MESE 1
Measurement in STEM Education (MESE1)
Naples, 30-31 January, 1 February

EXPLORING THE IMPACT OF STUDENT CHARACTERISTICS AND SOCIAL CONTEXT ON MATHEMATICAL LITERACY

Cristina Davino Rosaria Romano Domenico Vistocco

Department of Economics and Statistics
cristina.davino@unina.it
rosaroma@unina.it

Department of Political Science
domenico.vistocco@unina.it

University of Naples Federico II

OUTLINE

Aim

INVALSI data

Methodological framework

Main results

Conclusions

OUTLINE

Aim

INVALSI data

Methodological framework

Main results

Conclusions

Outline

Aim

INVALSI data

Methodological framework

Main results

Conclusions

Outline

Aim

INVALSI data

Methodological framework

Main results

Conclusions

OUTLINE

Aim

INVALSI data

Methodological framework

Main results

Conclusions

Mathematical literacy

Definition

Mathematical literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgements and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen (OECD/PISA, 2003)

Mathematical literacy

Mathematical literacy has received increasing attention in many countries over the last few years:

- driven by concerns of employers that too many students leave school unable to function mathematically at the level needed in the modern world of work
- it is increasingly recognised that people can only tackle many of the challenges of modern life effectively if they are mathematically literate in key areas (planning in personal finance, assessment of risk, design in the home or on the computer screen, and critical appraisal of the flood of statistical information from advertising, politicians and the press (Steen, Turner, Burkhardt, 2007)

Factors affecting the learning process

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 (Costanzo, Desimoni, 2017):

- gender differences and the impact of students' socioeconomic conditions on learning achievement explored by international comparative studies (IEA, OECD, NAEP).
- 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

FACTORS AFFECTING THE LEARNING PROCESS

The results achieved by each student are affected by different components:

- The outcomes of the learning-teaching process
- Some individual characteristics of the student (gender, the field of study attended, regularity in studies, the economic-social-cultural context of the family of origin, etc.)
- The environment in which they live (geographical area of residence, the economic-social-cultural context of the school, etc.)

AIM

Aim of the talk

Exploring the impact of student characteristics and social context on mathematical literacy highlighting heterogeneity:

- unobserved
- territorial
- context

High heterogeneity is often more realistic for modeling the messy real world and may give better results or identify subpopulations

AIM OF THE TALK

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

AIM OF THE TALK

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

Methodological framework:

Quantile regression (QR)

(Koenker R., Basset G. 1978)
(Koenker R. 2005)
(Koenker R. `quantreg R package` 2018)
(Davino C., Furno M., Vistocco D. 2013)
(Furno M., Vistocco D. 2018)

Handling heterogeneity among units

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

CLUSTERING & MODELING: Identifying a typology in a dependence model

- Identifying groups of units characterized by similar dependence structures
- Discovering the best model for each group
- Testing differences among groups

Handling heterogeneity among units

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

Research questions?

- How to identify unobserved heterogeneity?
- How to partition the units according to the dependence relationship?
- How many groups?
- What is the best model for each group?

Handling heterogeneity among units

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

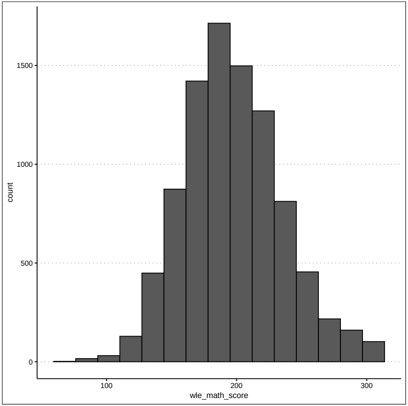
Comparison with alternative methods

- Estimation of different models for each group
- Introduction of a dummy variable
- Multilevel modeling

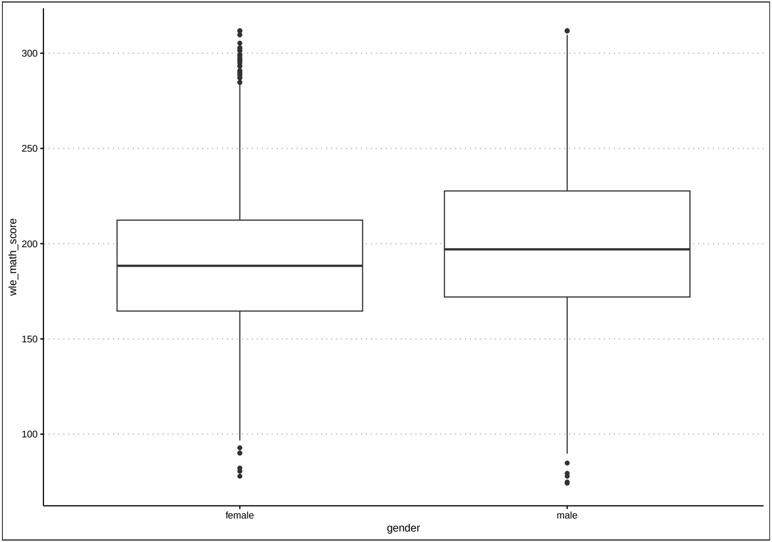
INVALSI DATA

INVALSI MATHEMATICS TESTS

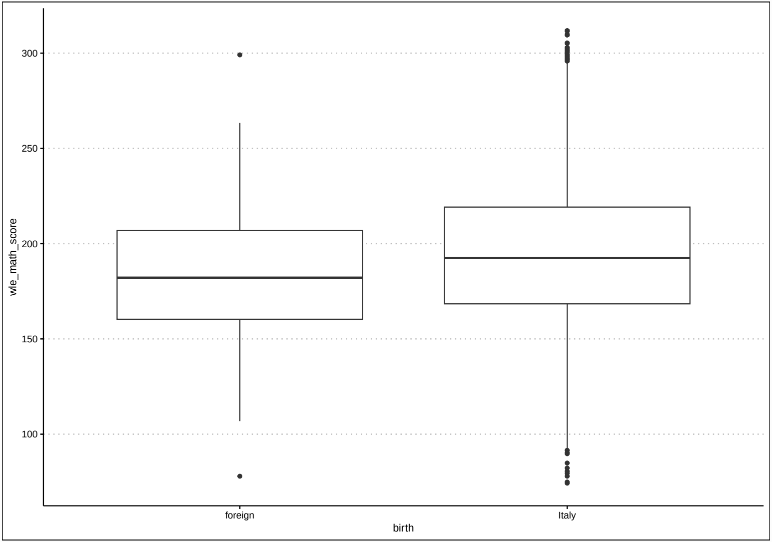
- Sample data
- **13 grade** students (at the end of upper secondary school)
- Outcome variable: **ability math score** (*wle_math_score*)
- **Factors**: school, gender, age, place of birth, regularity, origin, area, escs (Economic, Social and Cultural Status) index



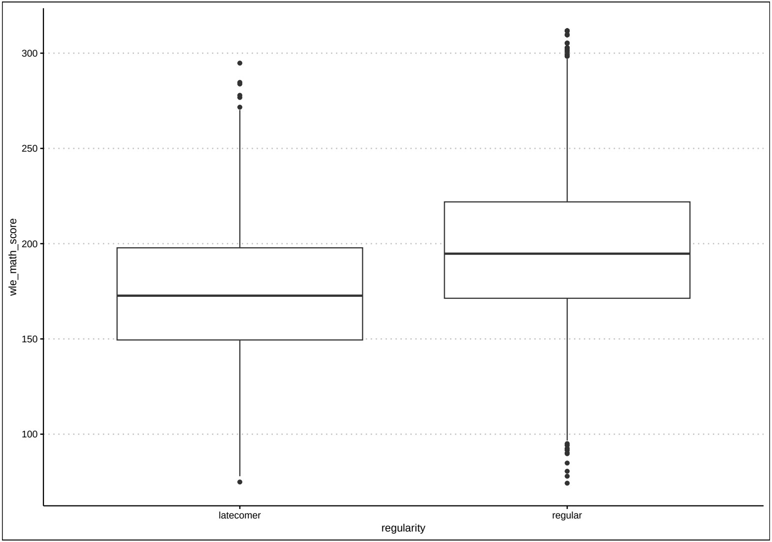
FACTORS AFFECTING MATHS ABILITY: GENDER



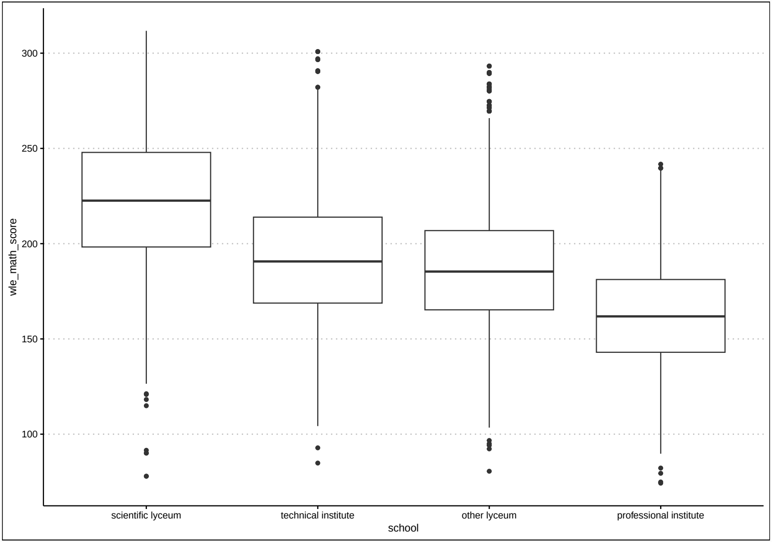
FACTORS AFFECTING MATHS ABILITY: PLACE OF BIRTH



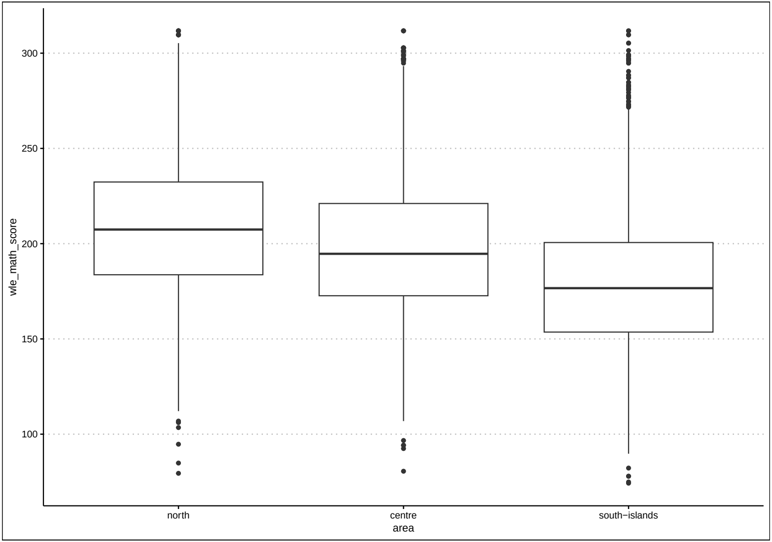
FACTORS AFFECTING MATHS ABILITY: REGULARITY OF SCHOOL CAREER



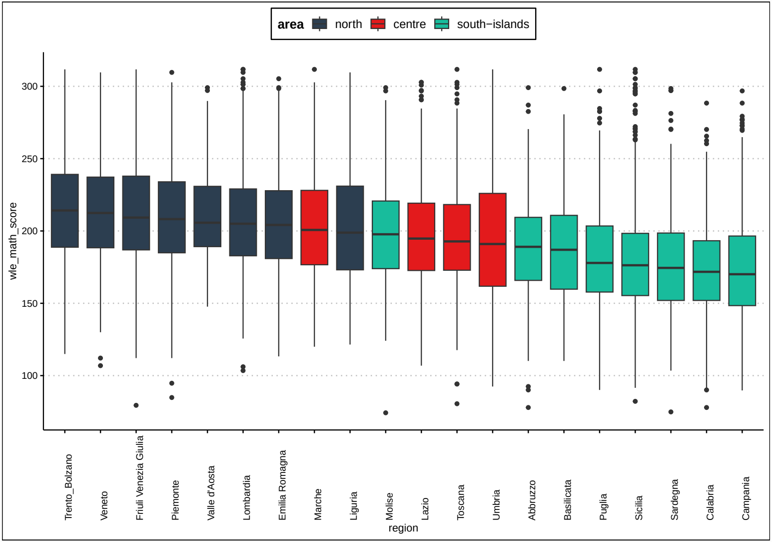
FACTORS AFFECTING MATHS ABILITY:
TYPE OF SCHOOL



FACTORS AFFECTING MATHS ABILITY: GEOGRAPHICAL AREA



FACTORS AFFECTING MATHS ABILITY: REGION



METHODOLOGICAL FRAMEWORK

Methodological framework
Quantile Regression

Quantile regression

QR has become a popular alternative to least squares regression for modeling heterogeneous data

Mosteller and Tukey (1977)

What the regression curve does is give a grand summary for the averages of the distributions corresponding to the set of **X**'s.
We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set.
Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions.

(Koenker R W and Basset G, Regression Quantiles. *Econometrica* **46**(1), 1978)

QUANTILE REGRESSION

- QR gained popularity in applied economics by the end of the 90’s, when people realize the importance of heterogeneity
- Application fields:
  - astrophysics
  - chemistry
  - ecology
  - economics
  - finance
  - food science
  - genomics
  - medicine
  - meteorology
  - sociology
  - marketing

Classical vs Quantile regression

Classical linear regression (conditional expected value)

estimation of the conditional mean of a response variable (Y) as a function of a set X of predictor variables

Quantile regression (conditional quantiles)

estimation of the conditional quantiles of a response variable (Y) as a function of a set X of predictor variables

CLASSICAL VS QUANTILE REGRESSION

Classical linear regression (conditional expected value)

estimation of the conditional mean of a response variable (Y) as a function of a set X of predictor variables

Quantile regression (conditional quantiles)

estimation of the conditional quantiles of a response variable (Y) as a function of a set X of predictor variables

Quantile Regression

QR allows to handle:

- heteroscedasticity
- skewness
- kurtosis
- outliers in Y

QR:

- generalizes univariates quantiles for conditional distributions
- analyses regressor effects on the whole dependent variable
- is equivariant to monotone transformations distribution

(Koenker R., Basset G. 1978) (Koenker R. 2005)

(Davino C., Furno M., Vistocco D. 2013) (Furno M., Vistocco D. 2018)

Quantile Regression Model

$$y_i = \mathbf{x}_i \beta(\theta) + \epsilon_i(\theta)$$

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

where

- $\mathbf{x}_i$ a generic row of the regressor matrix $\mathbf{X}$
- $\mathbf{y}$: dependent variable
- $0 < \theta < 1$: a generic quantile
- $Q_\theta(.|.)$: conditional quantile function
- ϵ: error term such that $Q_\theta(\epsilon|\mathbf{X}) = 0$.

Interpretation

$$\hat{\beta}_i(\theta) = \frac{\partial Q_\theta(\mathbf{y}|\mathbf{X})}{\partial \mathbf{x}_i}$$

Rate of change in the θ^{th} quantile of the dependent variable distribution for a one-unit change in the value of the i^{th} regressor, taking constant all the other regressors

Main results

Aim of the talk

Exploring the impact of student characteristics and social context on mathematical literacy highlighting heterogeneity:

- unobserved
- territorial
- context

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

AIM OF THE TALK

Exploring the impact of student characteristics and social context on mathematical literacy highlighting heterogeneity:

- unobserved
- territorial
- context

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

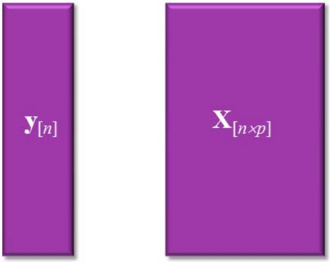
The main steps of the unsupervised approach

1. Identification of the global dependence structure
2. Identification of the best model for each unit
3. Clustering units
4. Modeling groups
5. Testing differences among groups

BASIC NOTATION

The data structure

- n units
- p regressors
- *1* quantitative or ordinal dependent variable



Basic notation

The data structure

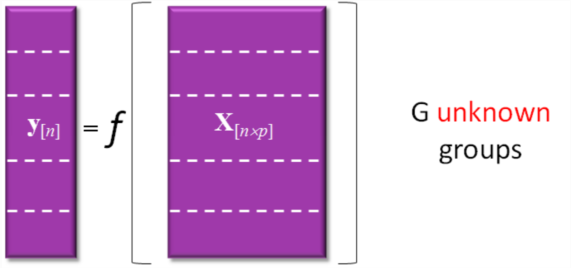
- n units
- p regressors
- 1 quantitative or ordinal dependent variable

$$\mathbf{y}_{[n]} = f\left[\mathbf{X}_{[n \times p]}\right]$$

Basic notation

The data structure

- n units
- p regressors
- 1 quantitative or ordinal dependent variable



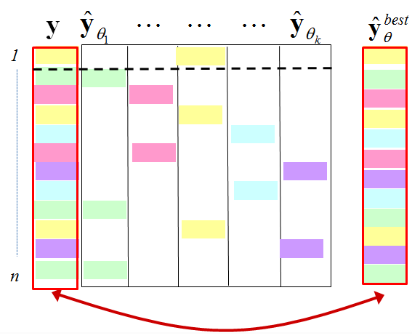
THE PROPOSED APPROACH

1.Identification of the global dependence structure

$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}(\theta) \quad \theta = 1, \ldots, k$

2.Identification of the best model for each unit

- estimated values
  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$
- best model identification
  $\theta_i^{best} : \underset{\theta=1,\ldots,k}{\text{argmin}} \ |\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta)|$
- best estimates identification
  $\hat{\mathbf{y}}_\theta^{best}$



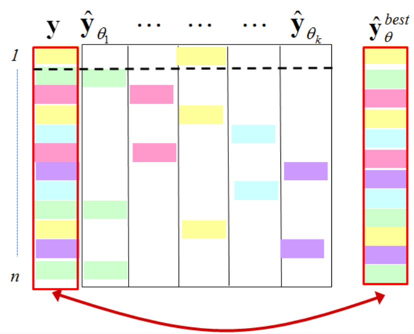
THE PROPOSED APPROACH

1.Identification of the global dependence structure

$$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}(\theta) \quad \theta = 1, \dots, k$$

2.Identification of the best model for each unit

- estimated values
  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$
- best model identification
  $\theta_i^{best} : \underset{\theta=1,\dots,k}{\text{argmin}} \ |\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta)|$
- best estimates identification
  $\hat{\mathbf{y}}_\theta^{best}$



INVALSI RESULTS

1. Global estimation

$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}(\theta)$

2. Identification of the best model for each unit

1. estimated values
   $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$
2. best model identification
   $\theta_i : \underset{\theta=1,\ldots,k}{\operatorname{argmin}} |y_i - \hat{y}_i(\theta)|$
3. best estimates identification
   $\hat{\mathbf{y}}_\theta^{best}$

INVALSI RESULTS

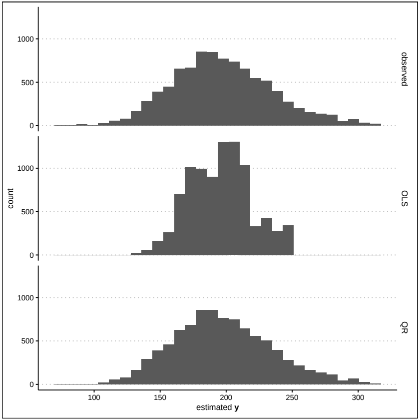
1. Global estimation

$$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}(\theta)$$

2. Identification of the best model for each unit

1. estimated values
   $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$
2. best model identification
   $\theta_i : \underset{\theta=1,\dots,k}{\text{argmin}} \; |y_i - \hat{y}_i(\theta)|$
3. best estimates identification
   $\hat{\mathbf{y}}_\theta^{best}$

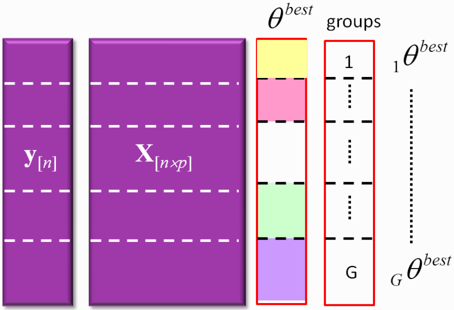
Distribution of the dependent variable: observed (left panel), LS estimated (middle panel), best QR estimated (right panel)



THE PROPOSED APPROACH

3. Clustering units

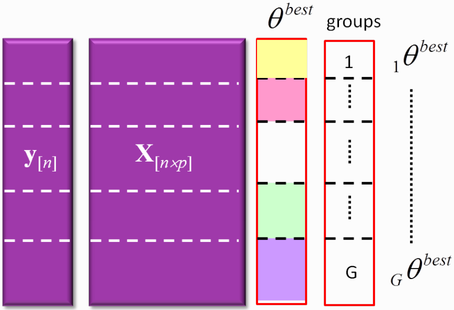
- finding the best partition of the θ^{best} vector
- identification of the group reference quantile ${}_g\theta^{best}$, for $g = 1, G$



THE PROPOSED APPROACH

3. Clustering units

- finding the best partition of the θ^{best} vector
- identification of the group reference quantile $_g\theta^{best}$, for $g = 1, G$



3. CLUSTERING UNITS

Finding the best partition of the θ^{best} vector

- θ^{best} is partitioned into D groups (e.g. according to the deciles)
- identification of a reference quantile for each of the D groups:

$$_{d}\overline{\theta}^{best} = \frac{\sum_{i=1}^{n_d} \theta_i^{best}}{n_d}$$

$(d = 1, \ldots, D)$

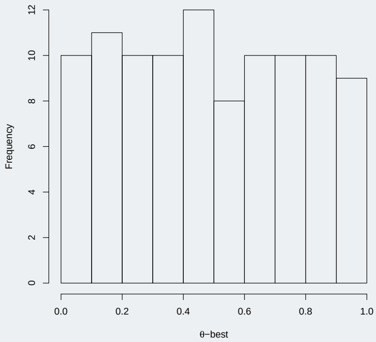
- estimate D quantile regression models with $\theta = \left[{}_{1}\overline{\theta}^{best}, \ldots, {}_{D}\overline{\theta}^{best} \right]$

INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector: a solution

- θ^{best} is partitioned according to its deciles ($d = 1, \ldots, D$)

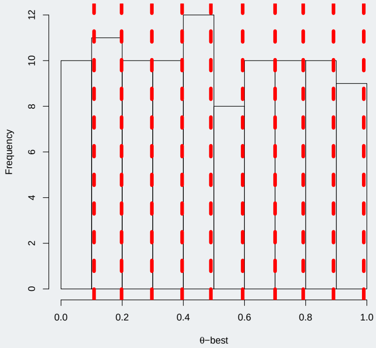


INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector

- θ^{best} is partitioned according to its deciles ($d = 1, \ldots, D$)



INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector

- identification of a reference quantile for each of the D groups:

quantile	$_d\bar{\theta}^{best}$
0.1	0.053
0.2	0.159
0.3	0.264
0.4	0.371
0.5	0.470
0.6	0.570
0.7	0.670
0.8	0.770
0.9	0.864

- estimate D quantile regression models

3. Clustering units

Finding the best partition of the θ^{best} vector

- test whether the slopes of pairs of consecutive models are identical

  Joint Test of Equality of Slopes

  Koenker R.W. and Basset G. 1982 Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**(1)

- group units if their reference quantiles do not provide significantly different coefficients

- identification of the group reference quantile ${}_g\theta^{best}$, for $g = 1, G$

HETEROSCHEDASTICITY TEST

$Q_{\theta_i}(\hat{\mathbf{y}}|\mathbf{x}) = \hat{\beta}_0(\theta_i) + \hat{\beta}_1(\theta_i)\mathbf{x}$
$Q_{\theta_j}(\hat{\mathbf{y}}|\mathbf{x}) = \hat{\beta}_0(\theta_j) + \hat{\beta}_1(\theta_j)\mathbf{x}$

$H_0 : \beta_1(\theta_i) = \beta_1(\theta_j)$

Test Statistic:

$$T = \frac{\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]^2}{var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]} \sim \chi^2_{1gdl} \tag{1}$$

where $var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right] =$
$var\left[\hat{\beta}_1(\theta_i)\right] + var\left[\hat{\beta}_1(\theta_j)\right] - 2cov\left[\hat{\beta}_1(\theta_i)\,\hat{\beta}_1(\theta_j)\right]$
A possible solution to estimate $var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right]$: bootstrap

INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector

- sequentially test if the slope coefficients of the models are identical

quantile	${}_d\overline{\theta}^{best}$	p-value
0.1	0.053	0.008
0.2	0.159	0.092
0.3	0.264	0.102
0.4	0.371	0.151
0.5	0.470	0.006
0.6	0.570	0.002
0.7	0.670	0.193
0.8	0.770	0.000
0.9	0.864	0.127

INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector

- group units if their reference quantiles provide not significantly different coefficients

quantile	$_d\bar{\theta}^{best}$	p-value	group	n_g
0.1	0.053	0.008	1	898
0.2	0.159	0.092	2	3548
0.3	0.264	0.102		
0.4	0.371	0.151		
0.5	0.470	0.006		
0.6	0.570	0.002	3	876
0.7	0.670	0.193		
0.8	0.770	0.000	4	1882
0.9	0.864	0.127	5	1946

INVALSI RESULTS

3. Clustering units

Finding the best partition of the θ^{best} vector

- identification of the group reference quantile

quantile	${}_d\overline{\theta}^{best}$	p-value	group	n_g	${}_g\theta^{best}$
0.1	0.053	0.008	1	898	0.053
0.2	0.159	0.092	2	3548	0.305
0.3	0.264	0.102			
0.4	0.371	0.151			
0.5	0.470	0.006			
0.6	0.570	0.002	3	876	0.554
0.7	0.670	0.193			
0.8	0.770	0.000	4	1882	0.705
0.9	0.864	0.127	5	1946	0.903

THE PROPOSED APPROACH

4. Modeling groups

$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}({}_g\theta^{best})$

5. Testing differences among groups

- Testing if all the slope coefficients of the groups are identical
- Separate testing on each slope coefficient

Koenker R.W. and Basset G. 1982 Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**(1)

THE PROPOSED APPROACH

4. Modeling groups

$$Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}({}_g\theta^{best})$$

5. Testing differences among groups

- Testing if all the slope coefficients of the groups are identical
- Separate testing on each slope coefficient

Koenker R.W. and Basset G. 1982 Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**(1)

INVALSI RESULTS

Step 4: Modeling groups

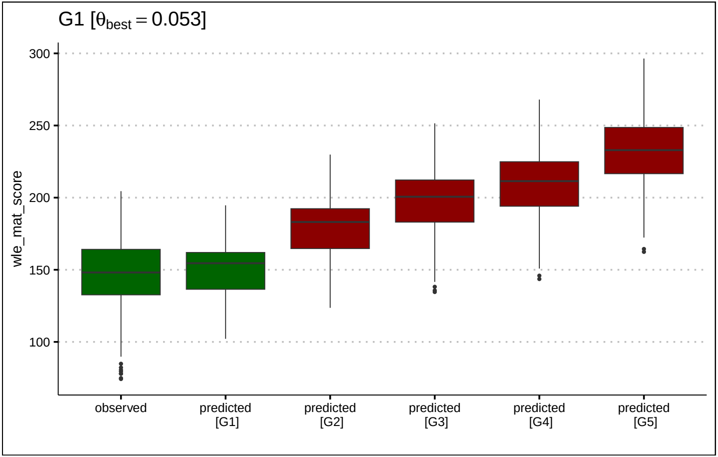
QR coefficients with group effects

Variable	OLS	G1 $\theta = 0.053$	G2 $\theta = 0.305$	G3 $\theta = 0.554$	G4 $\theta = 0.705$	G5 $\theta = 0.903$
(Intercept)	**213.67**	**145.82**	**205.22**	**222.53**	**216.38**	**248.64**
technical institute	**-30.08**	**-26.95**	**-28.45**	**-28.20**	**-31.85**	**-37.48**
other lyceum	**-32.99**	**-29.94**	**-31.23**	**-32.22**	**-34.11**	**-38.24**
professional institute	**-54.03**	**-49.31**	**-48.57**	**-52.42**	**-54.35**	**-64.48**
male	**11.03**	**6.10**	**8.73**	**11.98**	**14.36**	**16.57**
age	0.22	1.59	-0.13	-0.07	0.89	0.35
birth_Italy	3.31	3.33	2.38	-0.18	0.11	6.31
regular career	**12.45**	8.89	**12.30**	**14.12**	**15.67**	**12.45**
foreigner	-1.79	-2.76	**-4.08**	**-3.57**	-1.76	1.48
centre	**-14.54**	**-13.83**	**-15.01**	**-14.64**	**-15.34**	**-11.30**
south-islands	**-30.91**	**-26.91**	**-30.92**	**-31.52**	**-32.37**	**-31.12**
escs	**3.16**	1.31	**2.45**	**3.04**	**3.83**	**4.35**

Step 4: Modeling groups

Group 1

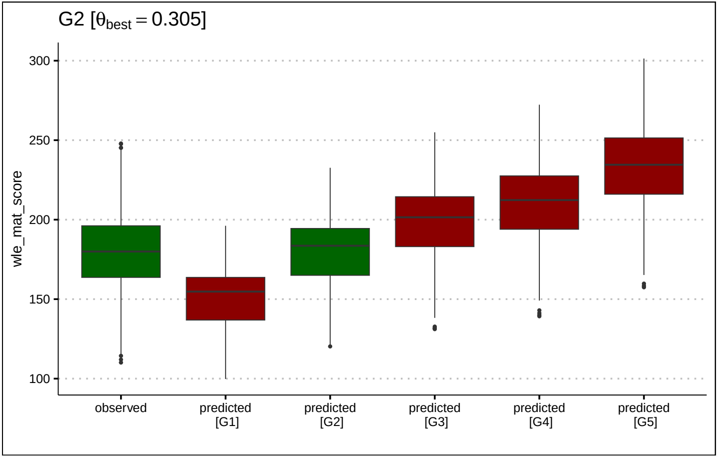
Observed and estimated response distributions using the reference quantile of G1



Step 4: Modeling groups

Group 2

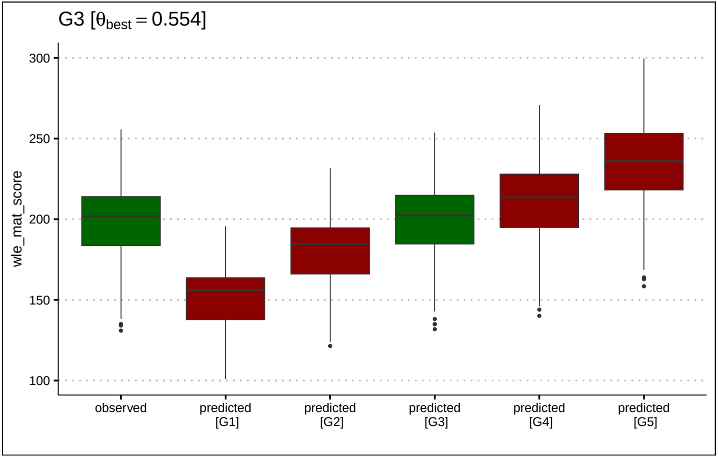
Observed and estimated response distributions using the reference quantile of G2



Step 4: Modeling groups

Group 3

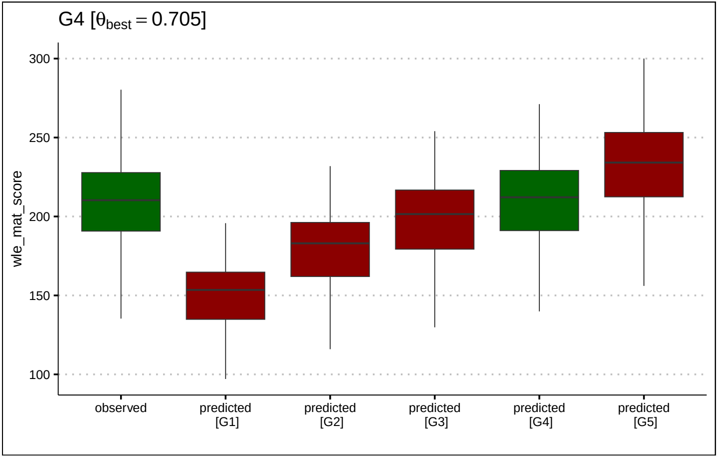
Observed and estimated response distributions using the reference quantile of G3



Step 4: Modeling groups

Group 4

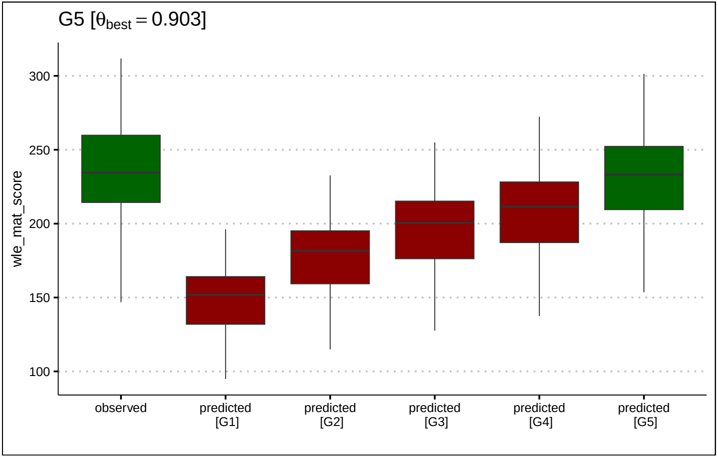
Observed and estimated response distributions using the reference quantile of G4



STEP 4: MODELING GROUPS

Group 5

Observed and estimated response distributions using the reference quantile of G5



Step 5: Testing differences among groups

Testing if all the slope coefficients of the groups are identical

p-values

	p-value
G1 vs G2	0.0003
G2 vs G3	0.0000
G3 vs G4	0.0008
G4 vs G5	0.0000

Separate testing on each slope coefficient

	G1 vs G2	G2 vs G3	G3 vs G4	G4 vs G5
technical institute	0.360	0.826	0.001	0.001
other lyceum	0.397	0.317	0.061	0.017
professional institute	0.733	0.007	0.136	0.000
male	0.044	0.000	0.002	0.075
age	0.316	0.958	0.377	0.749
birth_Italy	0.849	0.364	0.890	0.040
regular career	0.219	0.369	0.395	0.233
foreigner	0.604	0.764	0.178	0.063
centre	0.444	0.724	0.439	0.003
south-islands	0.002	0.497	0.303	0.358
escs	0.074	0.169	0.048	0.403

RECAP & PROS

Clustering units taking into account the dependence structure

- Estimation of the group dependence structure using the whole sample
- Impact of the regressors on the entire conditional distribution
- Clarity of the final results
- Availability of classical inferential procedures to test differences among groups
- Number of groups defined by the procedure
- Exact solution method

Aim of the talk

Exploring the impact of student characteristics and social context on mathematical literacy highlighting heterogeneity:

- unobserved
- territorial
- context

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

AIM OF THE TALK

Exploring the impact of student characteristics and social context on mathematical literacy highlighting heterogeneity:

- unobserved
- territorial
- context

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

The main steps of the unsupervised approach

1. Identification of the global dependence structure
2. Identification of the best model for each unit
3. ~~Clustering units~~ Identification of the best model for each group
4. Modeling groups
5. Testing differences among groups

SUPERVISED APPROACH

3. Identification of the best model for each group

Geographical area

	θ^{best}
South	0.378
Center	0.521
North	0.610

Gender

	θ^{best}
Female	0.466
Male	0.521

School

	θ^{best}
Scientific lyceum	0.715
Technical institute	0.490
Other lyceum	0.442
Professional institute	0.253

4. Modeling groups
5. Testing differences among groups

Supervised approach: School differences

Step 4: Modeling groups

QR coefficients with *school* effects

Variable	G_{prof} Professional inst. $\theta = 0.253$	G_{oth} Other lyceum $\theta = 0.0.442$	G_{tech} Technical inst. $\theta = 0.490$	G_{sci} Scientific lyc. $\theta = 0.715$
(Intercept)	**162. 71**	**221.52**	**227.84**	**269.42**
centre	**-13.58**	**-14.03**	**-13.70**	**-13.94**
south-islands	**-28.14**	**-31.15**	**-30.93**	**-32.34**
male	**10.63**	**14.01**	**15.17**	**16.36**
age	0.23	-2.32	-2.41	-3.66
birth_Italy	-1.80	1.83	1.01	1.52
regular career	**19.08**	**18.40**	**17.38**	**17.82**
foreigner	**-3.74**	-1.28	-0.01	-2.69
escs	**6.64**	**8.32**	**9.01**	**9.59**

Step 5: Testing differences among groups

Testing if all the slope coefficients of the groups are identical

p-values

	p-value
G_{prof} vs G_{oth}	0.002
G_{prof} vs G_{tech}	0.022
G_{prof} vs G_{sci}	0.000
G_{oth} vs G_{tech}	0.468
G_{oth} vs G_{sci}	0.000
G_{tech} vs G_{sci}	0.000

Separate testing on each slope coefficient

	G_{prof} vs G_{sci}	G_{oth} vs G_{sci}	G_{tech} vs G_{sci}
centre	0.290	0.335	0.218
south-islands	0.169	0.780	0.673
male	0.000	0.008	0.017
age	0.103	0.394	0.450
birth_Italy	0.659	0.814	0.937
regular career	0.358	0.409	0.191
foreigner	0.870	0.877	0.457
escs	0.001	0.264	0.453

CONCLUSIONS

Concluding remarks:
back to motivation

Importance of knowledge of mathematics

Mathematical competence is one of the critical skills for personal fulfilment, active citizenship, social inclusion and lifelong learning, both nationally and internationally. (INVALSI, 2021)

Mathematical literacy, like literacy in language, is empowering

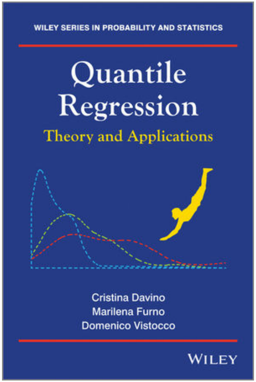
CONCLUDING REMARKS:
BACK TO MOTIVATION

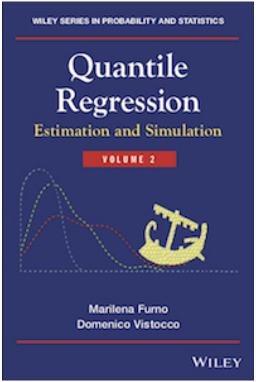
QR is capable of providing a more complete, more nuanced view of heterogeneous covariate effects (Koenker et al., 2017)

MAIN REFERENCES

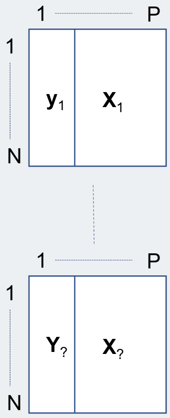
- Koenker R.W., Basset G. (1978) Regression quantiles. *Econometrica*, Vol. 46, No. 1.
- Koenker R. (2005) *Quantile Regression*. Econometric Society Monographs. Cambridge: Cambridge University Press.
- Koenker R. (2018). quantreg: Quantile Regression. R package version 5.35. https://CRAN.R-project.org/package=quantreg
- Mosteller F., Tukey J. (1977) *Data Analysis and Regression: A Second Course in Statistics*. Reading, MA: Addison–Wesley.

QUANTILE REGRESSION





UNSUPERVISED APPROACH / UNOBSERVED HETEROGENEITY



(e)

- *Methodological aim*: identifying group effect through a quantile regression model
- *Students'performance*: investigating the impact of students'features on University outcome

STATISTICS AND ITS INTERFACE Volume 11 (2018) 541–556

Handling heterogeneity among units in quantile regression. Investigating the impact of students' features on University outcome

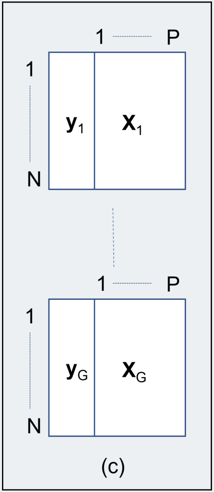
CRISTINA DAVINO* AND DOMENICO VISTOCCO

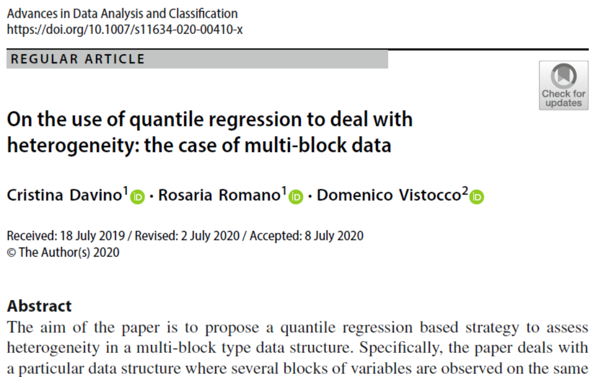
In many real data applications, statistical units belong to different groups and statistical models should be tailored to incorporate and exploit this heterogeneity among units. This

share the aim of inspecting ture affects the impact of variable, although they d ability to detect group effe The simplest approach

SUPERVISED APPROACH / OBSERVED HETEROGENEITY

- *Methodological aim*: clustering units according to the similarities in the dependence structure
- *Consumer studies*: clustering groups of consumers according to the similarities in the dependence structure among their overall liking and the liking for different drivers





Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-020-00410-x

REGULAR ARTICLE

Check for updates

On the use of quantile regression to deal with heterogeneity: the case of multi-block data

Cristina Davino[1] · Rosaria Romano[1] · Domenico Vistocco[2]

Received: 18 July 2019 / Revised: 2 July 2020 / Accepted: 8 July 2020

Abstract
The aim of the paper is to propose a quantile regression based strategy to assess heterogeneity in a multi-block type data structure. Specifically, the paper deals with a particular data structure where several blocks of variables are observed on the same

MAIN REFERENCES

- Davino C., Vistocco D. (2015) Quantile Regression for Clustering and Modeling Data. In I. Morlini, T. Minerva, M. Vichi (eds) *Advances in Statistical Models for Data Analysis*: Studies in Classification, Data Analysis, and Knowledge Organization. p. 85–96, Springer, Heidelberg.
- Davino C., Vistocco D. (2008) Quantile regression for the evaluation of student satisfaction. *Italian Journal of Applied Statistics* **20**, 179–196.
- Davino C., Vistocco D. (2015) Quantile Regression for Clustering and Modeling Data, In I. Morlini, T. Minerva, M. Vichi (eds), Advances in Statistical Models for Data Analysis, pp. 85–95, Springer.
- Davino C., Vistocco D. (2007) The evaluation of University educational processes: a quantile regression approach. *STATISTICA*, n.3, pp. 267-278.
- Davino C., Vistocco D. (2018) Handling heterogeneity among units in quantile regression. Investigating the impact of students' features on University outcome, *Statistics & Its Interface*, Vol. 11, pp. 541-556.
- Davino C., Romano R., Vistocco D. (2018) Modelling drivers of consumer liking handling consumer and product effects, *Italian Journal of Applied Statistics* (in press).

MAIN REFERENCES

- Davino C., Dolce P., Taralli S. (2017) Quantile Composite-based Model: a Recent Advance in PLS-PM. A Preliminary Approach to Handle Heterogeneity in the Measurement of Equitable and Sustainable Well-Being. In H. Latan, R. Noonan (eds) *Partial Least Squares Structural Equation Modeling - Basic Concepts, Methodological Issues and Applications*, Springer.
- Davino C., Dolce P., Esposito Vinzi V., Taralli S. (2016) A Quantile Composite-Indicator Approach for the Measurement of Equitable and Sustainable Well-Being: A Case Study of the Italian Provinces. *Social Indicators Research*, vol. xx, p. 1-318.
- Davino C., Dolce P., Taralli S., Vistocco D. (2020) Composite-Based Path Modeling for Conditional Quantiles Prediction. An Application to Assess Health Differences at Local Level in a Well-Being Perspective. *SOCIAL INDICATORS RESEARCH*, doi: 10.1007/s11205-020-02425-5.
- Davino C., Romano R., Vistocco D. (2020) On the Use of Quantile Regression to deal with Heterogeneity: the Case of Multi-block Data. *ADVANCES IN DATA ANALYSIS AND CLASSIFICATION*, vol. 14, p. 771-784.
- Carannante M., Davino C., Vistocco D. (2021) Modelling students'performance in MOOCs: a multivariate approach. *STUDIES IN HIGHER EDUCATION*, vol. 46, p.2371–2386.