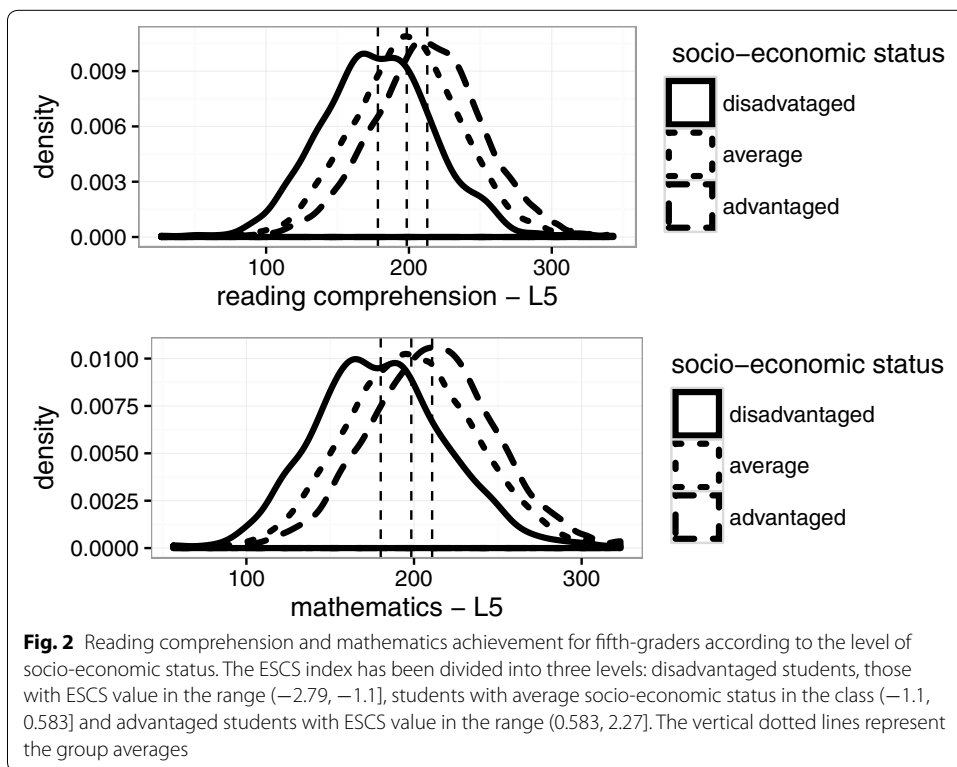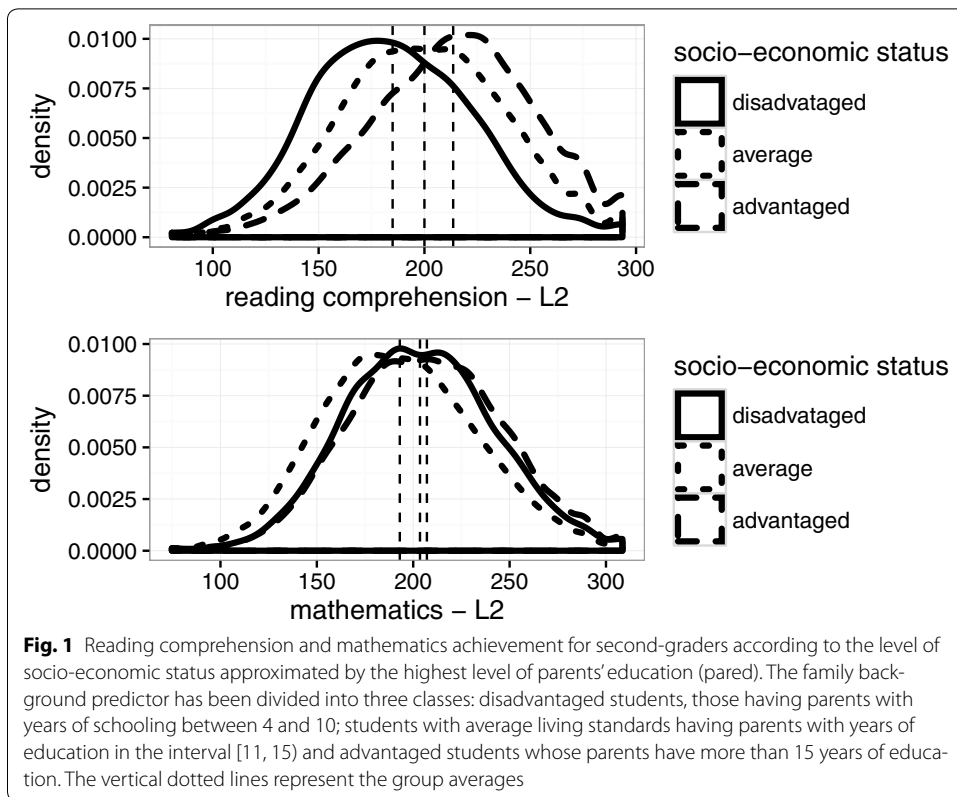
Results

Descriptive statistics

Table 1 reports the main summary statistics concerning pupils' performance in reading and mathematics' tests in second grade and in fifth grade. For the sake of simplicity, in the following tables and figures the symbols L2 and L5 are used to indicate second-graders and fifth-graders, respectively.

Figures 1 and 2 depict the density plot of the test scores in reading comprehension and mathematics in the second (L2) and in the fifth (L5) year of schooling with respect to the students' socio-economic status.

Table 1 Main summaries for students' performance in reading and mathematics in second grade (L2) and fifth grade (L5)—school year 2014–2015

	Test score distribution						
	Mean	SD	Min	q_1	Median	q_3	Max
L2							
Reading	201.3	40.1	80.7	170.3	198.2	228.3	293.6
Mathematics	201.1	40.6	74.8	170.0	202.6	225.7	308.5
L5							
Reading	200.4	39.7	26.5	175.9	199.4	224.4	343.2
Mathematics	201.0	39.7	56.5	172.2	201.2	227.4	323.2

Fig. 1 Reading comprehension and mathematics achievement for second-graders according to the level of socio-economic status approximated by the highest level of parents' education (pared). The family background predictor has been divided into three classes: disadvantaged students, those having parents with years of schooling between 4 and 10; students with average living standards having parents with years of education in the interval [11, 15] and advantaged students whose parents have more than 15 years of education. The vertical dotted lines represent the group averages

Fig. 2 Reading comprehension and mathematics achievement for fifth-graders according to the level of socio-economic status. The ESCS index has been divided into three levels: disadvantaged students, those with ESCS value in the range (−2.79, −1.1], students with average socio-economic status in the class (−1.1, 0.583] and advantaged students with ESCS value in the range (0.583, 2.27]. The vertical dotted lines represent the group averages

Figures 1 and 2 depict the density plot of the students' performance (WLE estimates) in reading and mathematics in second grade and fifth grade with respect to the socio-economic status.

Interestingly, in second grade the shape of pupils' achievement distribution in reading varies with the family background. On the other hand, shape changes appear not to be particularly pronounced for the mathematics achievement distribution.

Regarding pupils in fifth grade (Fig. 2), the shape of the performance distribution varies across the level of ESCS providing evidence of a particular pattern in the positive relationship between students' family background and the distribution of performance. For instance, for students belonging to the lowest class of the ESCS index (see Fig. 2), the distribution of mathematical achievement shows a bimodal pattern. Figures 3 and 4 depict the density plot of achievement scores in both subjects for males and females in second and fifth grades. Beyond the average results, it emerges that the gender gap can be better appreciated in terms of distributional differences. In particular, from a descriptive point of view, the gender gap in favor of males in mathematics seems particularly evident among high performers, i.e. students located at the upper tail of the conditional test score distribution from second grade and tends to worsen in fifth grade.

The estimated LS and QR models

As already stated, the advantages of using the QR perspective compared to the LS regression setting are explored with respect to two main aspects: the possibility to approximate the whole distribution of the response variable conditional on the values of the selected predictors and the ability to enrich results on the relationships between covariates and the dependent variable. Figure 5 reports the observed distribution of students' math achievement in fifth grade (straight line) and the estimated performance
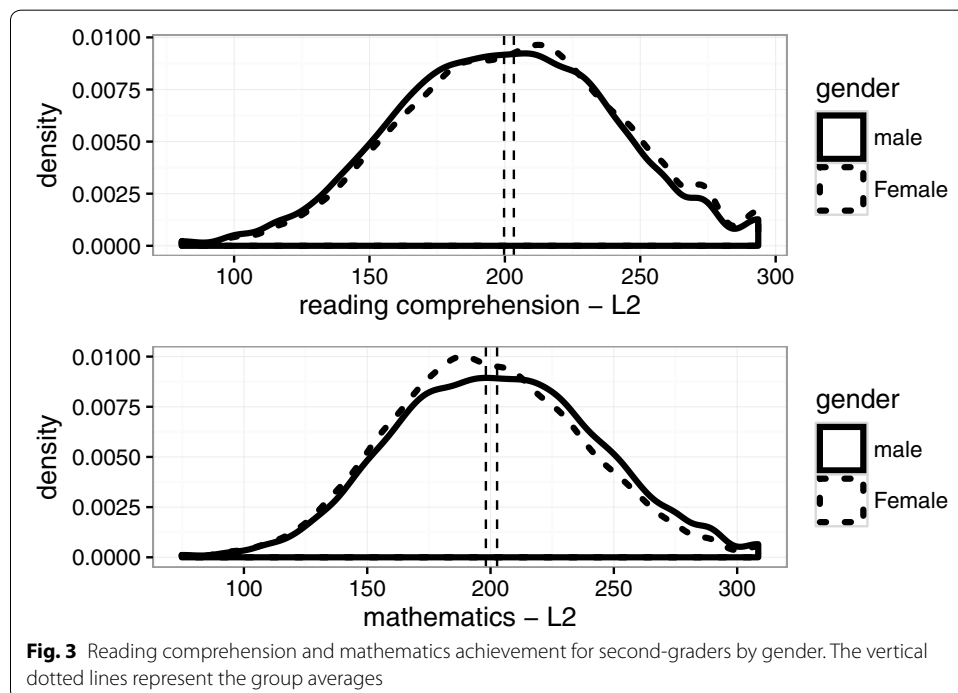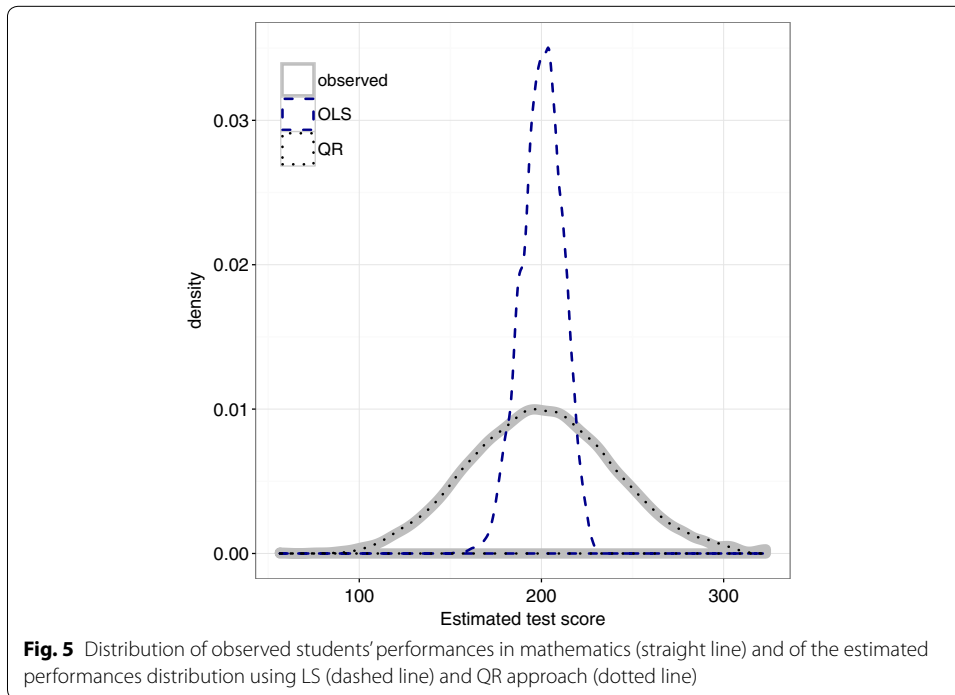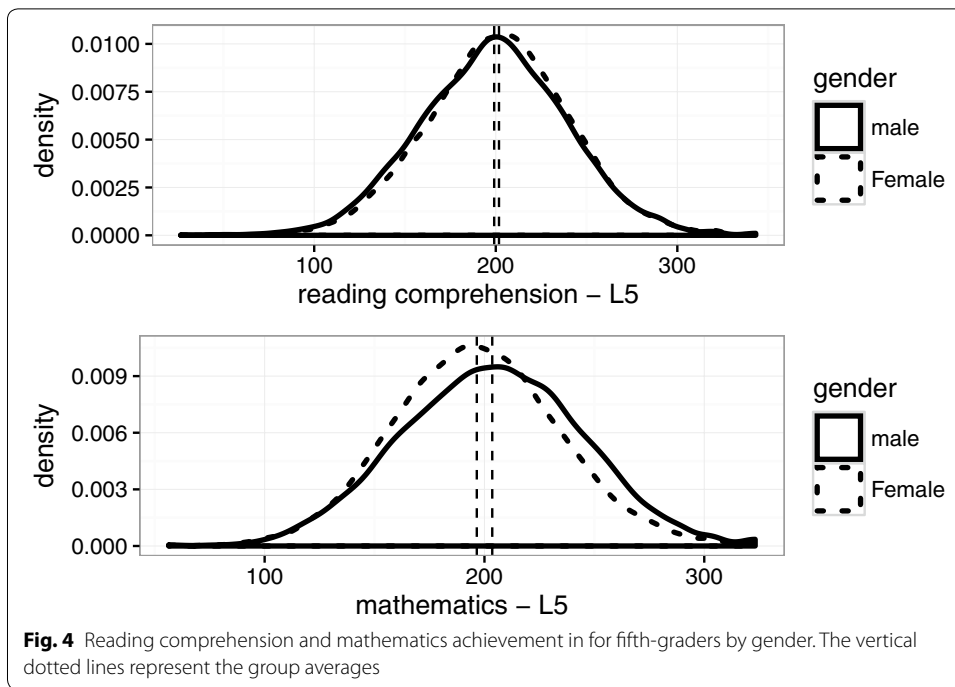

Fig. 3 Reading comprehension and mathematics achievement for second-graders by gender. The vertical dotted lines represent the group averages

Fig. 4 Reading comprehension and mathematics achievement in for fifth-graders by gender. The vertical dotted lines represent the group averages

Fig. 5 Distribution of observed students' performances in mathematics (straight line) and of the estimated performances distribution using LS (dashed line) and QR approach (dotted line)

distribution using LS estimates (dashed line) and QR estimates (dotted line). The QR ability to fully characterize the distributional features of the response variable emerges: the QR estimated density is, indeed, almost equivalent to the observed distribution of mathematical performance, while it is evident that the approximation obtained focused

Table 2 LS and QR estimates of the effects of covariate on reading comprehension in primary school—second grade

	LS	Reading comprehension—L2					
		QR					
		$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$	p value equiv.test
(Intercept)	170.03	127.15	142.24	167.13	193.70	212.61	
	(2.08)	(4.00)	(3.41)	(2.78)	(3.10)	(4.08)	
Female	*3.73*	*4.73*	*3.40*	*3.23*	*3.72*	*8.52*	< 0.01
	(0.64)	(1.29)	(0.91)	(1.39)	(1.72)	(2.47)	
Early enrolment	−5.20	−2.99	−3.20	−6.12	*−6.44*	*−12.92*	0.36
	(2.73)	(5.31)	(2.97)	(3.67)	(2.50)	(4.46)	
Late enrolment	*−15.52*	*−26.60*	*−21.53*	*−15.24*	*−14.00*	*−9.92*	< 0.01
	(2.53)	(3.73)	(3.44)	(3.66)	(3.54)	(4.24)	
Kindergarten:yes	*4.02*	3.11	3.70	3.71	*6.20*	*8.44*	0.13
	(1.61)	(3.20)	(2.80)	(2.12)	(2.33)	(3.38)	
Immigrant	*−13.96*	*−9.47*	*−14.46*	*−14.69*	*−14.89*	*−18.44*	< 0.01
	(1.14)	(1.94)	(1.53)	(1.87)	(1.70)	(1.66)	
Centre-Italy	1.02	0.01	−0.01	0.46	1.50	1.56	0.34
	(0.91)	(1.41)	(1.22)	(1.41)	(1.76)	(2.05)	
South-Italy	*−3.24*	*−4.73*	*−3.70*	*−4.39*	*−3.90*	*−0.01*	< 0.01
	(0.73)	(1.36)	(1.11)	(1.36)	(2.00)	(1.25)	
Pared	*1.96*	*1.56*	*2.16*	*2.20*	*2.07*	*2.00*	< 0.01
	(0.07)	(0.13)	(0.11)	(0.08)	(0.13)	(0.21)	

In italics estimates with p < 0.05. Standard errors in parenthesis computed with non-clustered bootstrap estimation. In the last column the p value of the equivalence test (equiv.test) results

only on the conditional mean. Analogous results emerged for the other outcomes (reading—L2; mathematics—L2; reading—L5).[1]

Next, the added value of the QR model as a statistical tool for completing the regression picture resulting from the LS approach, is examined in terms of the difference between conditional quantile parameters at different points of the distribution of reading and mathematics scores, and between those parameters and mean regression coefficients. Each regression coefficient measures the change of the reading or mathematics score deriving from a one-unit increase in continuous variables (e.g. pared; ESCS) or the change from 0 to 1 of dummy variables (e.g. from male to female), fixing all the other independent variables. LS coefficients measure a change in the conditional mean while QR coefficients measure a change on a given conditional quantile.

The estimates from LS and QR models used to investigate students' performance as a function of the selected covariates are shown in Tables 2, 3 for the second grade (L2) and in Tables 4, 5 for the fifth grade (L5). The last column in each table contains the p value associated with the equivalence test results for each covariate.

To facilitate interpretation, results are also shown in Figs. 6, 7, 8, and 9. In particular, each panel represents a covariate in the model, the horizontal axes display the quantiles while the estimated effects are reported on the vertical axes. The horizontal solid line parallel to the x-axis corresponds to LS coefficient along with the 95% confidence

---

[1] Results are available upon request.

Table 3 LS and QR estimates of the effects of covariate on mathematics achievement in primary school—second grade

	LS	Mathematics—L2					
		QR					
		$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$	p value equiv.test
(Intercept)	173.69	143.49	145.74	190.36	199.13	226.49	
	(1.90)	(2.56)	(2.68)	(2.25)	(3.31)	(3.74)	
Female	−4.37	−2.85	−5.36	−5.55	−6.22	−7.21	< 0.01
	(0.63)	(1.49)	(1.46)	(0.30)	(1.41)	(1.72)	
Early enrolment	−4.11	0.03	−0.03	−4.49	−4.44	−11.93	0.15
	(2.69)	(5.91)	(2.91)	(3.15)	(3.84)	(4.81)	
Late enrolment	−8.55	−15.62	−13.66	−11.39	−3.72	−0.05	< 0.01
	(2.49)	(4.10)	(3.78)	(3.98)	(4.77)	(3.71)	
Kindergarten:yes	3.25	3.57	3.44	5.23	4.29	2.29	0.70
	(1.55)	(2.53)	(2.12)	(2.94)	(2.81)	(2.62)	
Immigrant	−11.53	−10.38	−8.01	−11.58	−15.35	−14.22	< 0.01
	(1.10)	(2.12)	(1.71)	(1.79)	(2.17)	(2.07)	
Centre-Italy	−0.25	−0.72	0.01	0.02	0.01	0.00	0.90
	(0.88)	(1.56)	(0.83)	(0.47)	(1.61)	(1.89)	
South-Italy	−5.79	−6.81	−1.92	−1.31	0.40	2.29	< 0.01
	(1.42)	(0.59)	(1.78)	(2.35)	(1.40)	(1.82)	
Pared	1.98	1.82	2.03	2.11	2.02	1.90	< 0.05
	(0.07)	(0.14)	(0.07)	(0.06)	(0.13)	(0.2)	

In Italic estimates with p < 0.05. Standard errors in parenthesis computed with non-clustered bootstrap estimation. In the last column the p value of the equivalence test (equiv.test) results

interval. Each dot is the slope coefficient for the quantile indicated on the x-axis. Therefore, the solid polygonal path represents the QR pattern estimates along with the confidence bands.

The joint inspection of the QR coefficients and the corresponding confidence bands, along with the LS confidence intervals permits an understanding of whether the effect of predictors is significantly different across the conditional distribution of pupils' performance compared to the LS estimate. Furthermore, as the median ($\theta = 0.5$), like the mean estimate, describes the central tendency of data by comparing the size and the statistical significance of the median coefficient and the other conditional quantile estimates from Tables 2, 3, 4 and 5 it is possible to assess the differential effects of predictors on students' performance. Besides, the information about whether the estimated QR effects of predictors are significantly different across the conditional distribution of performances can be gathered from the equivalence test results reported in the last column.

The LS results indicate that, on average, after controlling for kindergarten attendance and enrolment in primary school, students scores are uniquely related to gender, socioeconomic background, immigrant status and school geographical area. As for gender, the LS regression coefficients indicate that, holding other variables constant, on average females outperform males in reading performance in second and in fifth school grade; conversely, males perform better than girls in mathematical tests.

By looking at the QR results, the equivalence test shows that the effect of gender differs across quantiles in all school grades both for mathematics (L2: F = 4.64; df = 4, 75,656;

Table 4 LS and QR estimates of the effects of covariate on reading comprehension—fifth grade

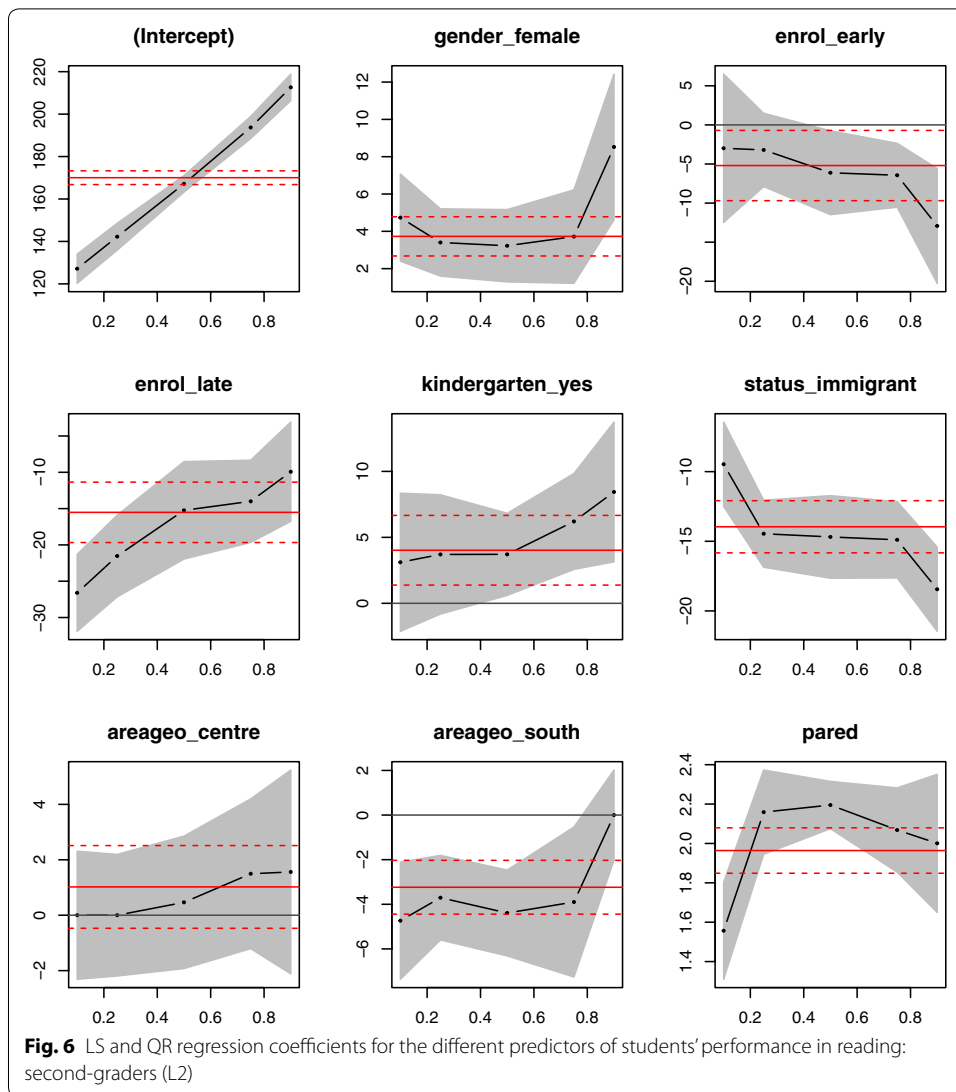|	LS	Reading comprehension—L5					
		QR					
		$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$	p value equiv.test
(Intercept)	198.06	150.92	171.60	198.76	224.45	245.66	
	(1.09)	(1.89)	(1.53)	(1.43)	(1.62)	(1.80)	
Female	*2.81*	*4.97*	*4.62*	*1.89*	*1.81*	*2.46*	< 0.01
	(0.55)	(0.89)	(0.76)	(0.63)	(0.84)	(1.02)	
Early enrolment	−0.51	0.30	−1.13	−0.36	−2.22	−4.64	0.80
	(2.60)	(3.73)	(3.46)	(3.92)	(3.85)	(5.14)	
Late enrolment	*−11.25*	*−14.39*	*−13.59*	*−12.26*	*−8.81*	*−4.39*	< 0.05
	(1.98)	(3.67)	(2.28)	(2.94)	(3.34)	(3.00)	
Kindergarten:yes	*3.08*	3.37	*5.01*	*2.71*	0.05	1.36	< 0.05
	(1.01)	(1.77)	(1.47)	(1.32)	(1.47)	(1.72)	
Immigrant	*−11.15*	*−11.09*	*−12.03*	*−11.58*	*−10.07*	*−11.15*	0.41
	(1.02)	(1.51)	(1.19)	(1.68)	(1.44)	(2.30)	
Centre-Italy	0.48	*−2.64*	0.04	0.92	1.57	1.29	< 0.05
	(0.78)	(1.08)	(1.01)	(0.91)	(1.06)	(1.25)	
South-Italy	*−5.77*	*−9.97*	*−8.55*	*−6.33*	*−3.04*	*−2.19*	< 0.01
	(0.64)	(1.17)	(1.07)	(0.75)	(0.91)	(1.07)	
ESCS	*11.19*	*11.13*	*11.28*	*11.18*	*11.49*	*11.47*	0.92
	(0.30)	(0.49)	(0.37)	(0.38)	(0.46)	(0.58)	

In Italic estimates with p < 0.05. Standard errors in parenthesis computed with non-clustered bootstrap estimation. In the last column the p value of the equivalence test (equiv.test) results

p < 0.01; L5: F = 18,77; df = 4; 95,541, p < 0.01) and reading (L2: F = 5.45; df = 4; 72,411, p < 0.01; L5: F = 4.91; df = 4, 91,936, p < 0.01). In second grade, the difference between males and females in reading performance is significantly wider at the tails of the conditional performance distribution (Table 2, Fig. 6) whereas in fifth grade, gender inequality narrows moving from the lower to the upper quantiles of the conditional distribution of students' performance (Table 4, Fig. 8). As for mathematics results, both in the second (Table 2, Fig. 7) and in the fifth grade (Table 5, Fig. 9), gender differences in favor of males tends to widen from the lower to the higher quantiles of the conditional performance distribution. With respect to the role of students' socio-economic background, LS estimates indicate that, holding other variables constant, on average the unique contribution of parents' education relative to the reading and mathematics scores in second grade is positive. In particular, the change in the conditional expectation of students' test scores resulting from the increase of 1-year of parent schooling is 1.96 for reading (Table 2) and 1.98 for mathematics (Table 3).

QR findings for the second grade indicate that the effect of parent education is significant and positive along the entire conditional distribution of performances in mathematics and reading. Moreover, the tests for equivalence of coefficients indicate that QR estimates significantly differ across the quantiles (mathematics: F = 2.41; df = 4; 75,656, p < 0.05; reading: F = 8.16 df = 4; 72,411, p < 0.01). In fact, the QR slope is rather constant across the conditional quantiles, except for a slight increase in the magnitude of

Table 5 LS and QR estimates of the effects of covariate on mathematics achievement in primary school—fifth grade

	LS	Mathematics—L5					
		QR					
		$\theta = 0.1$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 0.9$	p value equiv.test
(Intercept)	206.60	154.20	177.62	207.41	235.34	259.17	
	(1.07)	(2.05)	(1.50)	(1.56)	(1.62)	(2.14)	
Female	−7.74	−2.07	−5.91	−8.85	−10.94	−10.77	< 0.01
	(0.55)	(1.01)	(0.74)	(0.72)	(0.74)	(1.07)	
Early enrolment	−1.45	5.05	5.33	−4.13	−7.70	−5.69	< 0.01
	(2.61)	(3.63)	(2.26)	(2.25)	(2.88)	(5.44)	
Late enrolment	−13.93	−19.97	−15.44	−12.13	−9.07	−9.96	< 0.05
	(1.93)	(3.62)	(2.62)	(3.30)	(3.13)	(3.24)	
Kindergarten:yes	0.77	3.18	3.07	0.24	−1.78	−2.30	0.06
	(0.99)	(1.94)	(1.36)	(1.47)	(1.58)	(2.00)	
Non-native	−8.24	−8.57	−7.44	−9.04	−8.21	−7.05	0.55
	(1.01)	(1.66)	(1.44)	(1.35)	(1.75)	(1.82)	
Centre-Italy	−1.55	−0.79	−1.39	−1.36	−1.43	−2.75	0.80
	(0.75)	(1.27)	(0.89)	(1.01)	(1.08)	(1.52)	
South-Italy	−7.73	−9.71	−8.36	−8.59	−7.39	−5.59	< 0.05
	(0.75)	(1.19)	(0.83)	(0.72)	(0.85)	(1.18)	
ESCS	10.15	9.66	10.70	10.84	10.24	9.46	< 0.01
	(0.30)	(0.56)	(0.38)	(0.43)	(0.48)	(0.57)	

In Italic estimates with p < 0.05. Standard errors in parenthesis computed with non-clustered bootstrap estimation. In the last column the p value of the equivalence test (equiv.test) results

parents' education effect moving from the lower tail to the remaining part of the conditional performance distribution (Tables 2, 3).

In the fifth grade, mean reading and mathematics scores increase by 11.19 points and 10.15 points, respectively, with a one-unit increase of the ESCS index. As for reading, the QR results indicate that the ESCS effect is significant and homogeneous across all quantiles (test for equivalence: F = 0.23, df = 4; 91,936, p > 0.05). In mathematics, the effect of ESCS is non-monotonic across quantiles (test for equivalence: F = 3.06 df = 4; 95,541; p < 0.01) with a lower effect at the bottom ($\theta = 0.1$) and the top ($\theta = 0.9$) of the conditional distribution of the outcome compared to the mean estimate (Tables 5, Fig. 9).

Moving to results on discrepancies related to students' immigrant status, LS findings indicate that—controlling for the other selected covariates—on average immigrants show a lower performance than their non-immigrant peers in reading and mathematics tests both in second and fifth school grades.
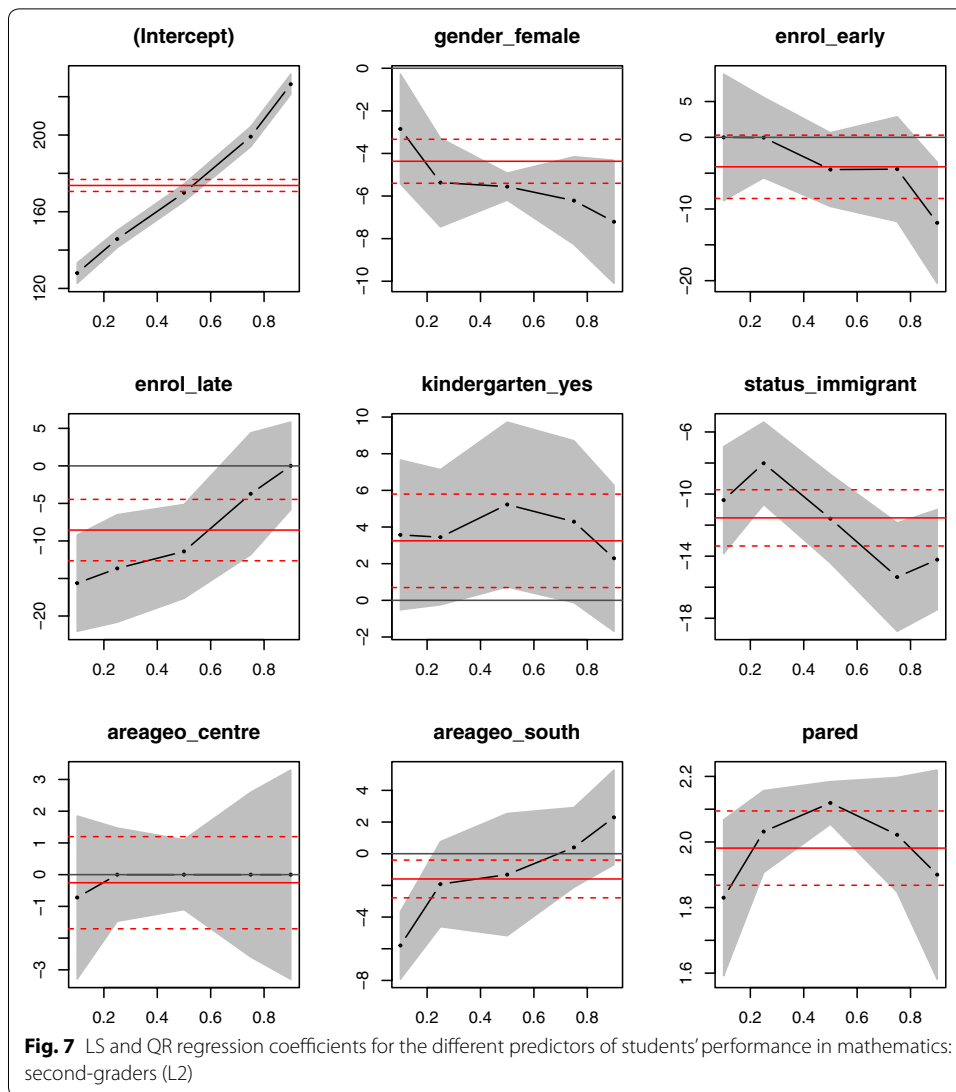
The equivalence tests indicate that the QR coefficients for immigrant status significantly differs across quantiles only in the second grade (mathematics: F = 9.21, df = 4; 75,656, p < 0.01; reading: F = 7.38, df = 4; 72,411, p < 0.01) and differences between immigrants and native students tend to increase moving from the lowest to the highest quantiles of the conditional performance distribution. On the other hand, in fifth grade, the QR coefficients are negative but the equivalence test does not approach significance level (mathematics: F = 0.75, df = 4; 95,541 p > 0.05; reading: F = 0.98, df = 4; 91,936,

Fig. 6 LS and QR regression coefficients for the different predictors of students' performance in reading: second-graders (L2)

$p > 0.05$): the QR slope is flat, meaning that inequality in the educational outcomes does not vary across quantiles.

The school geographical location exerts a significant unique effect on students' performance in reading and mathematics; in fact, the LS regression coefficients confirm that, on average, students in the southern regions obtain lower reading and mathematics test scores than students from the north. The difference in performances between students from Northern and Central Italy is statistically significant only in the mathematics test in fifth grade (Table 5, Fig. 9).

The QR results show that the intensity of the discrepancies between Northern and Southern Italy is not balanced along the outcomes distribution. In fact, the equivalence test for the north–south QR coefficients shows that the heterogeneous effect of the north–south duality on performances is statistically significant for reading and mathematics achievement in both school grades (reading—L2: $F = 8.16$, df = 4; 72,411, $p < 0.01$; mathematics—L2: $F = 9.21$, df = 4; 75,656 $p < 0.01$; reading—L5: $F = 9.70$, df = 4; 91,936 $p < 0.01$; mathematics—L5: $F = 2.50$, df = 4; 95,541 $p < 0.01$).

Fig. 7 LS and QR regression coefficients for the different predictors of students' performance in mathematics: second-graders (L2)

In the second grade, attending a school in Southern Italy is significantly and negatively associated with the lowest quantile ($\theta = 0.10$) of mathematics score distribution, while the discrepancy between Northern and Southern Italy is negligible in the remaining part of the conditional mathematical performance distribution. As for the reading performance, the estimated differences in educational outcomes between northern and southern regions are statistically significant in the lower part of the conditional performance distribution. In the fifth grade, all QR coefficients associated with the north–south discrepancy show a significantly negative effect both on reading (Table 4, Fig. 8) and mathematics achievements (Table 5, Fig. 9). However, the estimated gap tends to become less intensive moving from the lower to the upper tail of the conditional distribution of the outcomes.

Discussion and conclusions

Using INVALSI large-scale survey data on students' achievement, this paper aimed at exploring, on empirical grounds, the added value of quantile regression as a policy research tool to investigate the determinants and mechanism of inequalities in education

Fig. 8 LS and QR regression coefficients for the different predictors of students' performance in reading: fifth-graders (L5)
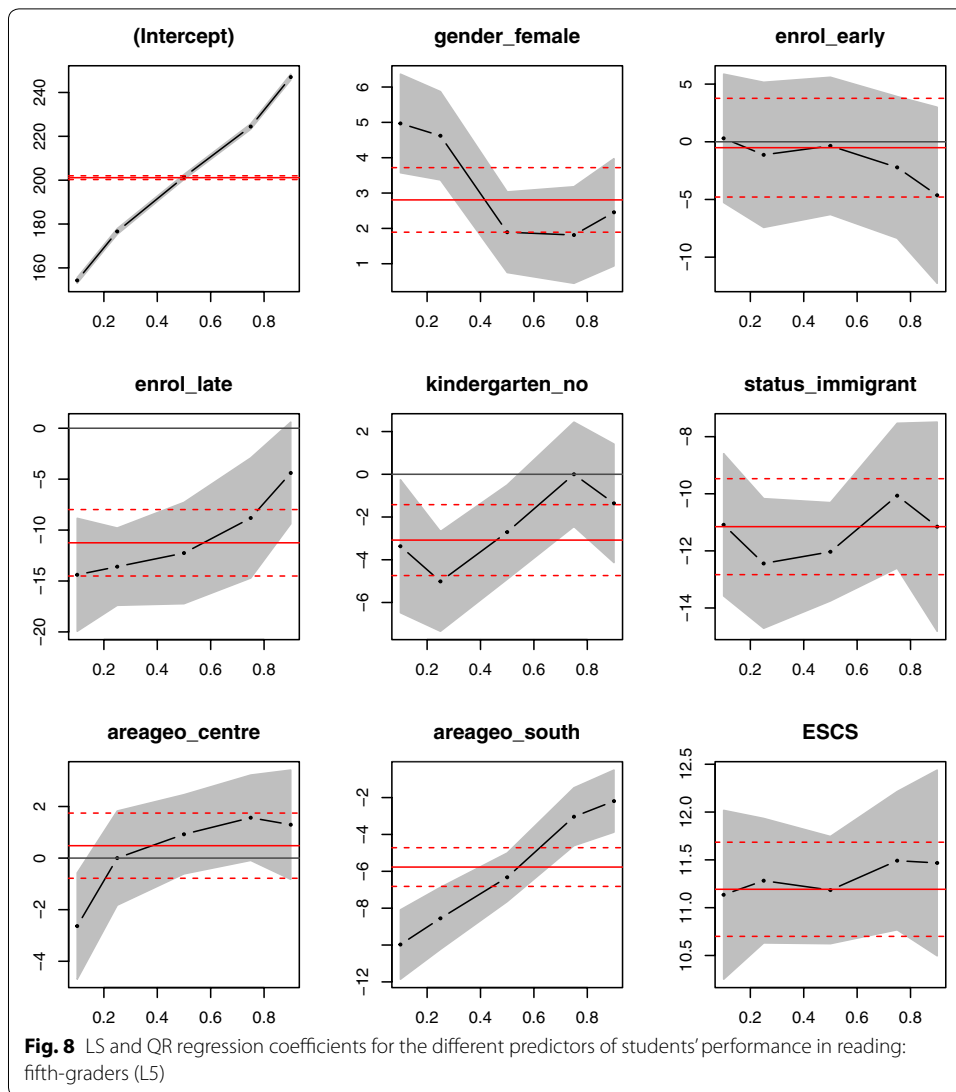
outcomes. This issue was addressed by focusing on data about primary school education in Italy.

Consistently with previous findings on data from the Italian context (Montanaro 2009; Montanaro and Sestito 2014; INVALSI 2015a, b; Gnaldi et al. 2015), LS results confirmed that—after controlling for covariates—on average inequalities by gender, immigrant status and socio-economic and cultural background in educational outcomes were statistically significant. Discrepancies in students' performances were also found with respect to the geographical location of the school (Northern Italy versus Southern Italy).

In the examined data, evidence of inequalities in primary school emerged very early; in fact, they were already found among students attending the second year of schooling. Although LS estimates provided straightforward information on whether inequalities matter on average, findings from the QR analysis suggested that they might tell only a part of the story.

Fig. 9 LS and QR regression coefficients for the different predictors of students' performance in mathematics: fifth-graders (L5)
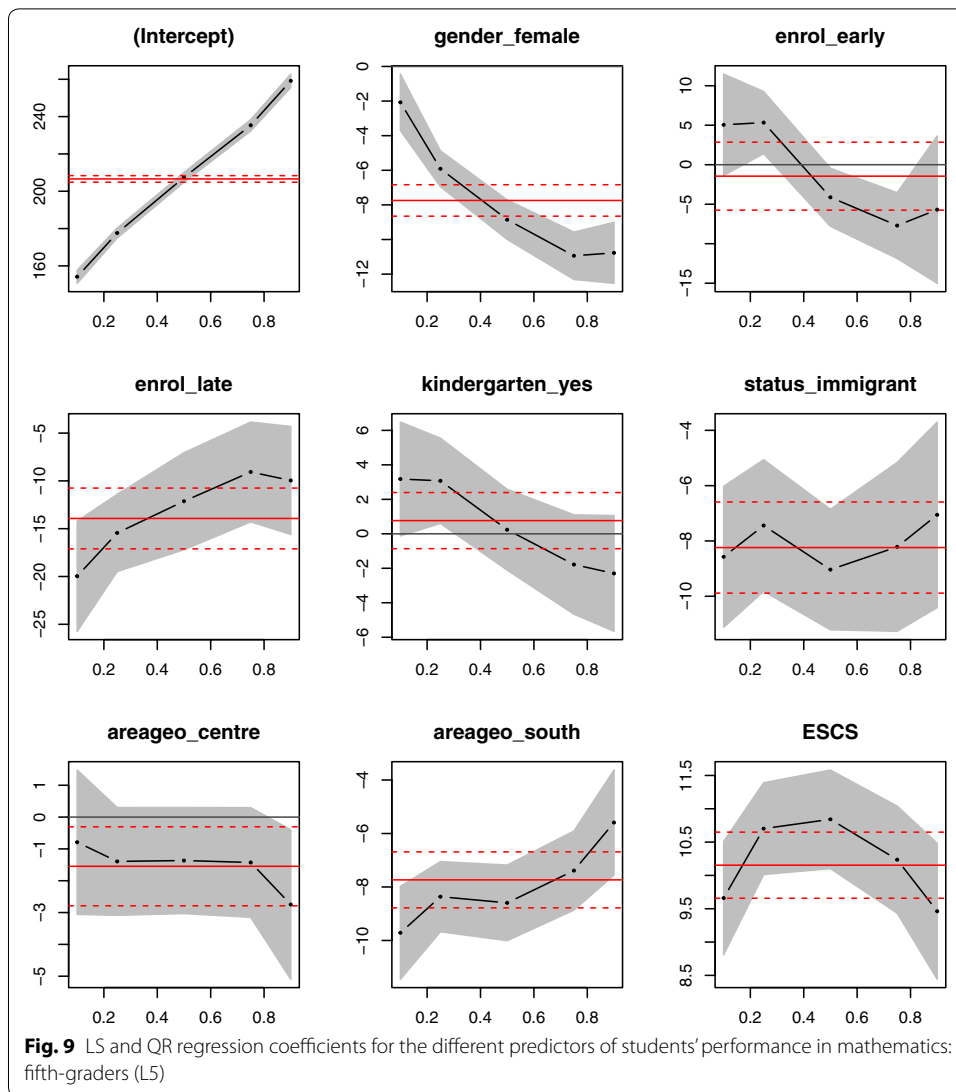
In fact, for some of the variables used in this study, heterogeneous effects emerged across the conditional distribution of pupils' achievement. This result suggests that, to expand knowledge about the inequality issue in the Italian educational system, it is even more important to assess—above and beyond the mean estimates-where and how the involved variables are effectively related with students performances.

This, in fact, is the case of gender inequalities in mathematics and reading. In mathematics, the QR results showed a steady increase of the gender gap from the lower to the upper tail of the conditional distribution of performances. Also, the gender disparities in reading were not uniform across the range of reading scores, with larger differences at the extreme tails of the distribution in second grade and in the lower tail of the distribution in fifth grade. Overall, these results confirm that gender gaps vary across the ability distribution, as observed in previous studies exploring gender differences using a quantile regression approach (Penner and Paret 2008; Robinson and Lubiensky 2011; Giambona and Porcu 2015; Contini et al. 2016) and in Baye and Monseur (2016)

meta-analysis on databases from IEA and PISA international surveys. Findings on the fifth-graders confirm that the male advantage in mathematics is more substantial at the upper tail of the ability distribution (e.g. Halpern et al. 2007; Baye and Monseur 2016) whereas the bottom of the reading achievement distribution is the part with the largest female advantage (Giambona and Porcu 2015; Baye and Monseur 2016).

As for the results on the second-graders, it is worth noting that this is one of the few studies investigating where the gender gaps are most prevalent in the reading and mathematics distribution of performances at the first years of primary school. Results in mathematics are consistent with those observed for fifth-graders; the advantage of males over females increases from the lower to the upper tail of the conditional performance distribution. Actually, this result confirms those of the relatively few studies examining the gender gap throughout the distribution of mathematics test score in early grades (e.g. Penner and Paret 2008; Robinson and Lubiensky 2011; Contini et al. 2016).

Findings on reading achievement in second grade suggest that gender gap varies, at least in part, with school grade. In accordance with the obtained results on fifth-graders and conforming to previous evidences on the upper grades of primary school (Baye and Monseur 2016), the advantage of females over males is larger among low performing students than at the median. However, the largest advantage of females emerges at the top of the reading performance distribution.

The latter result is at odd with Robinson and Lubiensky (2011) findings on United States students. The authors observed that, at the end of the early primary school grades (first and third grade), the lower quantiles of reading score distribution exhibit the largest gender differences while the difference between males and females is smaller in the upper quantiles. The discrepancy between results from the present study and those observed by Robinson and Lubiensky (2011) might be due to the difference between the degree of orthographic depth of the Italian language, that has a transparent orthography (i.e. the grapheme—phoneme correspondences are mainly one-to-one), and the English language, that has a deep orthography (i.e., several graphemes may correspond to the same phoneme and several phonemes may be represented by the same grapheme). A number of studies consistently found that the mechanisms underlying early reading skills development (word decoding, reading fluency, reading comprehension) as well as the cognitive predictors of reading acquisition vary, at least in part, with the orthographic context (e.g. Ziegler and Goswami 2005; Georgiou et al. 2008; Ziegler et al. 2010). The substantial advantage of the Italian high achieving females over the high achieving males in reading might be driven by some specific mechanisms or cognitive abilities implied in learning to read ortographies that are more transparent than the English one. Unfortunately, very few evidences are available on gender gap variability in early reading development, especially among students learning to read in transparent orthographies. More research is needed to further explore this issue and to identify the cognitive and educational factors through which gender inequality among young children in the Italian education system, as well as in other contexts, might operate.

To sum up, the QR results from this study suggest that although gender inequalities emerge for both mathematics and reading, the information behind the gender gap could be more complex and nuanced than it appears on average. In particular, it emerged that for reading, being a male is mostly associated with an increased risk of failure, whereas

in mathematics it is strongly associated with opportunities for outstanding success. This pattern of results is consistent with the Stoet and Geary (2013) statement that to understand and reduce gender inequalities, a different approach should be adopted for mathematics and reading. In line with the hypothesis, the QR results suggest that, in mathematics, the focus should be on the higher-achieving pupils. In particular, the QR estimates encourage investigation of the role of different factors which might be correlated with gender gap among top performers, i.e. the stereotype that mathematics is for boys, not for girls, or girls' mathematics self-concept. It is also worthwhile to mention that previous findings on US pupils (Cvencek et al. 2011) showed that "the math is for boys" stereotype emerged as early as second grade and influences emerging math self-concepts. Further research is needed to explore this issue in the Italian context. As for reading, the focus should be on the most vulnerable boys at the bottom of the reading performance continuum; e.g. on factors enhancing reading abilities in low performers.

Heterogeneous effects across the outcomes distribution also emerged for the geographical location of the school and the student immigrant status. The average north–south gap shown by the LS results is in line with previous findings in the Italian context based on data from international large-scale comparative studies, e.g. PISA, and on the INVALSI national data (e.g. Agasisti and Vittadini 2012; Agasisti and Cordero-Ferrara 2013; INVALSI 2015a).

The QR results showed that, after controlling for other covariates, inequalities related to the geographical location of the school are more pronounced at the bottom of the conditional distribution in both school grades. Furthermore, in second grade, the north–south gap did not reach statistical significance in the higher quantiles. These results are in line with Giambona and Porcu (2015) QR findings on older Italian students (15 year olds) and suggest that, from primary school, it is important to target programs to reduce regional inequalities in education outcomes for lower-performing pupils, who seem to be more penalized by the north–south disparities in the Italian education system.

Some insights also emerge from the QR results when addressing the gap between native and immigrant pupils in second grades. Results showed that, at least in early grades, the strength of the association between immigrant status and the outcomes steadily increase with the quantile values. However, given the relatively small number of non-native students in the examined data—especially at the tails of the conditional distribution of performances—estimates referring to the lowest and the highest quantiles should be carefully interpreted and further research is needed to explore the variability in the native–immigrant gap.

The widespread impact of family background along the entire distribution of pupils' performance is another important finding emerging from this study. Numerous researches in the field of education (Woessmann 2004; OECD 2007; Mullis et al. 2008, Gursakal et al. 2016) highlight that family background is strongly related to educational outcomes. Also, this relationship is considered by a large number of scholars as a proxy equality of opportunity (e.g. Schuetz et al. 2008). The significant impact of individual ESCS on students' performance on reading and mathematics tests emerged consistently in the Italian context (e.g. INVALSI 2015a). However, to our full knowledge, variability of the ESCS effect across the conditional performance distribution was examined by few studies. Among these, are included the secondary analysis of Giambona and Porcu

(2015) on PISA data on reading literacy and the OECD (2016) report on science assessment data. The former authors found that among Italian 15-year-old students, poor readers are the most sensitive to family background. Instead, the OECD results highlight significant differences between the impact of changes in socio-economic status on science scores at the 10th and 50th percentiles of the performance distribution and between top- and average-performing students.

A different picture emerged from the present study, based on younger pupils. In primary school grades, the children's background significantly affects reading and mathematics throughout the distribution. As for reading, a significant heterogeneity across quantiles emerged only in second grade, where the strength of the association between parent education and pupils' reading performance is slightly weaker, although statistically significant, at the bottom of the conditional reading score distribution. As for mathematics, the relationship between family background and students' performance is slightly weaker at the tails than in the remaining part of the conditional distribution. From a policy perspective, knowing that a low socio-economic status exerts a widespread negative effect on Italian students' attainment in early grades, it might be useful to focus on the reason why the basic needs of students, e.g. high-quality care, books, activities to encourage learning, remains unfulfilled.

To sum up, the main findings from this study suggest that integrating LR results with QR results provides a more nuanced view of inequalities in educational outcomes. Data exploited in the analysis show that the strength of the relationship among selected covariates and performances might change considerably at different locations representing the pupils' proficiency levels compared to the average results, e.g. gender gap or geographical areas. Therefore, QR results provide a measure of the degree of heterogeneity of the relationship between variables across the conditional performance distribution, allowing the disentanglement of inequalities that are significant and relatively homogeneous throughout the outcomes distribution (e.g. ESCS) from those which are heterogeneous (e.g. the gender gap). For the latter, the QR approach allows individualization of areas where inequalities emerge, thus providing important information on how to target programs to reduce inequalities. It is important to note that all the estimated LS and QR relationships between predictors and the conditional distribution of performances should not be interpreted as causal relations. This is, indeed, a descriptive paper because further methodological requirements should be specified to discuss the causal nature of the relationship between the selected covariates and the outcome (Pokropek 2016). However, exploring disparities in primary school permits reduction of the probability of including in the evaluation process some confounding factors which might interact with the observed outcomes, i.e. students' school track enrollment choice (e.g. technical-vocational versus academic programs). It is worth noting that this is one of the few studies examining inequalities at early grades of primary school (e.g. second-grade), using nationally representative large-scale survey data. This implies the opportunity to provide teachers and policy makers with useful information to overcome disadvantages and inequality from the earlier years of schooling.

This study and the corresponding inferences should also be interpreted in light of its limitations. Both the LS and QR models are estimated considering variables at student level, hence the multilevel structure of data (i.e. pupils within classes/school) is not

taken into account. In fact, future avenues could be to assess how class and school level variables might affect pupils' performance according to the different levels of ability by adopting a quantile multilevel regression perspective. Furthermore, this study is based on cross-sectional data. Future researches might investigate inequalities in their emerging, widening or narrowing from early grades to the end of primary school through a longitudinal study design.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Agasisti, T., & Cordero-Ferrara, J. M. (2013). Educational disparities across regions: A multilevel analysis for Italy and Spain. *Journal Policy Modelling*, *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.

Basset, G., & Koenker, R. (1982a). Test of linear hypotheses and L1 estimation. *Econometrica*, *50*, 1577–1583.

Basset, G., & Koenker, R. (1982b). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, *50*, 43–61.

Baye, A., & Monseur, C. (2016). Gender differences in variability and extreme scores in an international context. *Large-scale Assessments in Education*, *4*(1), 1. doi:10.1186/s40536-015-0015-x.

Campodifiori, E., Figura, E., Papini, M., & Ricci, R. (2008). Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia. In Working paper INVALSI. http://www.invalsi.it/download/wp/wp02_Ricci.pdf. Accessed 25 May 2017.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality*, *8*(4), 429–450.

Contini, D., Di Tommaso, M. L., & Mendolia, S. (2016). The gender gap in mathematics achievement: Evidence from Italian Data. IZA discussion paper 10053. http://ftp.iza.org/dp10053.pdf. Accessed 21 June 2017.

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math gender stereotypes in elementary school children. *Child Development*, *82*(3), 766–769.

Davino, C., Furno, M., & Vistocco, D. (2013). *Quantile regression: Theory and applications*., Series in probability and statistics New York: Wiley.

Falorsi, D. (2007). Nota metodologica sulla strategia di campionamento del sistema nazionale di valutazione delle competenze per le classi seconda e quinta del primo ciclo della scuola primaria. In Working paper INVALSI. http://www.invalsi.it/download/INVALSI_indagine_SNV_strategia.pdf. Accessed 20 June 2017.

Feingold, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist*, *50*(1), 5–13. doi:10.1037/0003-066X.50.1.5. Accessed 6 June 2017.

Ferrer-Esteban, G. (2013). Rationale and incentives for cheating in the standardised tests of the Italian assessment system. In Programma Education FGA, Working paper n. 50 (12/2013). http://www.fga.it/uploads/media/Ferrer_Esteban__Rationale_and_incentives_for_cheating_in_the_standardised_tests_of_the_Italian_assessment_system_FGA_WP50.pdf. Accessed 20 June 2017.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–40.

Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictorsof word decoding and reading fluency across languages varyingin orthographic consistency. *Journal of Educational Psychology*, *100*, 566–580.

Giambona, F., & Porcu, M. (2015). Student background determinants of reading achievement in Italy. A quantile regression analysis. *International Journal of Educational Development*, *44*, 95–107.

Gnaldi, M., Bartolucci, F., & Bacci, S. (2015). A multilevel finite mixture item response model to cluster examinees and schools. *Data Analysis Classification*. doi:10.1007/s11634-014-0196-0.

Gursakal, S., Murat, D., & Gursakal, N. (2016). Assessment of PISA 2012 results with quantile regression analysis within the context of inequality in educational opportunity. *Alphanumeric Journal*. doi:10.17093/aj.2016.4.2.5000186603.

Gutenbrunner, C., Jureckova, J., Koenker, R., & Portnoy, S. (1993). Tests of linear hypotheses based on regression rank scores. *Journal of Nonparametric Statistics*, *2*, 307–331.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51.

Hansen, K. Y., & Gustafsson, J. E. (2016). Determinants of country differences in effects of parental education on children's academic achievement. *Large-scale Assessments in Education*, *4*(1), 1.

Hao, L., & Naiman, D. (2007). *Quantile regression*., Series: Quantitative applications in the social sciences Newcastle upon Tyne: Sage.

INVALSI. (2015a). Rilevazioni nazionali sugli apprendimenti 2014–2015, National Report INVALSI. http://www.invalsi.it/invalsi/doc_evidenza/2015/034_Rapporto_Prove_INVALSI_2015.pdf. Accessed 20 June 2017.

INVALSI. (2015b). Rilevazioni nazionali sugli apprendimenti 2014–2015, Technical Report INVALSI. http://www.invalsi.it/invalsi/doc_evidenza/2015/024_Rapporto_tecnico_2015.pdf. Accessed 21 June 2017.

Koenker, R. (2005). *Quantile regression*. Cambdrige: Cambdrige University Press.

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

Montanaro, P. (2008). Learning divides across the Italian regions: Some evidence from National and International Surveys Occasional paper, 14, Bank of Italy. http://www.bancaditalia.it/pubblicazioni/qef/2008-0014/index.html?com.dotmarketing.htmlpage.language=1. Accessed 20 June 2017.

Montanaro, P. (2009). I divari regionali nell'apprendimento scolastico in Italia: Evidenze dalle indagini nazionali e internazionali, Rivista economica del Mezzogiorno, vol. XXIII, 3.

Montanaro, P., & Sestito, P. (2014). The quality of Italian education: A comparison between the International and the National Assessments. Occasional Paper, 218, Bank of Italy. http://www.bancaditalia.it/pubblicazioni/qef/2014-0218/index.html?com.dotmarketing.htmlpage.language=1. Accessed 20 June 2017.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD. (2007). *Education at a glance 2007: OECD indicators, PISA*. Paris: OECD Publishing Paris.

OECD. (2012). *Low-performing students: Why they fall behind and how To help them succeed*. Paris: PISA OECD Publishing.

OECD. (2015). *The ABC of gender equality in education: Aptitude, behavior, confidence*. Paris: PISA OECD Publishing.

OECD. (2016). PISA 2015 Results (volume I): Excellence and equity in education, PISA, OECD Publishing: Paris. doi:10.1787/9789264266490-en. Accessed 8 Aug 2017.

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, *37*(1), 239–253.

Pokropek, A. (2016). Introduction to instrumental variables and their application to large-scale assessment data. *Large-scale Assessments in Education 4*(1). doi:10.1186/s40536-016-0018-2.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Robinson, J. P., & Lubiensky, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 2268–2302.

Schuetz, G., Ursprung, H., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos*, *61*(2), 279–308.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within- and across-nation assessment of 10 years of PISA data. *PLoS ONE*, *8*(3), e57988. doi:10.1371/journal.pone.0057988.

Woessmann, L. (2004). How equal are educational opportunities? family background and student achievement in Europe and the United States. CESifo Working Paper, No. 1162.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29.

Ziegler, J. C., Bertrand, D., Tth, D., Cspe, V., Reis, A., Fasca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, *21*, 551–559.