



Testing heterogeneity in quantile regression: a multigroup approach

Cristina Davino¹ · Giuseppe Lamberti¹ · Domenico Vistocco²

Received: 1 June 2022 / Accepted: 15 May 2023 / Published online: 10 June 2023
© The Author(s) 2023

Abstract

The paper aims to introduce a multigroup approach to assess group effects in quantile regression. The procedure estimates the same regression model at different quantiles, and for different groups of observations. Such groups are defined by the levels of one or more stratification variables. The proposed approach exploits a computational procedure to test group effects. In particular, a bootstrap parametric test and a permutation test are compared through artificial data taking into account different sample sizes, and comparing their performance in detecting low, medium, and high differences among coefficients pertaining different groups. An empirical analysis on MOOC students' performance is used to show the proposal in action. The effect of the two main drivers impacting on performance, learning and engagement, is explored at different conditional quantiles, and comparing self-paced courses with instructor-paced courses, offered on the EdX platform.

Keywords Quantile regression · Multigroup approach · Testing heterogeneity

1 Introduction

Testing if a given model is different across groups of observations is relevant in many contexts. Groups' membership is usually coded in the dataset through one or more stratification variables, whose levels denote the different classes, introducing

✉ Cristina Davino
cristina.davino@unina.it

Giuseppe Lamberti
giuseppe.lamberti@unina.it

Domenico Vistocco
domenico.vistocco@unina.it

¹ Department of Economics and Statistics, University of Naples Federico II, Via Cintia 21, 80126 Naples, Italy

² Department of Political Sciences, University of Naples Federico II, Via Leopoldo Rodinò 22, 80133 Naples, Italy

an heterogeneity that must be considered in analyzing and modeling data. For example, in education, the impact of students' socioeconomic characteristics on their performance and learning achievement may differ by geographic origin (Hansen and Gustafsson 2016) or gender (Baye and Monseur 2016).

The classical approach estimates a single model on the whole dataset, introducing dummy variables to distinguish the different group effects (Gujarati 1970). Following this approach, an F -test is exploited to assess if groups are significantly different. Chow test (Chow 1960) and the Lebart test (Lebart et al. 1979) are two commonly used solutions. The two F -tests consist in comparing the restricted deviance with the unrestricted deviance, the former being related to a single model estimated on the whole sample, the latter related to a separate model for each group of observations associated with the source of heterogeneity. However, it is necessary to consider that there are two main drawbacks in pursuing such an approach. Firstly, these tests implicitly assume normality, independence, and homoscedasticity among groups. Furthermore, the specific impact of regressors is not immediate, especially in the case of multiple regression models. This unless interaction terms are included in the model, terms not always straightforward to interpret. A different approach, named multilevel modeling (Gelman 2006; Raudenbush and Bryk 2002; Snijders and Bosker 2011), estimates an additional coefficient, the between-cluster variability, to capture the hierarchical structure of the data, but is more suited to the case of a large number of groups induced by the stratification variable(s). An alternative to the analysis of group effects in statistical models has been recently introduced in the context of composite-based path modeling (Vinzi et al. 2013; Hair et al. 2016; Wold 1985). This approach exploits a multigroup perspective, separating data into segments according to the levels of the stratification variable(s), and estimating a separate model for each segment. Resampling and permutation methods are then used to test differences among the separate models (Hair et al. 2018). Since composite-based path modeling essentially consists of simple and multiple regressions aimed to estimate latent variables, the multigroup approach has been also used in the framework of ordinary least squares (OLS).

This paper extends the multigroup approach to quantile regression (QR), a class of models aiming to assess the effects of a set of explicative variables at different locations of the conditional distribution of a response variable. QR (Koenker and Bassett 1978; Davino et al. 2013; Furno and Vistocco 2018) provides a different model for each conditional quantile of interest, without introducing any parametric assumption on the response. The use of the multigroup approach in QR is promising, since it offers simple and interpretable tools to assess if and how heterogeneity impacts at different parts of the response distribution. Notwithstanding, the problems engendered from the use of separate models for each group of observations are amplified in the QR framework, since comparisons must be carried out both for models related to different groups at a given conditional quantile, and for models related to different conditional quantiles for a given group.

This paper focuses on two tests popular in composite-based path modeling, the parametric t -test (Keil et al. 2000), and the permutation test (Chin and Dibbern 2010), adapting and comparing them for the case of quantile regression. The two tests are detailed in Sect. 3, once the basic notation and the essential toolkit has

been introduced in Sect. 2. Section 4 validates the proposed multigroup approach through a study on artificial data whose design takes into account possible effects of different sample sizes, as well as the performance of the two tests in detecting low, medium, and high differences among coefficients pertaining different groups. The practical implications and the relevance of the approach is shown on real data in Sect. 5 through an empirical analysis on MOOC students' performance, one of the major challenges in learning analytics (Siemens and Long 2011). The empirical application allows to evaluate if and how the effect of learning and engagement, the two main drivers of student's performance (Carannante et al. 2020; de Barba et al. 2016; Moore and Wang 2021), changes according to the way the courses are offered, namely distinguishing self-paced courses and instructor-paced courses (Fianu et al. 2018; Goopio and Cheung 2020). Finally, a discussion on the main results, and the conclusions with some further research developments to be explored are included in Sect. 6.

2 The basic notation and the essential toolkit

The multigroup approach consists in testing statistical differences among coefficients of a given model estimated on different groups of observations. The different groups are associated with the levels of one or more stratification variables, making it possible to take into account heterogeneity corresponding to an a priori partition of the whole sample. The approach has been widely used for comparing OLS regression models, especially in the case of composite-based path modeling (Eslami et al. 2013; Hair et al. 2018). This section introduces the basic notation for extending the multigroup approach to compare QR models.

Let \mathbf{y} denote a response variable observed on $i = 1, \dots, n$ observations. Let \mathbf{X} be the matrix storing the explicative variables. Data are row-partitioned in G groups according to the levels of one or more stratification variables:

$$\begin{aligned}\mathbf{y}^T &= [\mathbf{y}_1, \dots, \mathbf{y}_J, \dots, \mathbf{y}_J, \dots, \mathbf{y}_G] \\ \mathbf{X}^T &= [\mathbf{X}_1, \dots, \mathbf{X}_J, \dots, \mathbf{X}_J, \dots, \mathbf{X}_G]\end{aligned}$$

where the subscripts I and J refer to two generic groups. Finally, let n_g be the cardinality of the generic group g , with $n = \sum_{g=1}^G n_g$. The multigroup approach allows to evaluate if the impact of the explicative variables is different across groups through a comparison of the coefficients estimated for each group. The number of possible pairwise comparisons clearly depends on the number G of groups, and on the number of coefficients included in the model, i.e. the number of explicative variables.

This paper introduces the multigroup approach to compare coefficients of QR models. QR extends classical regression to a set of quantile functions of a response variable \mathbf{y} , conditional on a set of covariates \mathbf{X} . QR, originally proposed by Koenker and Bassett (1978), is a a distribution free regression approach, since it does not pose any parametric assumption on the response distribution. It aims to estimate the effects of a set of regressors on the quantiles of a response variable. In particular, QR estimates separate models for different $\theta \in (0, 1)$, where θ denotes the conditional

quantile of interest. Unlike the classical regression model, where the conditional mean of the error $E(\epsilon | X) = 0$, in QR the θ -quantile of the error term is 0, namely $Q_\theta(\epsilon(\theta) | X) = 0$ where $Q_\theta(\cdot | \cdot)$ is the conditional quantile function. The separate models provided by QR, one for each quantile of interest, are interpretable in terms of regression models for the associated conditional quantiles of the response. The QR model for a given conditional quantile θ can be formulated as follows:

$$Q_\theta(\hat{y} | X) = \mathbf{X}\hat{\beta}(\theta). \quad (1)$$

The conditional quantile estimator minimizes the sum of absolute deviations, asymmetrically weighting positive and negative residuals. The bootstrap procedure is typically used for inference (Koenker et al. 2017), so to avoid the assumptions required by asymptotic theory. Bootstrap offers the flexibility to obtain standard errors and confidence intervals for any estimates and combinations of estimates, keeping the distribution free nature of QR. The reader interested in the machinery of QR can refer to Koenker and Bassett (1978), Davino et al. (2013), Furno and Vistocco (2018) for details.

Since QR provides separate models for each conditional quantile, the number of possible pairwise comparisons in the multigroup approach is amplified. The multigroup approach tests the null hypothesis $H_0 : \beta_I(\theta) = \beta_J(\theta)$ versus the presence of a significant difference, $H_1 : \beta_I(\theta) \neq \beta_J(\theta)$, where $\beta_I(\theta)$ and $\beta_J(\theta)$ are the coefficients of a generic explicative variable related to group I and group J , respectively. As has been pointed out above, the parametric t -test and the permutation test are the two most widely used tests in such framework. The parametric t -test (Keil et al. 2000) exploits a bootstrap resampling procedure to compare coefficients. In particular, the bootstrap samples are used to approximate the sampling distributions of the coefficient estimators for each segment, providing a reasonable solution if sample size is sufficiently large. Instead, the permutation test (Chin and Dibbern 2010) evaluates differences between the coefficients of different segments through a permutation procedure: data of the two groups are permuted preserving the initial group sizes to obtain the sampling permutation distribution of the difference between groups. The difference observed on the two samples is then compared with such permutation distribution to test the null hypothesis of no significant differences between groups. Details on the two tests adapted to the QR context are offered in Sect. 3.

3 The multigroup approach for quantile regression

Starting from the QR model (1), the aim is to assess if units corresponding to two different groups I and J share the same dependence structure, namely if the impact of the regressors on the response is different with respect to different segments of units. Moreover, since the QR model provides estimates at different conditional quantiles, the comparison can be carried out also for different locations. Starting from two QR models estimated for group I and J , respectively, the multigroup approach tests if the observed difference between coefficients $\hat{\beta}_I(\theta)$ and $\hat{\beta}_J(\theta)$ is significant for a given conditional quantile θ . The two following subsections extend to QR the two main

tests proposed in multigroup literature. To simplify the notation, in the following we refer to a generic $\beta(\theta)$ coefficient for a given conditional quantile. Clearly, the tests can be carried out for any of $\beta_p(\theta)$ coefficient in case of multiple regression, as shown in the application in Sect. 5. The null and alternative hypotheses can be defined as follow for both the tests:

$$\begin{aligned} H_0 &: \beta_I(\theta) = \beta_J(\theta) \\ H_1 &: \beta_I(\theta) \neq \beta_J(\theta). \end{aligned}$$

Furthermore, following the QR logic, comparisons for different conditional quantiles are carried out separately for each quantile θ of interest.

3.1 The parametric test

The parametric t -test (Keil et al. 2000) combines a classical test statistic for comparing the means of two groups, with a bootstrap procedure used to estimate its standard errors. Bootstrap (Efron and Tibshirani 1998) does not pose distributional assumptions but entails a higher computational cost with respect to the standard parametric procedure. Computational costs can be reduced using newest efficient bootstrapping methods (Kleiner et al. 2014; Sengupta et al. 2016). Bootstrap estimates are unbiased, even if they introduce additional sources of variability in the process, i.e. sample variability, being based on one single sample from a given population, and resampling variability, exploiting a finite number of replications (Davino et al. 2013).

To our ends, the resampling procedure is simultaneously applied to the vector of response variable and to the matrix of regressors, resampling with replacement B times, separately for the two groups I and J , holding fixed the cardinality n_I and n_J of the original groups. This paired resampling allows to preserve the dependence structure among the variables.¹

The QR model (1) is estimated separately for group I and J , for each bootstrap sample and for each conditional quantile of interest. Therefore, a vector $\hat{\beta}_{boot}(\theta) = [\hat{\beta}^1(\theta), \hat{\beta}^2(\theta), \dots, \hat{\beta}^B(\theta)]$ is obtained in correspondence of each coefficient and each quantile of interest. Such vectors are used to estimate the standard errors of the coefficients $SE_{\hat{\beta}_{boot}}(\theta)$, exploited in the usual t -test statistic for comparing the means of the two groups:

$$t(\theta) = \frac{\hat{\beta}_I(\theta) - \hat{\beta}_J(\theta)}{\sqrt{\frac{(n_I-1)}{n_I} SE_{\hat{\beta}_{boot}}^2(\theta) + \frac{(n_J-1)}{n_J} SE_{\hat{\beta}_{boot}}^2(\theta)}}.$$

¹ In the QR literature this procedure is called xy-pair method (Kocherginsky et al. 2005) and it is used to estimate parameter standard errors without requiring any assumption on the error distribution.

The statistic is asymptotically t -distributed and the degrees of freedom (df) are determined by means of the Welch–Satterthwaite equation. The formula is derived following Sarstedt et al. (2011):

$$df = \left\| \frac{\left(\frac{(n_I-1)SE^2_{\hat{\beta}_{I_{boot}}(\theta)} + (n_J-1)SE^2_{\hat{\beta}_{J_{boot}}(\theta)} \right)^2}{\frac{(n_I-1)SE^4_{\hat{\beta}_{I_{boot}}(\theta)} + (n_J-1)SE^4_{\hat{\beta}_{J_{boot}}(\theta)}}{n_I^2 + n_J^2}} - 2 \right\|. \tag{2}$$

Algorithm 1 outlines the steps for carrying out the test.

Algorithm 1 The parametric t -test procedure

- 1: sample B times with replacement from $[\mathbf{y}_I \mid \mathbf{X}_I]$, using a sample size of n_I
- 2: sample B times with replacement from $[\mathbf{y}_J \mid \mathbf{X}_J]$, using a sample size of n_J
- 3: for each conditional quantile θ of interest:
 - ✓ estimate the QR model (1) so to obtain the vectors of coefficients in group I and J on the B bootstrap samples:

$$\begin{aligned} \hat{\beta}_{I_{boot}}(\theta) &= [\hat{\beta}_I^{(1)}(\theta), \hat{\beta}_I^{(2)}(\theta), \dots, \hat{\beta}_I^{(B)}(\theta)] \\ \hat{\beta}_{J_{boot}}(\theta) &= [\hat{\beta}_J^{(1)}(\theta), \hat{\beta}_J^{(2)}(\theta), \dots, \hat{\beta}_J^{(B)}(\theta)] \end{aligned}$$

- ✓ calculate the standard errors: $SE_{\hat{\beta}_{I_{boot}}}(\theta)$ and $SE_{\hat{\beta}_{J_{boot}}}(\theta)$
- 4: test the null $H_0 : \beta_I(\theta) = \beta_J(\theta)$ using the following t -statistic:

$$t(\theta) = \frac{\hat{\beta}_I(\theta) - \hat{\beta}_J(\theta)}{\sqrt{\frac{(n_I-1)SE^2_{\hat{\beta}_{I_{boot}}}(\theta) + (n_J-1)SE^2_{\hat{\beta}_{J_{boot}}}(\theta)}{n_I + n_J}}}$$

3.2 The permutation test

The permutation test (Chin and Dibbern 2010) evaluates the differences between the coefficients of two groups exploiting a permutation procedure. The original difference $DIF = \|\hat{\beta}_I(\theta) - \hat{\beta}_J(\theta)\|$ between the coefficients estimated on the two groups I and J at a given conditional quantile θ is compared with the permutation distribution of the differences, $PDIF$, computed on P permuted samples. The actual aim is to appraise how extreme DIF is under the null hypothesis. The two samples $[\mathbf{y}_I \mid \mathbf{X}_I]$ and $[\mathbf{y}_J \mid \mathbf{X}_J]$, corresponding to the two groups, are merged. Then, P samples of size n_I and n_J are obtained permuting data. The assumption of exchangeability holds, and the units can be rearranged without substantially altering the process under H_0 . That is done by randomly assigning these units to one group or the other, preserving the cardinalities of the two initial segments. Starting from the two new groups of units, the permutation difference is

computed: $PDIF^{(i)} = \left\| \hat{\beta}_{I_{perm}}(\theta) - \hat{\beta}_{J_{perm}}(\theta) \right\|$ (for $i = 1, \dots, P$). Fixing a quantile of interest, model (1) is estimated after each permutation, and the differences between coefficients are calculated. Finally, the vector **PDIF** of permuted differences is compared with the original differences, computing p -value as 1 minus the proportion of times in which the original difference is larger than the permuted one. The steps of the permutation test procedure are summarized in Algorithm 2.

Algorithm 2 The permutation test procedure

- 1: For each conditional quantile θ of interest:
- 2: compute the original difference (DIF) between the coefficients of model (1) estimated on groups I and J :

$$DIF = \left\| \hat{\beta}_I(\theta) - \hat{\beta}_J(\theta) \right\|$$

- 3: merge the two samples $[\mathbf{y}_I \mid \mathbf{X}_I]$ and $[\mathbf{y}_J \mid \mathbf{X}_J]$ in

$$\begin{bmatrix} \mathbf{y}_I & \mid & \mathbf{X}_I \\ \mathbf{y}_J & \mid & \mathbf{X}_J \end{bmatrix}$$

- 4: for each conditional quantile θ of interest:
 - ✓ permute data in two groups $[\mathbf{y}_{I_{perm}} \mid \mathbf{X}_{I_{perm}}]$ and $[\mathbf{y}_{J_{perm}} \mid \mathbf{X}_{J_{perm}}]$ of cardinality n_I and n_J respectively.
 - ✓ estimate the coefficients $\hat{\beta}_{I_{perm}}(\theta)$ and $\hat{\beta}_{J_{perm}}(\theta)$ on permuted data
 - ✓ compute the difference between the coefficients:

$$PDIF^{(i)} = \left\| \hat{\beta}_{I_{perm}}(\theta) - \hat{\beta}_{J_{perm}}(\theta) \right\| \quad (i = 1, \dots, P)$$

- ✓ store the resulting difference in the vector **PDIF** of permuted differences:

$$\mathbf{PDIF} = [PDIF^{(1)}, PDIF^{(2)}, \dots, PDIF^{(P)}]$$

- 5: compute the p -value comparing DIF with **PDIF**:

$$p\text{-value} = 1 - \frac{\#[DIF > \mathbf{PDIF}]}{P}$$

Following Hair et al. (2012), to guarantee the stability of the results, should be convenient to maintain the number of permutations large (500 or 1000 permutations are commonly employed thresholds (Kherad-Pajouh and Renaud 2010)). However, depending also on the number of comparisons, a larger number of permutations could produce an increase in the computation time.

4 Simulation study

This section presents a simulation study aimed to show the proposed quantile multigroup approach in action, and to compare the parametric and the permutation test varying sample size, and evaluating the ability to capture group differences of different magnitude². We focus here on the case of one regressor and two groups, I and J . This setting represents a first step in investigating the behavior of the two tests in the context of quantile regression. For the considered simple regression model, the focus is on possible effects of the quantile of interest, the sample size and the separability of the groups. The introduction of additional variables would require an extremely complicated simulation design, considering both the case of independent and correlated predictors. Indeed, the problem of multicollinearity must be considered also for QR (Davino et al. 2022). Therefore, a study that considers multiple correlated regressor is postponed to a future work. Besides, just considering a multiple regression model under the assumption of uncorrelated regressors would not be relevant to any issues for a first investigation on the response performance of the two tests regarding sample size, quantile of interest, and difference between coefficients. Furthermore, in order to study the asymptotic distributions in case of multiple QR, we plan to compare classical QR with composite QR (Zou and Yuan 2008).

The data generating process exploits a uniform distribution $U(a = 0, b = 4)$ for the regressor, generated independently for the two segments. Error terms, ϵ_I and ϵ_J , were generated exploiting a normal standardized distribution, a skew normal distribution with shape parameter α equal to 4, and a uniform distribution $U(a = 0, b = 4)$, assuming homoscedasticity within the two subgroups. This allowed us to consider three different scenarios of normal, asymmetrical, and non normal residuals. Finally, since it is well known that QR is very useful in the case of heteroscedasticity, the error component was introduced in the model through a multiplicative term with the regressor. Different scenarios have been considered in order to assess the effect of the sample size, the degree of overlapping between the groups (namely, magnitude of the difference between the coefficients of the two segments), and the quantile of interest. More specifically, the design of the simulation study considers the following factors and levels:

Sample size The effect played by the cardinality of each group is explored by hypothesizing both equal group sizes {50, 100, 250, and 500}, and unbalanced segments {75 vs 25, 150 vs 50, 375 vs 125, and 750 vs 250}.

Difference between coefficients Following Lamberti et al. (2016a, b), this study considers the case of no difference between the coefficients estimated in groups I and J , and the cases of small, medium, or large differences. Table 1 shows these four levels specifying the size of the differences (second column) and the value of the coefficients in the two groups (last two columns) used to simulate data.

Quantile of interest The sensitivity of the tests is evaluated for the QR model estimated at $\theta = \{0.1, 0.25, 0.50, 0.75, 0.9\}$.

² All computations and simulations were developed using the statistical software R (Team 2002), along with the `quantreg` package (Koenker 2022). The source code is available upon request.

Table 1 Different profiles according to the coefficient differences between group I and J

Difference between coefficients		$\beta_I(\theta)$	$\beta_J(\theta)$
Equal	(0)	0.6	0.6
Small	(± 0.2)	0.7	0.5
Medium	(± 0.4)	0.8	0.4
Large	(± 0.6)	0.9	0.3

Error term distribution Error terms were generated exploiting a normal standardized distribution, a skew normal distribution with shape parameter α equal to 4, and a uniform distribution $U(a = 0, b = 4)$.

Considering all possible combinations of sample sizes (balanced and unbalanced), differences between coefficients, quantile of interest, and error term distribution, 480 scenarios ($2 \times 4 \times 4 \times 5 \times 3$) results.. Concerning the number of permutations, we employed 100 permutations to control the computation time.

In order to show the different degrees of separability between the segments, Fig. 1 provides the graphical visualizations of the models estimated for both groups (I and J) for the case of two balanced sample of 500 observations and normal errors. Each panel refers to a given coefficient difference (equal, small, medium, large, top-left by row). Different colors/grey levels distinguish the two segments, while different lines are used for each of the five considered QR models. Analyzing the graphs starting from the top-left, it is evident that the effect of the regressor on both the mean and the conditional quantiles of y is the same in the two segments in the case of equal differences, and becomes increasingly different moving towards large differences. Therefore, it becomes important to extend the comparison between segments to parts other than the conditional mean of the dependent variable.

The performance of the parametric and permutation tests is evaluated through the p -values obtained on 100 replications for each experimental condition.

Figures 2, 3, and 4 show the results obtained using the two tests with unbalanced samples, and considering the different quantiles of interest for the three different scenarios (normal, asymmetrical normal, and not normal errors, respectively). The results with balanced samples are almost similar and therefore not shown. Each panel reports the boxplots of the p -values (vertical axis) for both the parametric test (first four columns) and the permutation test (last four columns), in case of small, medium, and large differences between coefficients (horizontal axis). The case of no differences between the coefficients of the two groups is reported in Appendix. The rows of each graph refer to the different quantile of interest (0.1, 0.25, 0.5, 0.75, and 0.9, from top to bottom), the columns to the different sizes considered for the unbalanced groups. A dotted horizontal line is drawn at the conventional significance level $\alpha = 0.05$. In all the cases, the parametric test provides a better performance than the permutation test. More generally, the effect of sample size is evident since the tests detect medium and large differences for smaller samples (100 observations). When sample size increases, the sensitivity of the tests improves, even in presence of small differences between the coefficients. Furthermore, the two tests show a good performance even in the case of extreme quantiles. Results do not substantially differ

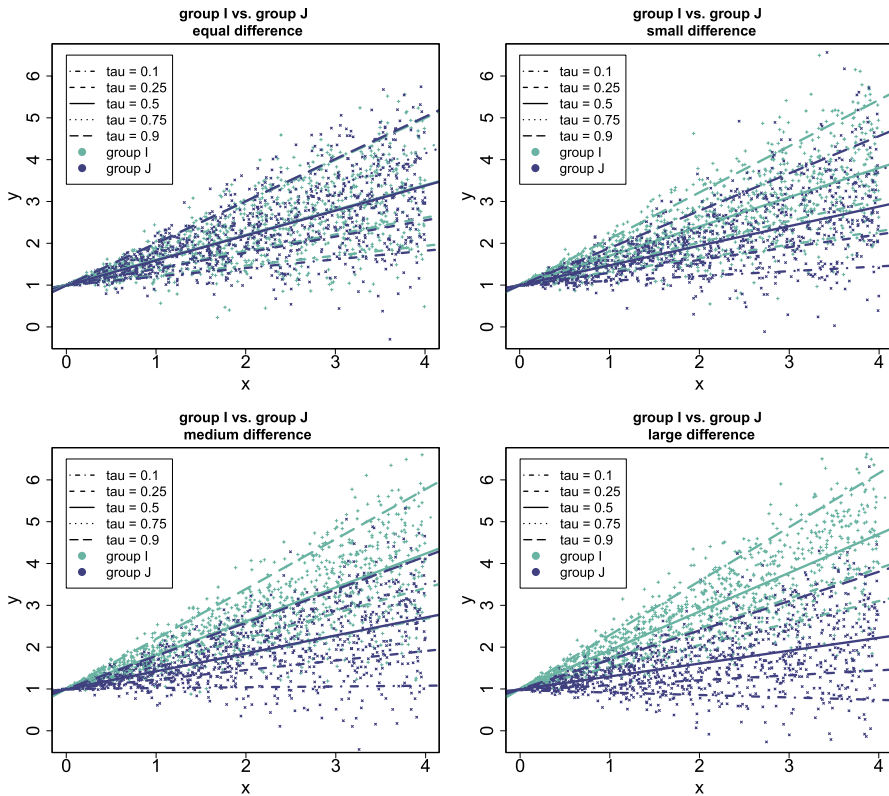


Fig. 1 Two artificial datasets of 500 observations, with normal errors. Different panels refers to the different magnitude of the difference between coefficients: equal, small, medium, and large, top-left, clockwise. The different lines correspond to the five QR models ($\theta = 0.1, 0.25, 0.5, 0.75,$ and 0.9)

in case of normal errors (Fig. 2), asymmetrical normal errors (Fig. 3), and uniform errors (Fig. 4). Finally, results suggest that at least one of the two samples must have an adequate size to identify also small differences.

5 A real data analysis

The proposed multigroup approach in QR is shown in action through a case study based on real data. The aim is to model students' performance in a particular type of course, the Massive Open Online Courses, also known as MOOCs. The data and model have recently been published by Carannante et al. (2020).

MOOCs are an increasingly common type of course in education, especially in higher education. The structure and delivery of such courses has a strong impact on the way students attend MOOCs and, inevitably, on their final performance. In the learning analytics framework (Siemens and Long 2011), predicting students' performance in MOOCs can be considered one of the main challenges. Indeed, there are

several elements affecting students' performance. Some are specifically related to the learning experience (student motivation, learning attitude, engagement), others can be defined as external, being related to personal characteristics of the student or to the specific features of the course.

In this study, we considered two main drivers of students' performance related to the learning experience: the learning attitude and the students' involvement in the planned MOOC activities (engagement), and one external factor describing the course type. In particular, the aim is to analyse if and how much the effects of learning and engagement on students' performance vary according to the course type. To this end, QR allows us to investigate whether the effect of this relationship varies for low, medium, or high performing students, and the proposed multigroup approach allows us to assess whether these effects vary by course type.

5.1 Data and measurements

Data refer to 3578 students who attended two courses in Political Science on the FedericaX platform, the EdX MOOC platform of the "Federica WebLearning" Center at University of Naples Federico II.³ Each course was offered in two versions: an instructor-paced version and a self-paced version. The instructor-paced course is strictly scheduled, with specific dates for assignments, course materials, exams, and a deadline for learners to complete the course and get a certification. Usually, this modality is integrated into an in-site course delivered in blended mode. Instead, the self-paced version provides all course materials as soon as the course starts, assignments and exams do not have due dates, and therefore a learner can progress through the course at its own speed and pass grade in the course, even without completing all of the course materials. Of the 3578 students, 73.1% followed in the self-paced modality, and 26.9% in instructor-paced modality.

The considered model uses performance as response variable, and learning and engagement as explicative variable. This in line with the model proposed in Carannante et al. (2020). Performance was measured as the proportion of correct answers to a set of questions. Learning was quantified by considering the quantity of actions undertaken to acquire knowledge. In particular, we exploit three dimensions for its measurement: frequency-based actions (count of activities spent studying), time-based actions (duration of time spent studying) and interactions (discussion on forums and social learning aspects). Engagement was analysed through two sub-dimensions: regularity (how a learner spends her/his time on the platform and how she/he organizes the learning road map), and no-procrastination (the ability of the learner in organizing the learning processes). For more details, consult Carannante et al. (2020), de Barba et al. (2016), Moore and Wang (2021).

Figure 5 shows the presence of a severe left-hand skewness both for the response and the regressors. In particular, the strong asymmetry of the response could lead to problems in using OLS, keeping in mind the classical assumptions of such a model.

³ <https://www.federica.eu>.

Fig. 2 Boxplots of p -values obtained on 100 replications of the parametric and permutation test. Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250). Each boxplot describes a different degree of overlapping among groups (small, medium, and large difference between coefficients). Simulations consider standard normal errors and unbalanced segments

A comparison of the response distribution between the two groups of students by course type is offered in Fig. 6 through violin plots (Hintze and Nelson 1998). They are a combination of a boxplot and a density plot, realized rotating and placing symmetrically on each side two density plots. The length of the horizontal axis allows to appreciate the range of the observed values, while the shape highlights how values are distributed in terms of variability and skewness. In particular, Fig. 6 reveals differences especially in the right tail of the distribution, showing a larger concentration of higher performance students in the group of instructor-based course.

5.2 Main results

The effect of engagement and learning on students' performance is explored comparing the results of classical regression and quantile regression both on the whole sample and on the two subgroups defined considering the course type. Results are reported in Table 2 and graphically summarized in Fig. 7. The first key to interpret such results lies in the comparison of OLS and QR coefficients estimated on the whole sample (third column of Table 2). Then, it is important to compare results obtained on students attending instructor-paced courses (fourth column) with students attending self-paced courses (fifth column). In all the three cases above, it is important to assess the sign and size of the coefficients but also their significance. The multigroup analysis then highlights any differences between the groups, for each regressor and for each model (sixth and seventh column of Table 2 report p -values for the parametric test and the permutation test, respectively).

The comparison of OLS and QR results on the whole sample shows different effects of engagement and learning on students' performance, although always with a positive sign. Considering the three conditional quantiles, this effect is increasing for learning (0.245, 0.420, 0.886, respectively), and particularly differentiated in the tails of the distribution when compared to the effect on the conditional mean, which is equal to 0.734. Regarding engagement, we have an inverse trend. The effect is decreasing (0.194, 0.183, 0.082), and not very dissimilar from the conditional mean, which is equal to 0.169. All coefficients are significant except the engagement estimated in correspondence of the quantile 0.75. From a practical perspective, this means that on the lower performing students, engagement plays an important role, at least more important than it does for the high performing students.

To analyse possible differences regarding course type, the same model is estimated separately on the two subgroups of students. The learning component has always a greater effect on performance than engagement, even if the impact is stronger among students who have chosen an instructor-paced course, and who perform better. Most of the coefficients are significant, especially at the considered extreme quantiles, which are very important for the practical use of the study.

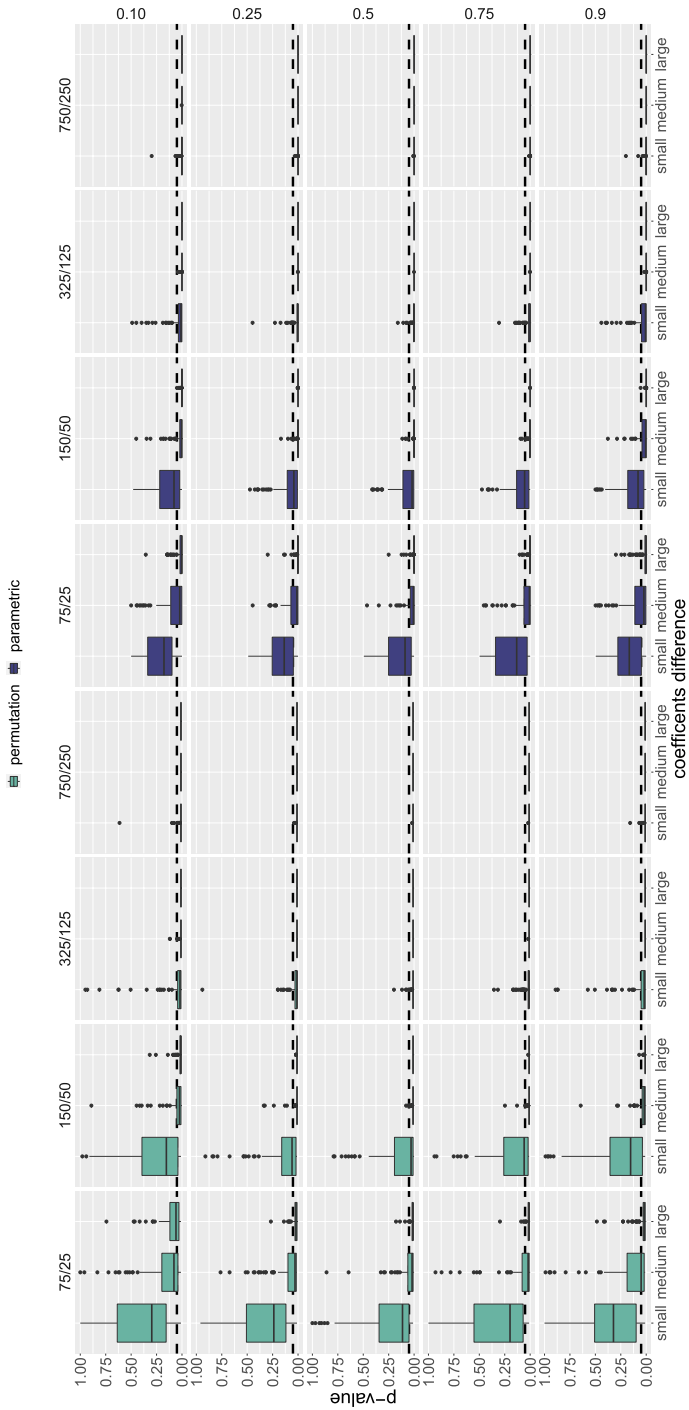


Fig. 3 Boxplots of p -values obtained on 100 replications of the parametric and permutation test. Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250). Each boxplot describes a different degree of overlapping among groups (small, medium, and large difference between coefficients). Simulations consider asymmetrical normal errors and unbalanced segments

Coefficients are compared by using both parametric and permutation test employing 500 permutations. The p -values reported in the last two columns of Table 2 highlight how QR can usefully complement OLS results. As an example, considering the effect of learning, OLS coefficients are affected by the high asymmetry of performance and hide a significant difference between the two groups which emerges looking at the QR results at the quantile 0.5.

Figure 7 provides an overview of the most important results, allowing a more immediate assessment. The two plots on the left column refer to the learning dimension, the two plots on the right column to engagement. The first row summarizes results of the parametric test, the second row results of the permutation tests. The coefficients are depicted on the vertical axis, the considered conditional quantiles on the horizontal axis. QR results on the whole sample is represented through solid lines, while results on specific groups using dashed lines. OLS results are depicted through three dots: the cross-shaped dots correspond to the coefficients estimated on the whole sample, triangle-shaped points correspond to coefficients estimated for the two groups of students, triangle with the vertex at the bottom refer to self-paced courses, while square-shaped dot to instructor-paced courses. Filled points indicate significant coefficients (p -value < 0.05) according to the permutation or the parametric tests. The four panels exploit a common scale, so as to make possible to visually appreciate the predominant impact of the learning component compared to the engagement component, pattern that becomes more relevant moving from lower to higher quantiles. However, both the t -test and the permutation test show that there is no significant difference between the two groups in the effect of learning on performance in the lowest 25% of students. The opposite occurs at the top of the performance distribution. The lower performing students in the two groups differ more with respect to engagement and, as already pointed out, engagement is more important among lower performing students that attend self-paced course.

6 Concluding remarks

Modelling is not an easy task both because of the possible complexity of the relationships among the components of the phenomenon under investigation, and because of the presence of possible heterogeneity in the dependency relationship. In some cases, this heterogeneity is known and defined through one or more stratification variables that identify groups of observations, each requiring different modelling. Obviously, for such different models it is necessary to test differences between

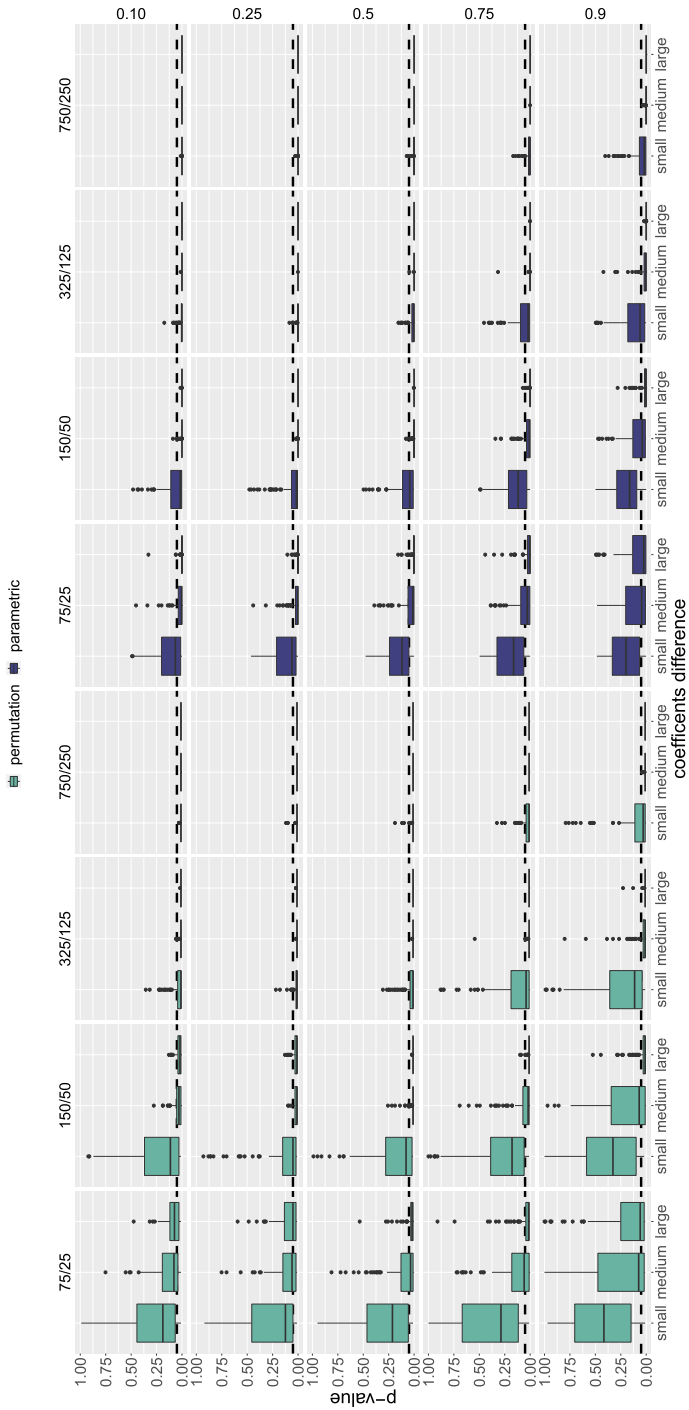


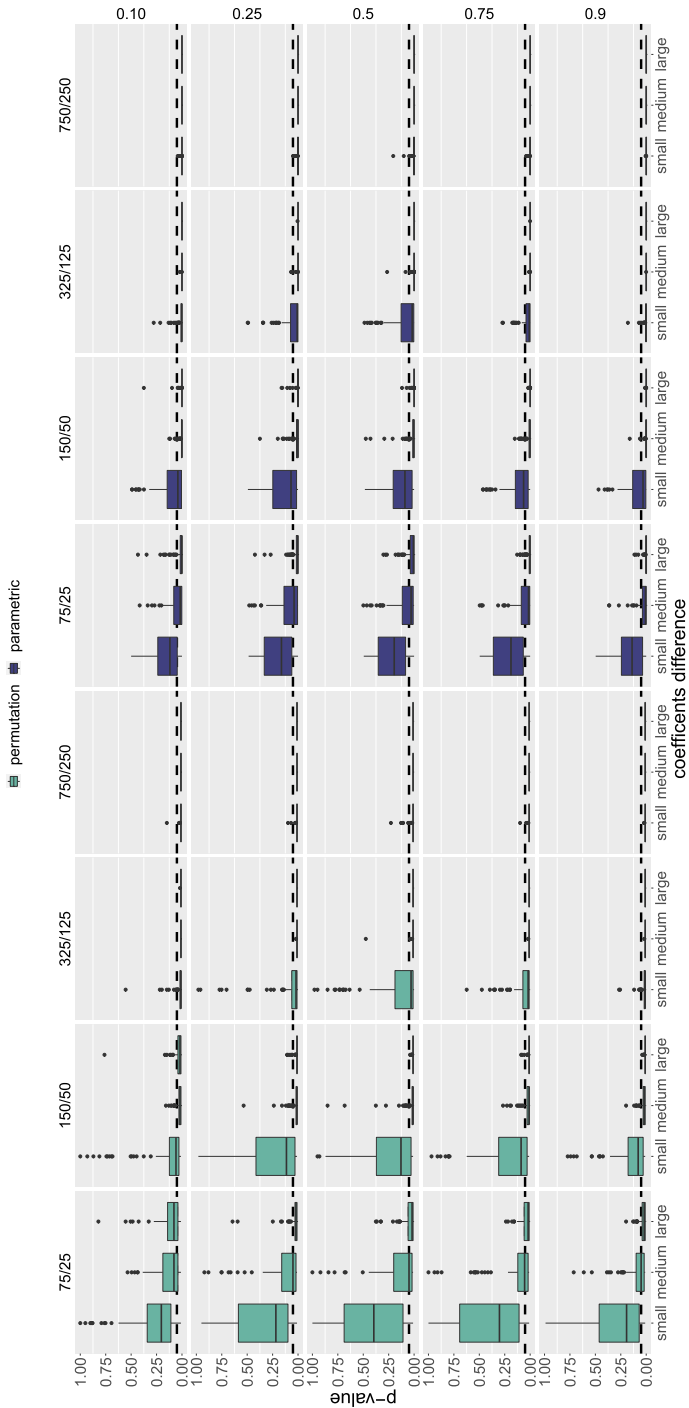
Fig. 4 Boxplots of p -values obtained on 100 replications of the parametric and permutation test. Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250). Each boxplot describes a different degree of overlapping among groups (small, medium, and large difference between coefficients). Simulations consider not normal errors and unbalanced segments

the dependence structure. The problem is amplified in case of QR, because we do not have a unique model, but separate models for each conditional quantile θ of interest. The challenge becomes larger with the analysis of group effects on different parts of the conditional distribution of the dependent variable.

In this paper, we extend the traditional approaches for handling heterogeneity from OLS regression to QR. In particular, we focus on two tests popular in the literature for multigroup analysis, introduced in composite-path modeling: the parametric t -test (Keil et al. 2000), and the permutation test (Chin and Dibbern 2010). The aim is to assess whether the presence of heterogeneity in the sample involves different effects at different parts of the conditional distribution of the response. Multigroup approach provides results easy to interpret, even if the number of comparisons is strictly related to the number of levels of the stratification variable. Indeed, such comparisons grow exponentially with the number of levels. Obviously, in case of more stratification variables, the number of comparisons is equal to the number of possible combinations of the levels of all the involved variables. Thus, we recommend to use the multigroup approach in case of stratification variables with a reduced number of levels.

The effectiveness of the parametric t -test and the permutation test has been illustrated through a real data application. Empirical analysis of MOOC students' performance showed that both engagement and learning are important drivers to explain the final performance. However, the effect of these variables is not uniform, but varies according to the conditional quantile of interest, and to the different ways in which the course is offered (self-paced and instructor-paced). Simulations confirm the ability of the proposed quantile multigroup approach in detecting differences in models. As expected, the sensitivity of both the considered tests was guaranteed by larger sample size and clearer differences between groups.

The multigroup approach focuses on observed heterogeneity, namely when the groups are defined a priori by the levels of one or more stratification variables. However, there are situations in which an a priori segmentation of the observations is not available. In such a case, we need to find some criterion to identify the comparisons to consider. The *pathmox* tree (Lamberti et al. 2016a) and *MOB* procedure (Zeileis et al. 2008) are possible procedures to identify the most important comparisons through the use of a recursive approach. They exploit multiple comparisons, and ranks the variables that produce differences in the model coefficients revealing the most significant comparisons. We postpone the feasibility of a recursive approach to the context of quantile regression to a future study.



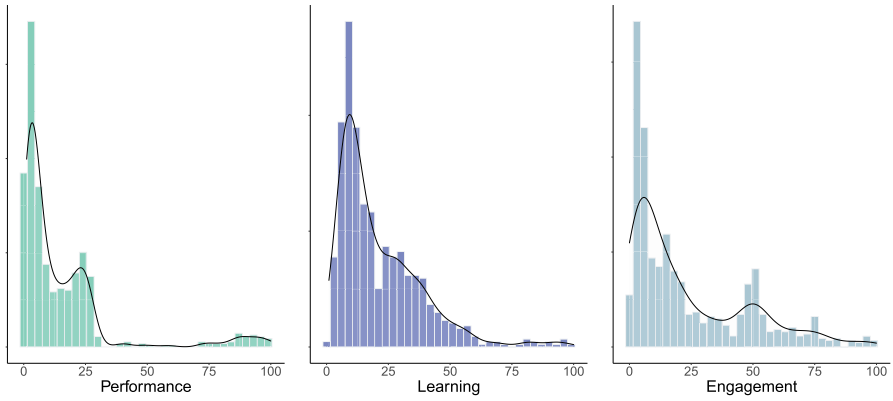


Fig. 5 Histogram and density plot for student's performance (left), engagement (middle) and learning (right)

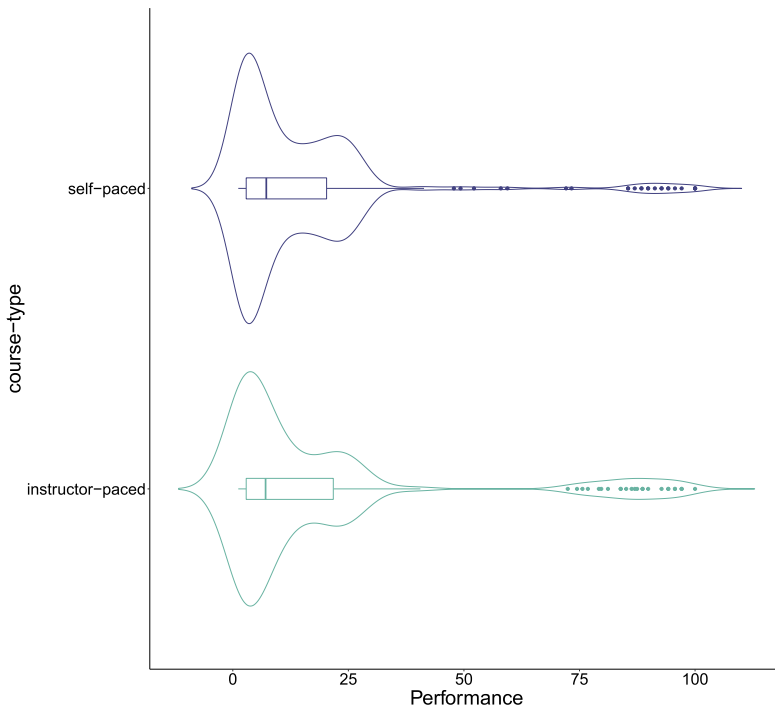


Fig. 6 Student's performance by course type

Table 2 Comparison of the QR and OLS coefficients (rows) for the global model (third column) and the two models according to the course type (fourth and fifth column). The *p*-values for the multigroup comparison using the parametric and the permutation test are in the last two columns

	Predictor	Coefficient			<i>p</i> -value	
		Global	Instructor	Self	Parametric Test	Permutation Test
			Paced	Paced		
$\tau = 0.25$	Learning	0.245***	0.365***	0.294***	0.339	0.495
	Engagement	0.194***	0.081 ^{NS}	0.231***	0.017	0.016
$\tau = 0.5$	Learning	0.420***	0.705***	0.348***	0.095	0.028
	Engagement	0.183***	-0.033 ^{NS}	0.267***	0.013	0.002
$\tau = 0.75$	Learning	0.886***	1.156***	0.610***	0.018	0.012
	Engagement	0.082 ^{NS}	-0.056 ^{NS}	0.185***	0.111	0.034
OLS	Learning	0.734***	0.962***	0.702***	0.070	0.180
	Engagement	0.169***	-0.014 ^{NS}	0.253***	0.018	0.070

*** < 0.001, ** < 0.01, * < 0.05, ^{NS} not significant

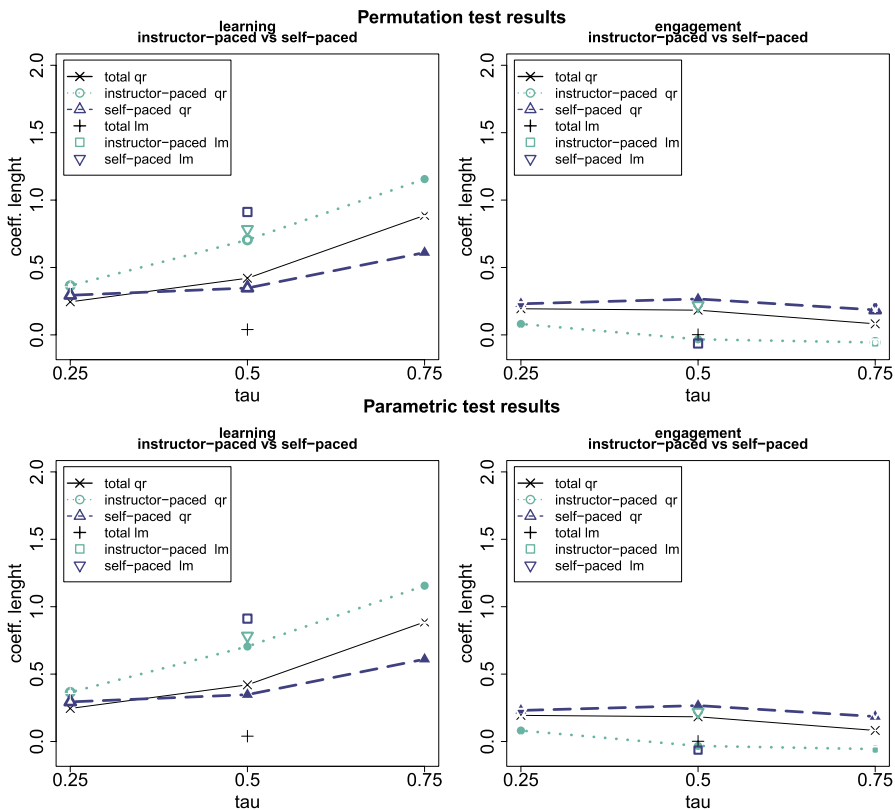


Fig. 7 Coefficient comparison of QR on the whole sample and on the two subgroups of students according to course type (different lines), the three conditional quartiles on the horizontal axis. The OLS coefficients are depicted using symbol at the conditional median for the sake of comparison. First row refers to parametric test, second row to permutation test; first column refers to learning component, second column to engagement

Appendix: Simulation results for equal coefficients between groups

See Figs. 8, 9 and 10.

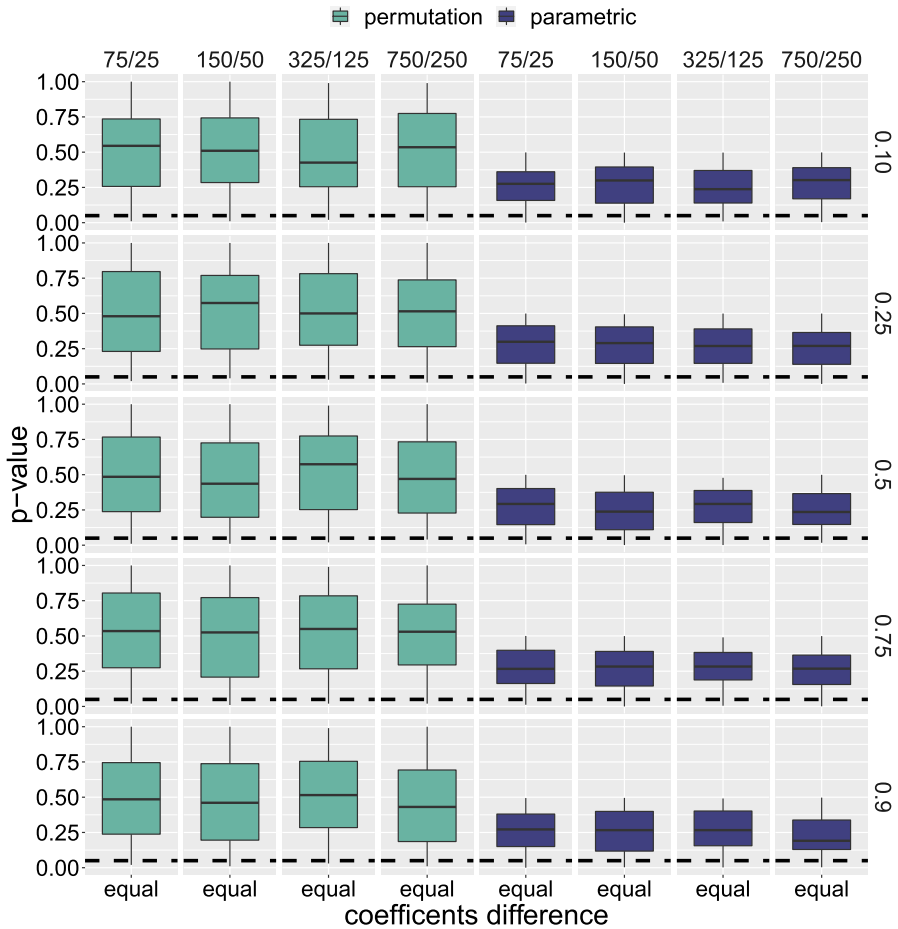


Fig. 8 Boxplots of p -values obtained on 100 replications of the parametric (first four columns) and permutation test (last four columns). Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250) by quantiles (from 0.1, 0.25, 0.5, 0.75, up 0.9), sample size (75 vs. 25, 150 vs. 50, 325 vs. 125, and 750 vs. 250). Simulations consider standard normal errors and unbalanced segments

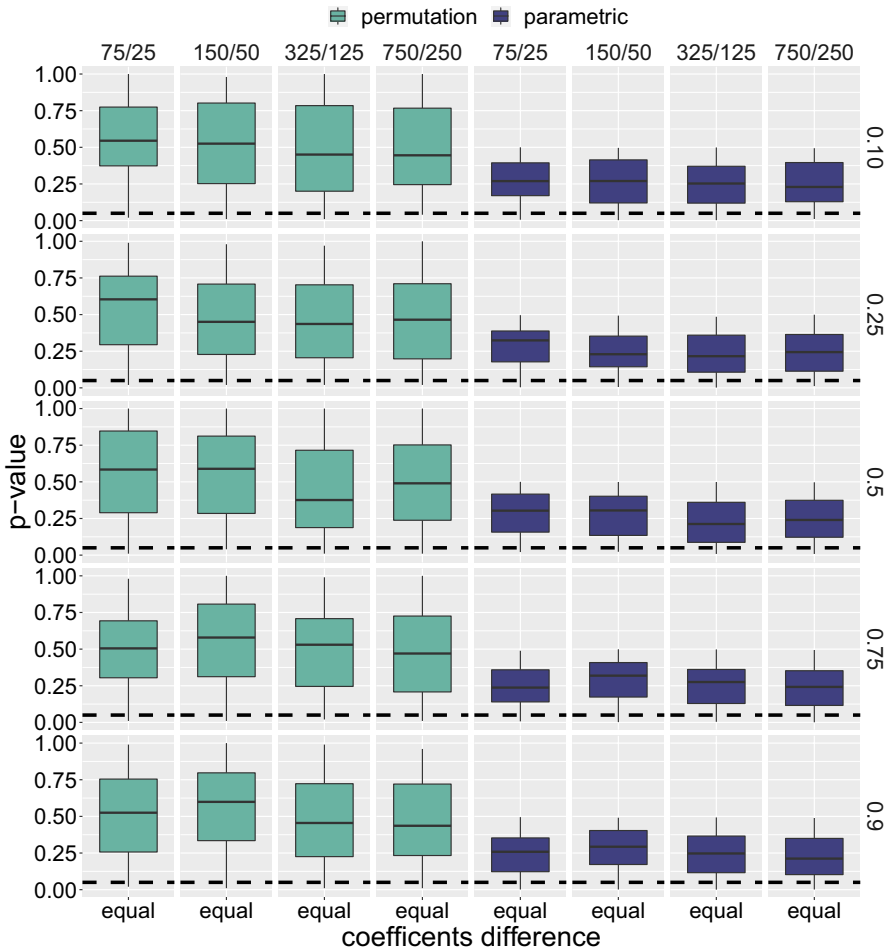


Fig. 9 Boxplots of p -values obtained on 100 replications of the parametric (first four columns) and permutation test (last four columns). Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250) by quantiles (from 0.1, 0.25, 0.5, 0.75, up 0.9), sample size (75 vs. 25, 150 vs. 50, 325 vs. 125, and 750 vs. 250). Simulations consider skew normal errors and unbalanced segments

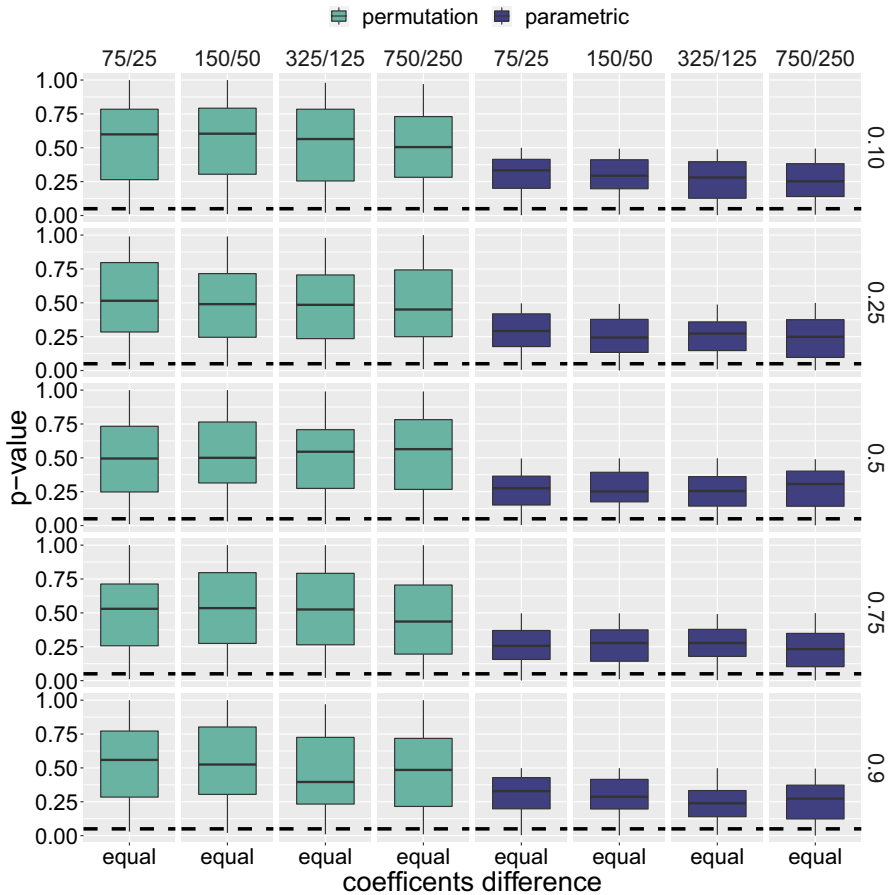


Fig. 10 Boxplots of p -values obtained on 100 replications of the parametric (first four columns) and permutation test (last four columns). Horizontal panels refer to quantiles (0.1, 0.25, 0.5, 0.75, 0.9), vertical panels to sample sizes (75 vs 25, 150 vs 50, 325 vs 125, and 750 vs 250) by quantiles (from 0.1, 0.25, 0.5, 0.75, up to 0.9), sample size (75 vs. 25, 150 vs. 50, 325 vs. 125, and 750 vs. 250). Simulations consider uniform errors and unbalanced segments

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baye A, Monseur C (2016) Gender differences in variability and extreme scores in an international context. *Large Scale Assess Educ* 4(1) <https://doi.org/10.1186/s40536-015-0015-x>
- Carannante M, Davino C, Vistocco D (2020) Modelling students' performance in MOOCs: a multivariate approach. *Stud High Educ* 32:453–468. <https://doi.org/10.1080/03075079.2020.1723526>
- Chin W, Dibbern J (2010) An introduction to a permutation based procedure for multi-group pls analysis: results of tests of differences on simulated data and a cross cultural analysis of the sourcing of information system services between germany and the usa. In: Vinzi VE, Chin W, Henseler J et al (eds) *Handbook of Partial Least Squares*. Springer Handbooks of Computational Statistics, Springer, Berlin, Heidelberg, pp 171–193
- Chow G (1960) Test of equality between sets of coefficients in two linear regressions. *Econometrica* 28:591–605. <https://doi.org/10.2307/1910133>
- Davino C, Romano R, Vistocco D (2022) Handling multicollinearity in quantile regression through the use of principal component regression. *METRON* 80:150–174. <https://doi.org/10.1007/s40300-022-00230-3>
- Davino C, Furno M, Vistocco D (2013) *Quantile Regression: Theory and Applications*. John Wiley & Sons
- de Barba P, Kennedy G, Ainley M (2016) The role of students' motivation and participation in predicting performance in a MOOC. *J Comput Assist Learn* 32:218–231. <https://doi.org/10.1111/jcal.12130>
- Efron B, Tibshirani R (1998) *Introduction to the Bootstrap*. Chapman & Hall
- Eslami A, Qannari E, Kohler A et al (2013) General overview of methods of analysis of multi-group datasets. *RNTI* 25:113–128
- Fianu E, Blewett C, Ampong G et al (2018) Factors affecting MOOC usage by students in selected Ghanaian universities. *Edu Sci* 8(2):70
- Furno M, Vistocco D (2018) *Quantile Regression: Estimation and simulation*. John Wiley & Sons
- Gelman A (2006) Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics* 48(3):432–435. <https://doi.org/10.1198/004017005000000661>
- Goopio J, Cheung C (2020) The MOOC dropout phenomenon and retention strategies. *J Teach Travel Tour* 21(2):177–97. <https://doi.org/10.1080/15313220.2020.1809050>
- Gujarati D (1970) Use of dummy variables in testing for equality between sets of coefficients in two linear regressions: a note. *Am Stat* 24(1):50–52. <https://doi.org/10.2307/2682300>
- Hair JJ, Sarstedt M, Ringle C et al (2012) An assessment of the use of partial least squares structural equation modeling in marketing research. *J Acad Mark Sci* 40(1):414–433. <https://doi.org/10.1007/s11747-011-0261-6>
- Hair JJ, Hult G, Ringle C et al (2016) *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications, Los Angeles
- Hair JJ, Sarstedt M, Ringle C et al (2018) *Advanced issues in partial least squares structural equation modeling*. Sage publications, Los Angeles
- Hansen KY, Gustafsson J (2016) Determinants of country differences in effects of parental education on children's academic achievement. *Large-scale assess educ* 4(1):1–13. <https://doi.org/10.1186/s40536-016-0027-1>
- Hintze J, Nelson R (1998) Violin plots: a box plot-density trace synergism. *Am Stat* 52:181–184. <https://doi.org/10.1080/00031305.1998.10480559>
- Keil M, Tan B, Wei K et al (2000) A cross-cultural study on escalation of commitment behavior in software projects. *MIS Q* 24(2):181–184. <https://doi.org/10.2307/3250940>
- Kherad-Pajouh S, Renaud O (2010) An exact permutation method for testing any effect in balanced and unbalanced fixed effect ANOVA. *Comput Stat Data Anal* 54(7):1881–1893. <https://doi.org/10.1016/j.csda.2010.02.015>
- Kleiner A, Talwalkar A, Sarkar P et al (2014) A scalable bootstrap for massive data. *J R Stat Soc Ser B Stat Methodol* 76(4):795–816
- Kocherginsky M, He X, Mu Y (2005) Practical confidence intervals for regression quantiles. *J Comput Graph Stat* 14:41–55
- Koenker R (2022) *Quantreg: quantile regression*. R Packag Vers 5:94
- Koenker R, Bassett J (1978) Regression quantiles. *Econometrica* pp 33–50. <https://doi.org/10.2307/1913643>
- Koenker R, Chernozhukov V, He X et al (2017) *Handbook of Quantile Regression*. Sage publications

- Lamberti G, Aluja T, Sanchez G (2016) The Pathmox approach for PLS path modeling. *Appl Stoch Models Bus Ind* 32:453–468. <https://doi.org/10.1002/asmb.2168>
- Lamberti G, Aluja T, Sanchez G (2016) The Pathmox approach for PLS path modeling: discovering which constructs differentiate segments. *Appl Stoch Models Bus Ind* 33(6):674–689. <https://doi.org/10.1002/asmb.2270>
- Lebart L, Morineau A, Fenelon J (1979) *Traitement des donnees statistiques*. Dunod, Paris
- Moore R, Wang C (2021) Influence of learner motivational dispositions on MOOC completion. *J Comput High Educ* 33(1):121–134. <https://doi.org/10.1007/s12528-020-09258-8>
- Raudenbush S, Bryk A (2002) *Hierarchical linear models: applications and data analysis methods*. Sage publications
- Sarstedt M, Henseler J, Ringle C (2011) Multi-group analysis in partial least squares (PLS) path modeling: alternative methods and empirical results. *Adv Int Mark* 22:195–218. [https://doi.org/10.1108/S1474-7979\(2011\)0000022012](https://doi.org/10.1108/S1474-7979(2011)0000022012)
- Sengupta S, Volgushev S, Shao X (2016) A subsampled double bootstrap for massive data. *J Am Stat Assoc* 111(515):1222–1232. <https://doi.org/10.48550/arXiv.1508.01126>
- Siemens G, Long P (2011) Penetrating the fog: analytics in learning and education. *EDUCAUSE Rev* 46(5):30–40
- Snijders T, Bosker R (2011) *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage publications
- Team RC (2002) *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna, Austria, <https://www.R-project.org/>
- Vinzi VE, Chin W, Henseler J et al (2013) *Handbook of partial least squares*. Springer Handbooks of Computational Statistics, Springer, Berlin
- Wold H (1985) Partial least squares. In: Kotz S, Johnson N (eds) *Encyclopedia of statistical sciences*. Wiley & Sons, New York, Heidelberg, pp 581–591
- Zeileis A, Hothorn T, Hornik K (2008) Model-based recursive partitioning. *J Comput Graph Stat* 17:492–514. <https://doi.org/10.1198/106186008X319331>
- Zou H, Yuan M (2008) Composite quantile regression and the oracle model selection theory. *Ann Statist* 36(3):1108–1126

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