Handling heterogeneity among units in Quantile Regression

Cristina Davino   Domenico Vistocco

University of Macerata
cristina.davino@unimc.it

University of Cassino
vistocco@unicas.it

Workshop on
"Recent Advances in Quantile and M-quantile Regression"
Pisa - July 15, 2016

all computations and graphics were done in the R language using the packages *quantreg* and *ggplot2*

Outline

Outline

Outline

Outline

Outline

Outline

Our interests in the QR framework

Application fields

- Educational process evaluation
- Financial data
- Customer satisfaction
- Consumer studies
- Equitable and Equitable and sustainable well-being

Handling heterogeneity among units

- Methodology: supervised and unsupervised
- Applications: educational process effectiveness
- Simulation study

Our interests in the QR framework

Quantile Composite-based Path Modeling

- Methodology (estimation, assessment)
- Applications: European and American Customer Satisfaction Index, Equitable and Equitable and sustainable well-being
- Simulation study

Comparison among estimators for linear regression

- Methodology: Prediction Intervals
- Simulation study

Our interests in the QR framework

Detection of structural breaks in quantile regression

- Methodology: modified Qu test for structural breaks
- Applications: Consumption data and OECD-PISA test
- Simulation study

Quantile style analysis

- Methodology: Constrained quantile regression and robust confidence intervals through subsampling
- Applications: Financial data
- Simulation study

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

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

Aim of the talk

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

The main goals of the paper:

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

Aim of the talk

Identification of group effects in a regression model

- Unsupervised approach
- Supervised approach

The main goals of the paper:

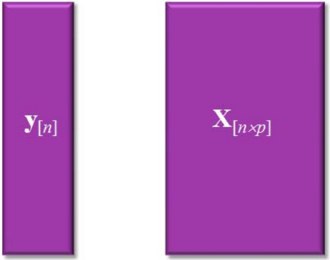
- Identifying groups of units characterized by similar dependence structures
- Discovering the best model for each group
- Testing differences among groups

Methodological framework: Quantile regression

Basic notation

The data structure

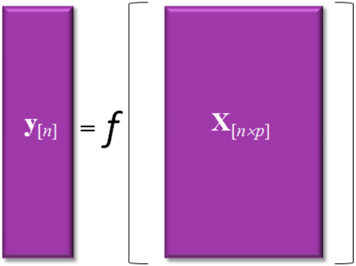
- *n* units
- *p* regressors
- *1* quantitative or ordinal dependent variable



Basic notation

The data structure

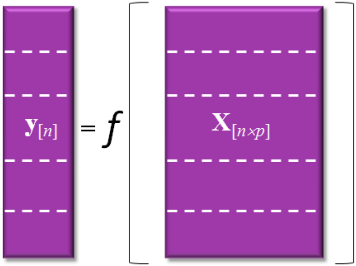
- *n* units
- *p* regressors
- *1* quantitative or ordinal dependent variable



Basic notation

The data structure

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



G unknown groups

The proposed approach

1. Identification of the global dependence structure
2. Identification of the best model for each unit
3. Identification of groups (partition) with different dependence structures
4. Estimation of the group dependence structure
5. Testing differences among groups

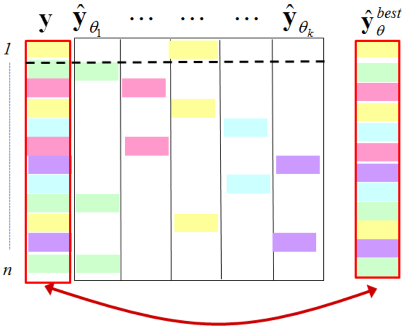
The proposed approach

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

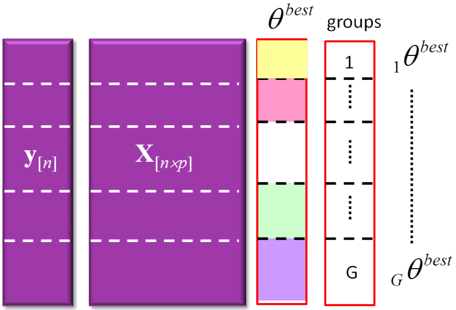
- estimated values
  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$
- best model identification
  $\theta_i^{best} : \underset{\theta=1,\ldots,k}{\text{argmin}} \; |\mathbf{y}_i - \hat{\mathbf{y}}_i(\theta)|$
- best estimates identification
  $\hat{\mathbf{y}}_\theta^{best}$



The proposed approach

3.Identification of a partition

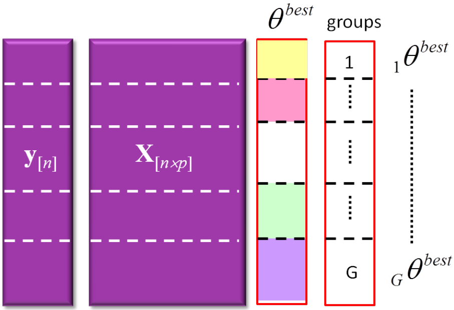
- finding the best partition of the θ^{best} vector
- identification of the group reference quantile ${}_g\theta^{best}$, for $g = 1, G$



The proposed approach

3.Identification of a partition

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



3.Identification of a partition

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:

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

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

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

3.Identification of a partition

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

- sequentially test if all the slope coefficients of the 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 $g = 1, G$

The proposed approach

4. Estimation of the group dependence structure

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

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. Estimation of the group dependence structure

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

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)

A working example: 2 groups

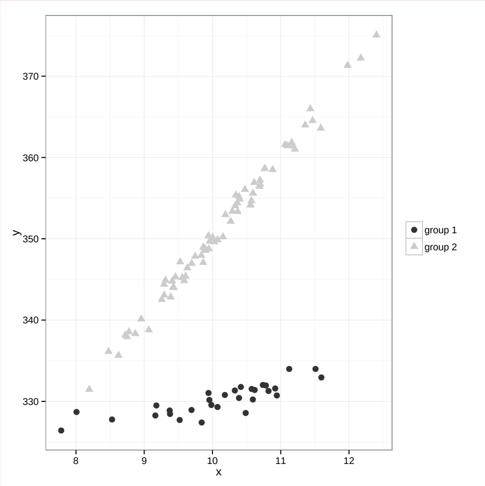
Structure of the two groups

	group 1	group 2
sample size	$n_1 = 30$	$n_2 = 70$
regressor	$\mathbf{x}_1 \sim N(10;1)$	$\mathbf{x}_2 \sim N(10;1)$
error	$\mathbf{e}_1 \sim N(0;1)$	$\mathbf{e}_2 \sim N(0;1)$
response variable	$\mathbf{y}_1$=310+2$\mathbf{x}_1$+$\mathbf{e}$	$\mathbf{y}_2$=250+10$\mathbf{x}_2$+$\mathbf{e}$

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \tag{1}$$

A working example

Structure of the two groups



A working example: 2 groups

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
   $\hat{\mathbf{y}}_\theta^{best}$

A working example: 2 groups

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
   $\theta_i : \underset{\theta=1,\ldots,k}{\text{argmin}} \; |y_i - \hat{y}_i(\theta)|$
3. best estimates identification
   $\hat{\mathbf{y}}_\theta^{best}$

A working example: 2 groups

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
   $\theta_i : \underset{\theta=1,\ldots,k}{\text{argmin}} \; |y_i - \hat{y}_i(\theta)|$
3. best estimates identification
   $\hat{\mathbf{y}}_\theta^{best}$

A working example: 2 groups

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
   $\theta_i : \underset{\theta=1,\ldots,k}{\text{argmin}} \; |y_i - \hat{y}_i(\theta)|$
3. best estimates identification
   $\hat{\mathbf{y}}_\theta^{best}$

A working example: 2 groups

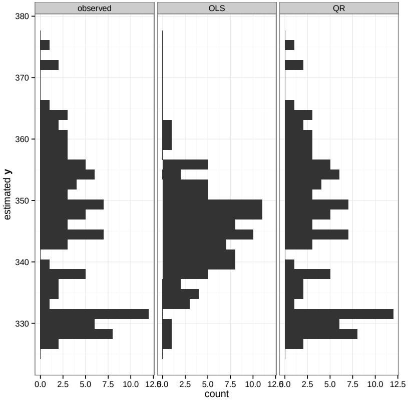
1. Global estimation

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

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)

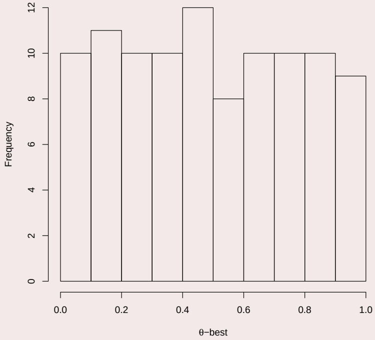


A working example: 2 groups

3.Identification of a partition

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

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

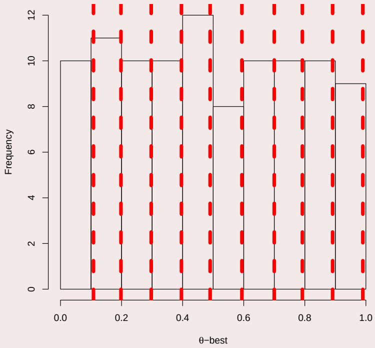


A working example: 2 groups

3.Identification of a partition

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

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



A working example: 2 groups

3.Identification of a partition

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

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

quantile	value	${}_d\overline{\theta}^{best}$
0.1	0.108	0.046
0.2	0.198	0.148
0.3	0.297	0.246
0.4	0.396	0.345
0.5	0.490	0.435
0.6	0.594	0.545
0.7	0.700	0.642
0.8	0.792	0.750
0.9	0.891	0.845

- estimate D quantile regression models

A working example: 2 groups

3.Identification of a partition

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

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

quantile	value	${}_d\overline{\theta}^{best}$	p-value
0.1	0.108	0.046	0.853
0.2	0.198	0.148	0.872
0.3	0.297	0.246	0.000
0.4	0.396	0.345	0.758
0.5	0.490	0.435	0.975
0.6	0.594	0.545	0.489
0.7	0.700	0.642	0.152
0.8	0.792	0.750	0.660
0.9	0.891	0.845	0.912

A working example: 2 groups

3.Identification of a partition

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

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

quantile	value	${}_d\overline{\theta}^{best}$	p-value	group	n_g
0.1	0.108	0.046	0.853	1	30
0.2	0.198	0.148	0.872		
0.3	0.297	0.246	0.000		
0.4	0.396	0.345	0.758	2	70
0.5	0.490	0.435	0.975		
0.6	0.594	0.545	0.489		
0.7	0.700	0.642	0.152		
0.8	0.792	0.750	0.660		
0.9	0.891	0.845	0.912		

A working example: 2 groups

3.Identification of a partition

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

- identification of the group reference quantile

quantile	value	$_d\overline{\theta}^{best}$	p-value	group	n_g	$_g\theta^{best}$
0.1	0.108	0.046	0.853	1	30	0.147
0.2	0.198	0.148	0.872			
0.3	0.297	0.246	0.000			
0.4	0.396	0.345	0.758	2	70	0.649
0.5	0.490	0.435	0.975			
0.6	0.594	0.545	0.489			
0.7	0.700	0.642	0.152			
0.8	0.792	0.750	0.660			
0.9	0.891	0.845	0.912			

A working example: 2 groups

4. Estimation of the group dependence structure

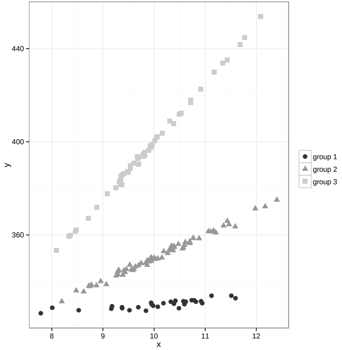
	$\theta = 0.145$ group 1	$\theta = 0.640$ group 2
intercept	313.11	248.19
x	1.71	10.19
original model	$\mathbf{y}_1$=310+2$\mathbf{x}_1$+$\mathbf{e}$	$\mathbf{y}_2$=250+10$\mathbf{x}_2$+$\mathbf{e}$

Percentage of Correct classification (%CC)=100%

A working example: 3 groups

Structure of the three groups

	group 1	group 2	group 3
sample size	$n_1 = 30$	$n_2 = 70$	$n_3 = 50$
regressor	$\mathbf{x}_1 \sim N(10;1)$	$\mathbf{x}_2 \sim N(10;1)$	$\mathbf{x}_3 \sim N(10;1)$
error	$\mathbf{e}_1 \sim N(0;1)$	$\mathbf{e}_2 \sim N(0;1)$	$\mathbf{e}_3 \sim N(0;1)$
response variable	$\mathbf{y}_1$=310+2$\mathbf{x}_1$+$\mathbf{e}$	$\mathbf{y}_2$=250+10$\mathbf{x}_2$+$\mathbf{e}$	$\mathbf{y}_3$=150+25$\mathbf{x}_2$+$\mathbf{e}$

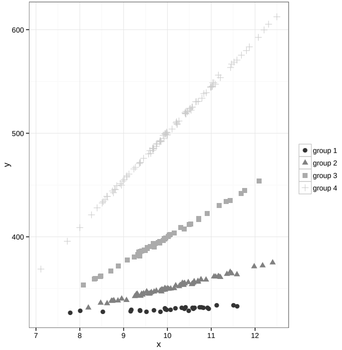


%CC=97%

A working example: 4 groups

Structure of the four groups

	group 1	group 2	group 3	group 4
sample size	$n_1 = 30$	$n_2 = 70$	$n_3 = 50$	$n_4 = 100$
regressor	$\mathbf{x}_1 \sim N(10; 1)$	$\mathbf{x}_2 \sim N(10; 1)$	$\mathbf{x}_3 \sim N(10; 1)$	$\mathbf{x}_4 \sim N(10; 1)$
error	$\mathbf{e}_1 \sim N(0; 1)$	$\mathbf{e}_2 \sim N(0; 1)$	$\mathbf{e}_3 \sim N(0; 1)$	$\mathbf{e}_4 \sim N(0; 1)$
response variable	$\mathbf{y}_1$=310+2$\mathbf{x}_1$+$\mathbf{e}_1$	$\mathbf{y}_2$=250+10$\mathbf{x}_2$+$\mathbf{e}_2$	$\mathbf{y}_3$=150+25$\mathbf{x}_3$+$\mathbf{e}_3$	$\mathbf{y}_4$=50+45$\mathbf{x}_4$+$\mathbf{e}_4$



%CC=97%

1000 replications of the data for each case

	group size	average %CC	wrong partition %
2 Groups	30; 70	97%	3% $\rightarrow$ 3 groups
3 Groups	30; 70; 50	95%	0.5% $\rightarrow$ 2 groups 3% $\rightarrow$ 4 groups
4 Groups	30; 70; 50; 100	96%	1% $\rightarrow$ 2 groups 20% $\rightarrow$ 3 groups 2% $\rightarrow$ 5 groups

An empirical analysis

The aim of the analysis

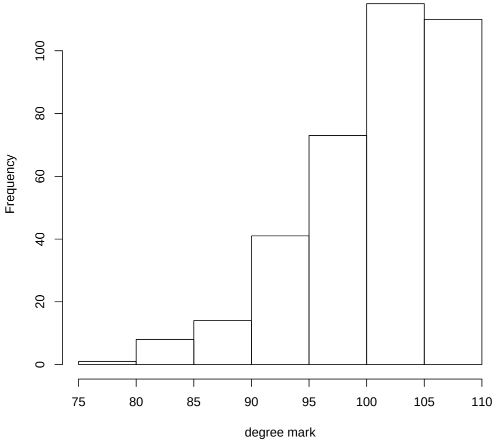
Evaluate if and how
the **student features**
(socio-demographic and University experience attributes)
affect the **outcome** of the University career (degree mark) in
case of unobserved group heterogeneity

The dataset

The evaluation of University educational processes

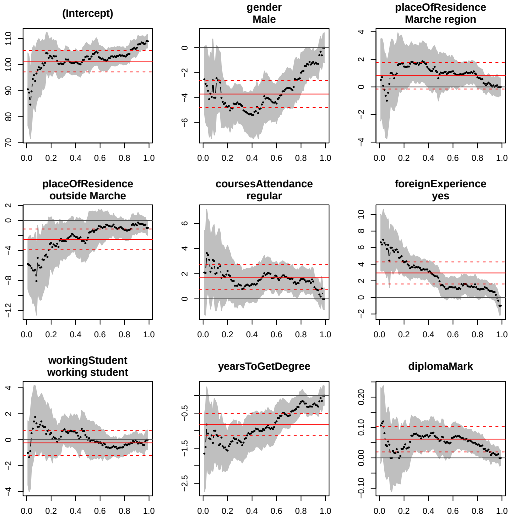
- random sample of 362 students graduated at University of Macerata (Italy)
- dependent variable: degree mark (110 scores excluded)
- 7 regressors related to the student profile:
  - gender
  - place of residence during University (Macerata and its province, Marche region, outside Marche)
  - course attendance (no attendance, regular)
  - foreign experience (yes, no)
  - working condition (full time student, working student)
  - number of years to get a degree
  - diploma mark

The dataset



Distribution of the dependent variable

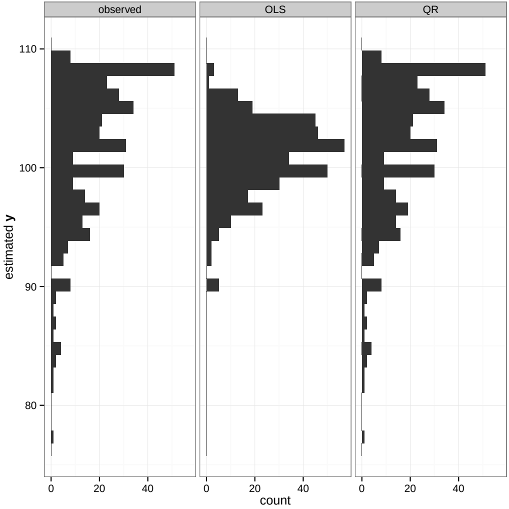
1.Identification of the global dependence structure



LS and QR coefficients

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

Step 2: Identification of the best model for each unit



Distribution of the:

- dependent variable (*left panel*)
- LS estimated dependent variable (*middle panel*)
- best QR estimated dependent variable (*right panel*)

Step 2:

- 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: $\hat{\mathbf{y}}_\theta^{best}$

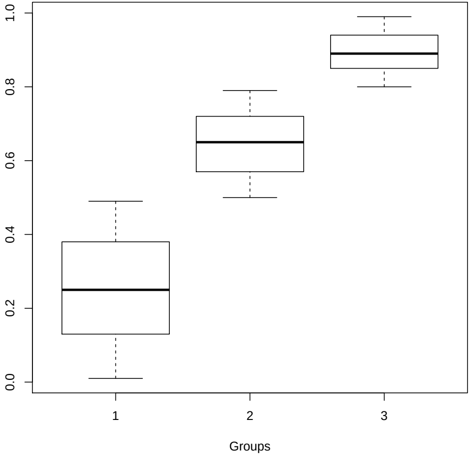
Step 3: Identification of a partition

quantile	value	${}_d\overline{\theta}^{best}$	p-value	group	n_g	${}_g\theta^{best}$
0.1	0.090	0.036	0.412	1	182	0.246
0.2	0.190	0.145	0.170			
0.3	0.293	0.250	0.842			
0.4	0.400	0.341	0.631			
0.5	0.490	0.444	0.000			
0.6	0.596	0.547	0.322	2	109	0.650
0.7	0.690	0.636	0.168			
0.8	0.790	0.747	0.008			
0.9	0.889	0.844	0.298	3	71	0.896

Step 3:

- partitioning of θ^{best}
- identification of the group reference quantile

Step 3: Identification of a partition

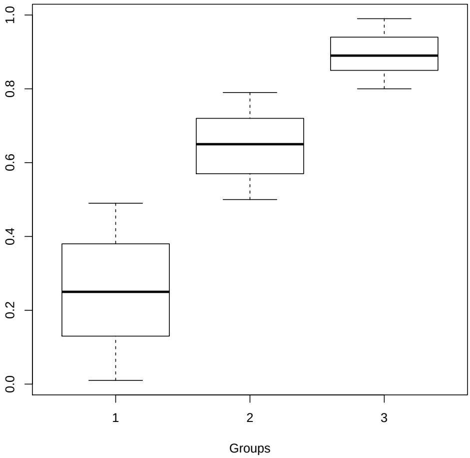


Distribution of the "best" quantiles in the groups

Step 3:

- partitioning of θ^{best}
- identification of the group reference quantile

Step 3: Identification of a partition



Reference 'best" quantile for each group:

Mean value of the "best" quantiles assigned to units belonging to the g^{th} group

- θ_1^{best}=0.246
- θ_2^{best}=0.649
- θ_3^{best}=0.896

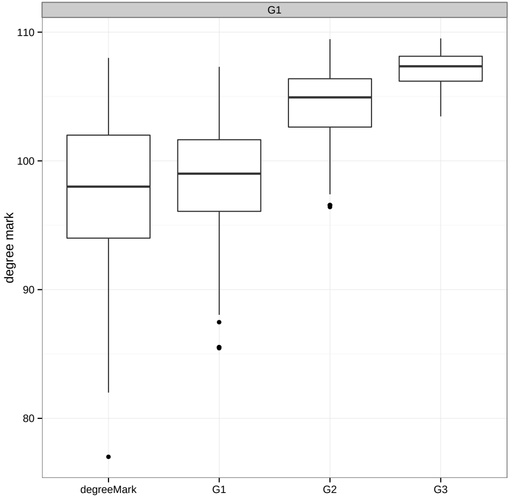
Step 4: Estimation of the group dependence structure

QR coefficients with group effects

Variable	OLS	G1	G2	G3
		$\theta = 0.246$	$\theta = 0.649$	$\theta = 0.896$
Intercept	**101.35**	**102.74**	**101.43**	**106.43**
gender (Male)	**-3.71**	**-5.04**	**-3.61**	**-1.14**
place of residence (Marche region)	0.81	**1.64**	**0.88**	0.25
place of residence (outside Marche)	**-2.53**	**-3.60**	-0.63	-0.64
courses attendance (regular)	**1.72**	0.99	**1.83**	**1.40**
foreign experience (yes)	**2.95**	**3.38**	1.09	0.76
working student	-0.24	-0.17	-0.49	-0.14
years to get a degree	**-0.83**	**-1.22**	**-0.52**	**-0.25**
diploma mark	**0.06**	0.04	**0.07**	0.02

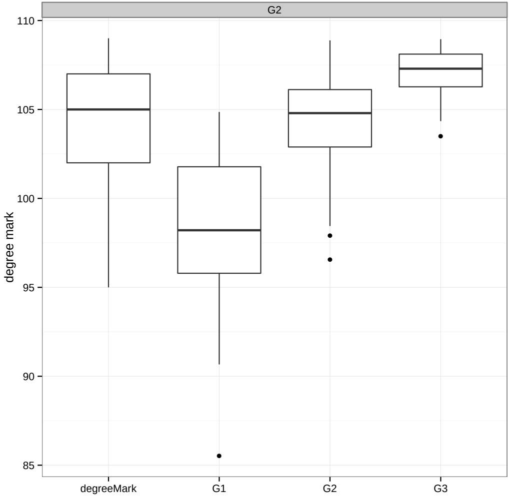
Step 4: $Q_\theta(\hat{\mathbf{y}}|\mathbf{X}) = \mathbf{X}\hat{\mathbf{B}}({}_g\theta^{best})$

Step 4: Estimation of the group dependence structure



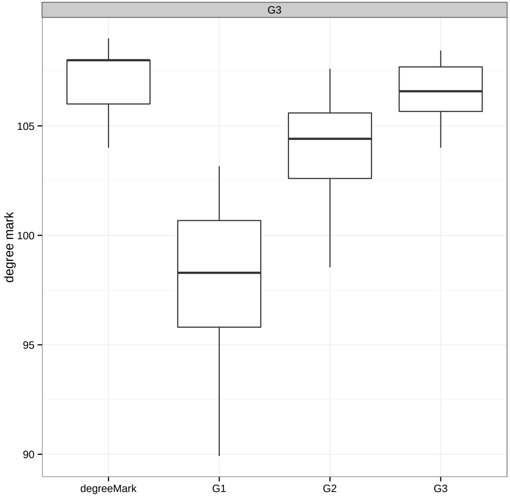
Group 1
Observed response distribution compared with the estimated distributions using the reference quantile of G1

Step 4: Estimation of the group dependence structure



Group 2
Observed response distribution compared with the estimated distributions using the reference quantile of G2

Step 4: Estimation of the group dependence structure



Group 3
Observed response distribution compared with the estimated distributions using the reference quantile of G3

Step 5: Testing differences among groups

Testing if all the slope coefficients of the groups are identical

p-values

	G1	G2	G3	G1;G2;G3
G1		0.001021	0.000000	
G2			0.000329	
G3				0.000000

Separate testing on each slope coefficient

	g1 vs g2	g2 vs g3	g1 vs g3
gender (Male)	0.114	0.003	0.000
place of residence (Marche region)	0.202	0.131	0.024
place of residence (outside Marche)	0.051	0.990	0.081
courses attendance (regular)	0.253	0.484	0.599
foreign experience (yes)	0.005	0.646	0.000
working student	0.609	0.436	0.969
years to get a degree	0.008	0.115	0.000
diploma mark	0.341	0.006	0.549

Concluding remarks

- Clustering units taking into account the dependence structure
- Estimation of the group dependence structure
- Impact of the regressors on the entire conditional distribution
- Clarity of the final results
- Availability of classical inferential procedures to test differences among groups

Further developments

- Explore alternatives to partition th θ_{best} vector
- Comparison with competitive methods
- Introduce cluster validation statistics
- Simulation study to explore robustness with respect to:
  - the degree and type of overlapping among the groups
  - the cardinality of each group (equal or unbalanced)
  - the sample size (from 100 to 1000 with step equal to 100)

Main references

Davino C., Furno M., Vistocco D. (2013) *Quantile Regression: Theory and Applications*. Wiley.

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. (2008) Quantile regression for the evaluation of student satisfaction. *Italian Journal of Applied Statistics* **20**, 179–196.

Davino C., Romano R., Naes T. (2015) The use of quantile regression in consumer studies. *Food Quality and Preference*, 40, pp. 230–239, Elsevier.

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.

Main references

- Davino C. (2016) Combining PLS path modeling and quantile regression for the evaluation of customer satisfaction. *Italian Journal of Applied Statistics*, vol. 26(2), pp. 93–116
- Davino C., Esposito Vinzi V. (2016) Quantile Composite-based Path modeling. *Advances in Data Analysis and Classification. Theory, Methods, and Applications in Data Science* .DOI 10.1007/s11634-015-0231-9
- Davino C., Esposito Vinzi V., Dolce, P. (2016) Assessment and validation in quantile composite-based path modeling. In Abdi, H., Esposito Vinzi, V., Russolillo, G., Saporta, G., and Trinchera, L. (eds) *The Multiple Facets of Partial Least Squares Methods*, Springer Proceedings in Mathematics & Statistics, Springer
- Davino C., S. Taralli, Dolce P., Esposito Vinzi V.(accepted) A quantile composite-indicator approach for the measurement of equitable and sustainable well-being: a case study of the Italian provinces *Social Indicators Research*.
- Davino C., Romano R., Naes T. (2015) The use of quantile regression in consumer studies. *Food Quality and Preference*, 40, pp. 230–239, Elsevier.

Main references

- Furno M., Vistocco D. (2015) A comparison among estimators for linear regression methods. In: *Cladag 2015, 10th Scientific Meeting of the Classification and Data Analysis Group*. e-Book of Short Paper.
- Furno M., Vistocco D. (2013) Qu test for structural breaks in quantile regression. *International Journal of Statistics and Probability*, vol. 2: 44-45.
- Conversano, C. e D. Vistocco (2010) Analysis of Mutual Funds' Management Styles: A Modeling, Ranking and Visualizing Approach. *JOURNAL OF APPLIED STATISTICS* 37, 1825-1845.
- Attardi, L. e D. Vistocco (2010) An index for ranking financial portfolio according to internal turnover In: *Data Analysis and Classification*. A cura di C. L. F. P. G. M. EDS. BERLIN HEIDELBERG: Springer, pp. 445-452. doi: 10.1007/978-3-642-03739-950.
- La Rocca, M. e D. Vistocco (2010) Robust estimation of style analysis coefficients. In: *Mathematical and Statistical Methods for Actuarial Sciences and Finance*. I. C. M. P. C. EDS. Milano: Springer-Verlag, pp. 163-172. doi: 10.1007/978-88-470-1481-7.
- Attardi, L. e D. Vistocco (2008) Estimating portfolio conditional returns distribution through style analysis models. In: *Mathematical and Statistical Methods for Insurance and Finance* P. C. S. M. EDS. MILAN: Springer, pp. 11-17. doi: 10.1007/978-88-470-0704-8$_2$.

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:

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

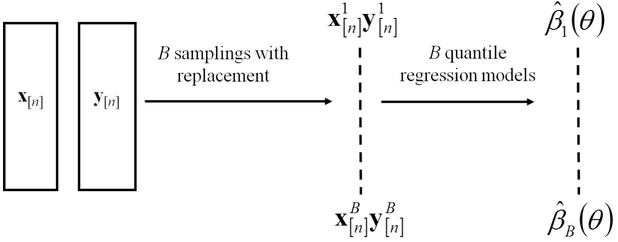
where $var\left[\hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j)\right] =$
$var\left[\hat{\beta}_1(\theta_i)\right] + var\left[\hat{\beta}_1(\theta_j)\right] - 2cov\left[\hat{\beta}_1(\theta_i)\hat{\beta}_1(\theta_j)\right]$

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

xy-pair method: a single quantile θ

Simple quantile regression model

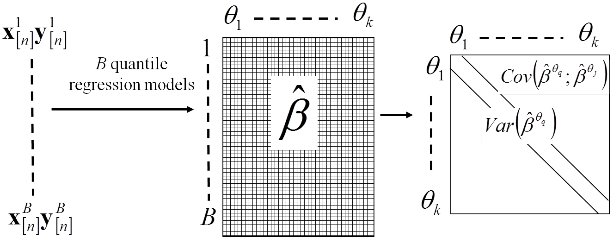
$$Q_\theta(\hat{\mathbf{y}}|\mathbf{x}) = \hat{\beta}_0 + \hat{\beta}_1(\theta)\mathbf{x} \tag{3}$$



Bootstrap estimate: $\overline{\hat{\beta}}(\theta) = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}_b(\theta)$

Bootstrap standard error: $se\left(\hat{\beta}_j(\theta_q)\right)$

xy-pair method: *k* quantiles



Estimated variance covariance matrix:

$$\hat{V}_{q,j} = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\beta}_{b,j}(\theta_q) - \overline{\hat{\beta}}_j(\theta_q)\right)\left(\hat{\beta}_{b,j}(\theta_q) - \overline{\hat{\beta}}_j(\theta_q)\right)^\top \quad (4)$$

where $j = 1, \ldots, p$; $q = 1, \ldots, k$ bootstrap estimates:

$\overline{\hat{\beta}}_j(\theta_q) = \frac{1}{B}\sum_{b=1}^{B}\hat{\beta}_{b,j}(\theta_q)$