Handling heterogeneity in Quantile Regression

Cristina Davino
Department of Economics and Statistics
University of Naples Federico II
Italy
cristina.davino@unina.it
Cnam - Paris
April 5, 2019

Outline

Heterogeneity

Part 1: Quantile Regression
- Basic insights
- Estimation
- Inference
- Properties
- Assessment

Part 2: My recent research on handling heterogeneity
- Unsupervised approach
- Supervised approach
- Quantile Composite-based Path Model

All computations and graphics were done in the R language using the packages quantreg and plspm.
Heterogeneity

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

Part 1: Quantile Regression
Motivation

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

The flaw of Averages: a rationale for quantile regression

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.
Quantile regression

- QR has become a popular alternative to least squares regression for modeling heterogeneous data
- QR gained popularity in applied economics by the end of the 90’s, when people realize the importance of heterogeneity
- Fields of application:
  - astrophysics
  - chemistry
  - ecology
  - economics
  - finance
  - genomics
  - medicine
  - meteorology
  - sociology
  - marketing
  - food science

Classical linear regression

- Classical linear regression (conditional expected value)
  estimation of the conditional mean of a response variable (y) distribution as a function of a set X of predictor variables

Classical vs quantile linear regression

- Classical linear regression (conditional expected value)
  estimation of the conditional mean of a response variable (y) distribution as a function of a set X of predictor variables
  \[ E(y \mid X) = X\beta \]

- Quantile regression (conditional quantiles)
  estimation of the conditional quantiles of a response variable (y) distribution as a function of a set X of predictor variables
  \[ Q_\theta(y \mid X) = X\beta(\theta) \]
  where: \( 0 < \theta < 1 \)

- Classical regression focuses on \( E(y \mid X) \)
- QR extends this approach to study the conditional distribution of a response variable
- \( \theta \) regression lines are estimated
- The estimation of coefficients for each quantile regression is based on the whole sample, not just the portion of the sample at that quantile

Quantile Regression model

QR model for a given conditional quantile \( \theta \) (linear regression):

\[ Q_\theta(y \mid X) = X\beta(\theta) \]

where

- \( 0 < \theta < 1 \)
- \( Q_\theta(\cdot \mid \cdot) \) denotes the conditional quantile function for the \( \theta \)th quantile
Two examples with simulated data

**homogeneous model**
\[ y_1 = 1 + 2x + e \]
\[ x \sim N(10; 1) \quad e \sim N(0; 1) \]

**heterogeneous model**
\[ y_2 = 1 + 2x + (1 + x)e \]
\[ x \sim N(10; 1) \quad e \sim N(-1 + 20x; e^{x/3}) \]

OLS results

**homogeneous model**

**heterogeneous model**

OLS and QR results

**homogeneous model**

**heterogeneous model**

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>OLS</th>
<th>( \theta = 0.1 )</th>
<th>( \theta = 0.25 )</th>
<th>( \theta = 0.5 )</th>
<th>( \theta = 0.75 )</th>
<th>( \theta = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>0.5</td>
<td>-0.5</td>
<td>-0.7</td>
<td>0.4</td>
<td>1.6</td>
<td>1.2</td>
</tr>
<tr>
<td>( x )</td>
<td>2.0</td>
<td>2.0</td>
<td>2.1</td>
<td>2.1</td>
<td>2.0</td>
<td>2.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>OLS</th>
<th>( \theta = 0.1 )</th>
<th>( \theta = 0.25 )</th>
<th>( \theta = 0.5 )</th>
<th>( \theta = 0.75 )</th>
<th>( \theta = 0.9 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>2092.0</td>
<td>-697.2</td>
<td>-1312.7</td>
<td>-1772.2</td>
<td>-2340.6</td>
<td>-2709.7</td>
</tr>
<tr>
<td>( x )</td>
<td>432.1</td>
<td>247.1</td>
<td>331.8</td>
<td>398.3</td>
<td>480.4</td>
<td>538.3</td>
</tr>
</tbody>
</table>
OLS and QR results

**homogeneous model**

<table>
<thead>
<tr>
<th>Intercept</th>
<th>quantiles</th>
<th>coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.5</td>
<td>0.0</td>
<td>0.5</td>
</tr>
<tr>
<td>1.0</td>
<td>1.5</td>
<td>2.0</td>
</tr>
<tr>
<td>2.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slope</th>
<th>quantiles</th>
<th>coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8</td>
<td>1.9</td>
<td>2.0</td>
</tr>
<tr>
<td>2.1</td>
<td>2.2</td>
<td></td>
</tr>
</tbody>
</table>

**heterogeneous model**

<table>
<thead>
<tr>
<th>Intercept</th>
<th>quantiles</th>
<th>coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2500</td>
<td>-2000</td>
<td>-1500</td>
</tr>
<tr>
<td>-1000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Slope</th>
<th>quantiles</th>
<th>coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>350</td>
<td>450</td>
</tr>
<tr>
<td>550</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Quantile Regression model

**Interpretation**

\[
\hat{\beta}_i(\theta) = \frac{\partial Q_\theta(y|X)}{\partial x_i}
\]

- Rate of change of the \(\theta\)th quantile of the dependent variable per unit change in the value of the \(i\)th quantile
- Fitted values reconstruct the conditional quantiles
- QR generalizes univariates quantiles for conditional distributions

**QR pros:**

- Regressor effects on the whole dependent variable distribution
- Heteroscedastic relationships
- Presence of outliers
- Skewed dependent variable

A simple example: the ‘93cars’ dataset

- 93 new cars for the 1993 model year
- Selected measures: Price, Origin (USA, non-USA), Horsepower

<table>
<thead>
<tr>
<th>Price</th>
<th>USA</th>
<th>non-USA</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>40</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

Slopes: rate of change of the \(y\) \(\theta\)th conditional quantile per unit change of the regressor

Fitted values reconstruct the conditional quantiles

QR generalizes univariates quantiles for conditional distributions

<table>
<thead>
<tr>
<th>USA</th>
<th>non-USA</th>
<th>uncond.</th>
<th>intercept</th>
<th>slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>18.6</td>
<td>20.5</td>
<td>19.5</td>
<td>18.6</td>
</tr>
<tr>
<td>(\theta=0.25)</td>
<td>13.5</td>
<td>11.6</td>
<td>12.2</td>
<td>13.4</td>
</tr>
<tr>
<td>(\theta=0.5)</td>
<td>16.3</td>
<td>19.1</td>
<td>17.7</td>
<td>16.3</td>
</tr>
<tr>
<td>(\theta=0.75)</td>
<td>20.7</td>
<td>26.7</td>
<td>23.3</td>
<td>20.8</td>
</tr>
</tbody>
</table>
The quantile process and the selection of the quantiles

- QR solutions are typically computed for a selected number of quantiles
- It is possible to obtain estimates across the entire interval of conditional quantiles
- A dense grid of equally spaced quantiles provides a fairly accurate approximation of the whole quantile regression pattern
- The number of distinct quantiles is related to: the number of units and the number of variables

Unconditional mean and quantiles

QR is to classical regression what quantiles are to mean in terms of describing locations of a distribution

Let \( Y \) be a generic random variable:
- Mean (and its objective function): \( \mu = \arg \min_c E(Y - c)^2 \)
- Median (and its objective function): \( Me = \arg \min_c E|Y - c| \)
- Generic quantile \( \theta \) (and its objective function):
  \[
  q_\theta = \arg \min_c E[\rho_\theta(Y - c)]
  \]

- \( \hat{\mu} \) and \( Me \) denote the sample estimators for such centers
- \( \rho_\theta(\cdot) \) denotes the following location functions:
  \[
  \rho_\theta(y) = [\theta - I(y < 0)]y \\
  = [(1 - \theta)I(y \leq 0) + \theta I(y > 0)]y
  \]
- \( \rho_\theta(\cdot) \) is an asymmetric absolute loss function; that is a weighted sum of absolute deviations, where a \((1 - \theta)\) weight is assigned to the negative deviations and a \(\theta\) weight to the positive deviations.

Part 1: Quantile Regression

Estimation

On optimal criteria

![Graph showing objective functions for different quantile regression patterns]
Conditional mean and conditional quantiles estimation

Least squares linear regression estimator

$$\hat{\beta} = \arg \min_{\beta} E [(y - X\beta)^2]$$

Conditional quantile linear regression estimator

$$\hat{\beta}(\theta) = \arg \min_{\beta} E [\rho_{\theta}(y - X\beta)]$$

Note: The \((\theta)\)-notation denotes that the parameters and the corresponding estimators are for a specific quantile \(\theta\).

\(\rho_{\theta}(.)\) is an asymmetric absolute loss function; that is a weighted sum of absolute deviations, where a \((1 - \theta)\) weight is assigned to the negative deviations and a \(\theta\) weight is used for the positive deviations.

$$\rho_{\theta} = \begin{cases} \theta(u) & \text{if } u > 0 \\ (\theta - 1)u & \text{if } u \leq 0 \end{cases}$$

On the objective function

\(\theta = 0.25\)

- 75% of points above the QR line and 25% below
- unbalanced weighting system: 0.75 (0.25) for sum of negative (positive) deviations
- \(m=2\) points lies exactly on the line \((m=\text{number of model parameters})\)

On the objective function

\(\theta = 0.75\)

- 25% of points above the QR line and 75% below
- unbalanced weighting system: 0.25 (0.75) for sum of negative (positive) deviations
- \(m=2\) points lies exactly on the line \((m=\text{number of model parameters})\)

The linear programming formulation of the QR problem

- Wagner (1959) proved that the least absolute deviation criterion can be formulated as a linear programming technique and then solved efficiently exploiting proper methods and algorithms
- Koenker and Basset (1978) pointed out how conditional quantiles could be estimated by an optimization function minimizing a sum of weighted absolute deviations, using weights as asymmetric functions of the quantiles
- The linear programming formulation of the problem was therefore natural, offering researchers and practitioners a tool for looking inside the whole conditional distribution apart from its center
Methods for solving the linear programming problem

- The simplex method (Dantzig, 1947) is the widespread solution for the linear programming problem.
- It is an iterative process, starting from a solution that satisfies the imposed constraints and looking for new and better solution.
- The process iterates until a solution that cannot be further improved is reached, moving along the edges of the simplex corresponding to the feasible set.
- For the QR problem, the efficient version of the simplex algorithm, proposed by Barrodale and Roberts (1974) and adapted by Koenker and D'Orey (1987) to compute conditional quantiles, is typically used with a moderate size problem.
- The simplex method is the default option in most of the QR software.
- A completely different method approaches the solution from the interior of the feasible set rather than on its boundary, that is starting in the zone where all the inequalities are strictly satisfied.
- Such methods, called interior–point methods, have their roots in the seminal paper of Karmakar (1984) and are usually superior on very large problems.
- The QR solution using interior–point methods has been proposed by Portnoy and Koenker (1997).
- A heuristic approach (finite smoothing algorithm) has been proposed by Chen (2004, 2007): it is faster and more accurate in the presence of a large number of covariates.

Main approaches to inference in QR

- Small sample theory
  (Koenker and Basset, 1978)
  “The practical of this theory would entail a host of hazardous assumptions and an exhausting computational effort” (Koenker, 2005)

- Asymptotic theory
  (Koenker and Basset, 1978, 1982a,b)

- Rank–based theory
  (Gutenbrunner and Jureckova, 1992) (Gutenbrunner, 1993)

- Resampling methods
Asymptotic theory

\[ Q_\theta(\hat{y} | x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x \]

“under mild regularity conditions”

Asymptotic distribution of the estimator:

- case of i.i.d. errors

\[ \sqrt{n} \left[ \hat{\beta}(\theta) - \beta(\theta) \right] \rightarrow N \left( 0, \sigma^2(\theta) J^{-1} \right) \]

- case of i.n.i.d. errors

\[ \sqrt{n} \left[ \hat{\beta}(\theta) - \beta(\theta) \right] \rightarrow N \left( 0, \theta (1 - \theta) H(\theta)^{-1} J H(\theta)^{-1} \right) \]

The error distribution affects the variance–covariance matrix of the QR estimator.

Main approaches to inference in QR

Asymptotic theory

\[ \frac{\hat{\beta}(\theta) - \beta(\theta)}{SE(\hat{\beta}(\theta))} \rightarrow N(0, 1) \]

- standard errors are simpler and easier to describe under the i.i.d. model
- it is quite complex to deal with the ni.i.d. case, as the errors no longer have a common distribution

Bootstrap approach

- useful when the assumptions for the asymptotic procedure do not hold
- easy to compute standard errors
- flexible to obtain standard error and confidence interval for any estimates and combinations of estimates

Resampling methods in QR

- *xy-pair* or *design matrix bootstrap method* (Kocherginsky, 2003)

- method based on pivotal estimation functions (Parzen, 1979)

- markov chain marginal bootstrap (He and Hu, 2002) (Kocherginsky et al. 2005)

**xy-pair method: a single quantile \( \theta \)**

Simple quantile regression model

\[ Q_\theta(\hat{y} | x) = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x \]  

(1)

Bootstrap estimate:

\[ \bar{\beta}(\theta) = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_b(\theta) \]

Bootstrap standard error: \( se\left( \bar{\beta}(\theta_q) \right) \)
Equivariance properties

Part 1: Quantile Regression

Equivariance properties

- scale equivariance
- shift or regression equivariance
- equivariance to reparametrization of design
- equivariance to monotone transformations

Equivariance to monotone transformations

\[ Q_{\theta} (\hat{y} | x) = \hat{\beta}_0 (\theta) + \hat{\beta}_1 (\theta) x \]

where \( h(.) \) is a non decreasing function in \( \mathbb{R} \)

\[ Q_{\theta} \left[ h(\hat{y}) | x \right] = h \left( Q_{\theta} (\hat{y} | x) \right) \]

- The quantiles of the transformed \( y \) variable are the transformed quantiles of the original ones
- appropriate selection of \( h(.) \) corrects different kinds of skewness
- The logarithmic transformation might be very hazardous in terms of the inference results of an OLS regression (Manning 1998) whereas it may aid the statistical inference of QR (Cade and Noon 2003)
Quantile regression models are estimated minimizing the absolute values of weighted residuals, as opposed to minimizing the sum of squared errors in OLS.

- The R2 is not an applicable goodness-of-fit measure.
- Methods available for evaluating goodness-of-fit in quantile regression allow to compare model fit among nested model but they are not comparable to standard coefficients of determination.


Part 1: Quantile Regression

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

OLS and QR coefficients

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>(\theta=0.10)</th>
<th>(\theta=0.25)</th>
<th>(\theta=0.50)</th>
<th>(\theta=0.75)</th>
<th>(\theta=0.90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>101.78</td>
<td>100.12</td>
<td>101.08</td>
<td>102.19</td>
<td>103.60</td>
<td>106.45</td>
</tr>
<tr>
<td>Gender = Male</td>
<td>-3.42</td>
<td>-1.94</td>
<td>-3.92</td>
<td>-4.12</td>
<td>-2.60</td>
<td>-4.12</td>
</tr>
<tr>
<td>Place of residence = Marche region</td>
<td>0.95</td>
<td>0.89</td>
<td>1.69</td>
<td>1.33</td>
<td>1.05</td>
<td>0.17</td>
</tr>
<tr>
<td>Place of Residence = outside Marche</td>
<td>-2.51</td>
<td>-1.27</td>
<td>-1.42</td>
<td>-0.88</td>
<td>-0.35</td>
<td>-0.17</td>
</tr>
<tr>
<td>Courses attendance = regular</td>
<td>1.87</td>
<td>2.52</td>
<td>0.92</td>
<td>2.34</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Working student = yes</td>
<td>-0.20</td>
<td>0.06</td>
<td>0.42</td>
<td>-0.21</td>
<td>-0.60</td>
<td>-0.31</td>
</tr>
<tr>
<td>Numbers of years to get a degree</td>
<td>-0.82</td>
<td>-1.27</td>
<td>-1.42</td>
<td>-0.88</td>
<td>-0.35</td>
<td>-0.17</td>
</tr>
<tr>
<td>Diploma mark</td>
<td>0.06</td>
<td>0.01</td>
<td>0.08</td>
<td>0.07</td>
<td>0.05</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Outline

Heterogeneity

Part 1: Quantile Regression
- Basic insights
- Estimation
- Inference
- Properties
- Assessment

Part 2: My recent research on handling heterogeneity
- Unsupervised approach
- Supervised approach
- Quantile Composite-based Path Model
Outline

Heterogeneity

Part 1: Quantile Regression
- Basic insights
- Estimation
- Inference
- Properties
- Assessment

Part 2: My recent research on handling heterogeneity
- Unsupervised approach
- Supervised approach
- Quantile Composite-based Path Model

A simple example

Percentage of votes for democrats for 50 states of the US

http://rcarbonneau.com/ClusterwiseRegressionDatasets.htm
A simple example

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?
A simple example

![Percentage of votes for democrats for 50 states of the US](image)

**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?

**The main steps**

- Identification of the global dependence structure
- Identification of the best model for each unit
- Clustering units
- Modeling groups
- Testing differences among groups

**Basic notation**

**The data structure**

- \(n\) units
- \(p\) regressors
- 1 quantitative or ordinal dependent variable

\[
y_{[n]} = f(x_{[n \times p]})
\]
Basic notation

The data structure
- $n$ units
- $p$ regressors
- 1 quantitative or ordinal dependent variable

A working example: 2 groups

Structure of the two groups

<table>
<thead>
<tr>
<th></th>
<th>group 1</th>
<th>group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample size</td>
<td>$n_1 = 30$</td>
<td>$n_2 = 70$</td>
</tr>
<tr>
<td>regressor</td>
<td>$x_1 \sim N(10; 1)$</td>
<td>$x_2 \sim N(10; 1)$</td>
</tr>
<tr>
<td>error</td>
<td>$e_1 \sim N(0; 1)$</td>
<td>$e_2 \sim N(0; 1)$</td>
</tr>
<tr>
<td>response variable</td>
<td>$y_1 = 310 + 2x_1 + e$</td>
<td>$y_2 = 250 + 10x_2 + e$</td>
</tr>
</tbody>
</table>

$$
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}
$$

(2)

A working example

Structure of the two groups

The proposed approach

1. Identification of the global dependence structure

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

2. Identification of the best model for each unit

   - estimated values
     $$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{B}}(\theta)$$
   - best model identification
     $$\theta_{j}^\text{best} : \arg \min_{\theta = 1, \ldots, k} |y_j - \hat{y}_j(\theta)|$$
   - best estimates identification
     $$\hat{\mathbf{y}}_{\theta}^\text{best}$$
A working example: 2 groups

1. Global estimation

\[ Q(\hat{y}|X) = XB(\theta) \]

2. Identification of the best model for each unit

- estimated values
  \[ \hat{Y} = XB(\theta) \]
- best model identification
  \[ \theta_i : \arg\min_{\theta=1,\ldots,k} |y_i - \hat{y}_i(\theta)| \]
- best estimates identification
  \[ \hat{y}_{best} \]

The proposed approach

3. Clustering units

- finding the best partition of the \( \theta_{best} \) vector
- identification of the group reference quantile
  \[ \theta_{best}^g, ~ \text{for} ~ g = 1, G \]

3. Clustering units

Finding the best partition of the \( \theta_{best} \) vector

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

\[ \theta_{best}^g = \frac{1}{n_d} \sum_{i=1}^{n_d} \theta_i \]

\( (d = 1, \ldots, D) \)
- estimate D quantile regression models with \( \theta = [1\theta_{best}^1, \ldots, D\theta_{best}^D] \)
3. Clustering units

Finding the best partition of the $\theta^{\text{best}}$ vector

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

<table>
<thead>
<tr>
<th>quantile</th>
<th>$d^{\theta_{\text{best}}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.108</td>
</tr>
<tr>
<td>0.2</td>
<td>0.198</td>
</tr>
<tr>
<td>0.3</td>
<td>0.297</td>
</tr>
<tr>
<td>0.4</td>
<td>0.396</td>
</tr>
<tr>
<td>0.5</td>
<td>0.490</td>
</tr>
<tr>
<td>0.6</td>
<td>0.594</td>
</tr>
<tr>
<td>0.7</td>
<td>0.700</td>
</tr>
<tr>
<td>0.8</td>
<td>0.792</td>
</tr>
<tr>
<td>0.9</td>
<td>0.891</td>
</tr>
</tbody>
</table>

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

- estimate $D$ quantile regression models

---

Heteroscedasticity test

\[ Q_i(\hat{y}|x) = \hat{\beta}_0(\theta_i) + \hat{\beta}_1(\theta_i)x \]
\[ Q_j(\hat{y}|x) = \hat{\beta}_0(\theta_j) + \hat{\beta}_1(\theta_j)x \]

\[ H_0 : \beta_1(\theta_i) = \beta_1(\theta_j) \]

Test Statistic:

\[ T = \frac{\left[ \hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j) \right]^2}{\text{var} \left[ \hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j) \right]} \sim \chi^2_{gdf} \] (3)

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

A possible solution to estimate \[ \text{var} \left[ \hat{\beta}_1(\theta_i) - \hat{\beta}_1(\theta_j) \right] : \text{bootstrap} \]
A working example: 2 groups

3. Clustering units

Finding the best partition of the $\theta^{\text{best}}$ vector

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

<table>
<thead>
<tr>
<th>quantile</th>
<th>value</th>
<th>$\theta^{\text{best}}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.108</td>
<td>0.046</td>
<td>0.853</td>
</tr>
<tr>
<td>0.2</td>
<td>0.198</td>
<td>0.148</td>
<td>0.872</td>
</tr>
<tr>
<td>0.3</td>
<td>0.297</td>
<td>0.246</td>
<td>0.000</td>
</tr>
<tr>
<td>0.4</td>
<td>0.396</td>
<td>0.345</td>
<td>0.758</td>
</tr>
<tr>
<td>0.5</td>
<td>0.490</td>
<td>0.435</td>
<td>0.975</td>
</tr>
<tr>
<td>0.6</td>
<td>0.594</td>
<td>0.545</td>
<td>0.489</td>
</tr>
<tr>
<td>0.7</td>
<td>0.700</td>
<td>0.642</td>
<td>0.152</td>
</tr>
<tr>
<td>0.8</td>
<td>0.792</td>
<td>0.750</td>
<td>0.660</td>
</tr>
<tr>
<td>0.9</td>
<td>0.891</td>
<td>0.845</td>
<td>0.912</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>quantile</th>
<th>value</th>
<th>$\theta^{\text{best}}$</th>
<th>p-value</th>
<th>group</th>
<th>$n_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.108</td>
<td>0.046</td>
<td>0.853</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>0.2</td>
<td>0.198</td>
<td>0.148</td>
<td>0.872</td>
<td>2</td>
<td>70</td>
</tr>
<tr>
<td>0.3</td>
<td>0.297</td>
<td>0.246</td>
<td>0.000</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>0.4</td>
<td>0.396</td>
<td>0.345</td>
<td>0.758</td>
<td>4</td>
<td>70</td>
</tr>
<tr>
<td>0.5</td>
<td>0.490</td>
<td>0.435</td>
<td>0.975</td>
<td>5</td>
<td>70</td>
</tr>
<tr>
<td>0.6</td>
<td>0.594</td>
<td>0.545</td>
<td>0.489</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>0.7</td>
<td>0.700</td>
<td>0.642</td>
<td>0.152</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>0.8</td>
<td>0.792</td>
<td>0.750</td>
<td>0.660</td>
<td>8</td>
<td>70</td>
</tr>
<tr>
<td>0.9</td>
<td>0.891</td>
<td>0.845</td>
<td>0.912</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

C. Davino (University of Naples) Handling heterogeneity in QR Paris, April 2019 73 / 110

The proposed approach

4. Modeling groups

$$Q_0(\hat{y}|X) = XB(\theta^{\text{best}})$$

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

4. Modeling groups

<table>
<thead>
<tr>
<th></th>
<th>group 1</th>
<th>group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\theta)</td>
<td>0.145</td>
<td>0.640</td>
</tr>
<tr>
<td>intercept</td>
<td>313.11</td>
<td>248.19</td>
</tr>
<tr>
<td>x</td>
<td>1.71</td>
<td>10.19</td>
</tr>
</tbody>
</table>

original model \(y_1 = 310 + 2x_1 + e\) \(y_2 = 250 + 10x_2 + e\)

Percentage of Correct classification (%CC)=100%

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

Distribution of the dependent variable

**The dataset**

- 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

Distribution of the dependent variable

**The dataset**

- 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

Distribution of the dependent variable
1. Identification of the global dependence structure

Step 1: Identification of the global dependence structure

\[
Q_0(y|x) = x\hat{\theta}(q) \quad \theta = 1, \ldots, k
\]

Step 2: Identification of the best model for each unit

\[
\hat{y} = \hat{X}\hat{\theta}(\hat{\phi})
\]

Best model identification
\[
\hat{\phi}_{best} = \arg\min_{\phi_{\text{best}}} \sum_{i=1}^{n} |y_i - \hat{y}_i(\phi_{\text{best}})|
\]

Best estimates identification: \( \hat{y}_{best} \)

Step 3: Clustering units

<table>
<thead>
<tr>
<th>quantile</th>
<th>value</th>
<th>( \hat{\phi}_{best} )</th>
<th>p-value</th>
<th>group</th>
<th>( n_g )</th>
<th>( \hat{y}_{best} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.090</td>
<td>0.036</td>
<td>0.412</td>
<td>1</td>
<td>182</td>
<td>0.246</td>
</tr>
<tr>
<td>0.2</td>
<td>0.190</td>
<td>0.145</td>
<td>0.170</td>
<td>2</td>
<td>209</td>
<td>0.575</td>
</tr>
<tr>
<td>0.3</td>
<td>0.293</td>
<td>0.250</td>
<td>0.842</td>
<td>3</td>
<td>71</td>
<td>0.896</td>
</tr>
<tr>
<td>0.4</td>
<td>0.400</td>
<td>0.341</td>
<td>0.631</td>
<td>4</td>
<td>109</td>
<td>0.560</td>
</tr>
<tr>
<td>0.5</td>
<td>0.490</td>
<td>0.444</td>
<td>0.000</td>
<td>5</td>
<td>71</td>
<td>0.896</td>
</tr>
</tbody>
</table>

Step 3:
- partitioning of \( \hat{\phi}_{best} \)
- Identification of the group reference quantile \( \hat{y}_{best} \) for \( g = 1, G \)

Distribution of the “best” quantiles in the groups

Step 3:
- partitioning of \( \hat{\phi}_{best} \)
- Identification of the group reference quantile \( \hat{y}_{best} \), for \( g = 1, G \)
Step 3: Clustering units

Reference ‘best’ quantile for each group:
Mean value of the ‘best’ quantiles assigned to units belonging to the $g^{th}$ group
- $\theta_{1}^{\text{best}} = 0.246$
- $\theta_{2}^{\text{best}} = 0.649$
- $\theta_{3}^{\text{best}} = 0.896$

Step 4: Modeling groups

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>101.35</td>
<td>102.74</td>
<td>101.43</td>
<td>106.43</td>
</tr>
<tr>
<td>gender (Male)</td>
<td>-3.71</td>
<td>-5.04</td>
<td>-3.61</td>
<td>-1.14</td>
</tr>
<tr>
<td>place of residence (Marche region)</td>
<td>0.81</td>
<td>1.64</td>
<td>0.88</td>
<td>0.25</td>
</tr>
<tr>
<td>place of residence (outside Marche)</td>
<td>-2.53</td>
<td>-3.60</td>
<td>-0.63</td>
<td>-0.64</td>
</tr>
<tr>
<td>courses attendance (regular)</td>
<td>1.72</td>
<td>0.99</td>
<td>1.83</td>
<td>1.40</td>
</tr>
<tr>
<td>foreign experience (yes)</td>
<td>2.95</td>
<td>3.38</td>
<td>1.09</td>
<td>0.76</td>
</tr>
<tr>
<td>working student</td>
<td>-0.24</td>
<td>-0.17</td>
<td>-0.49</td>
<td>-0.14</td>
</tr>
<tr>
<td>years to get a degree</td>
<td>-0.83</td>
<td>-1.22</td>
<td>-0.52</td>
<td>-0.25</td>
</tr>
<tr>
<td>diploma mark</td>
<td>0.06</td>
<td>0.04</td>
<td>0.07</td>
<td>0.02</td>
</tr>
</tbody>
</table>

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

Group 2
Observed response distribution compared with the estimated distributions using the reference quantile of G2
Step 4: Modeling groups

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

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G1:G2;G3</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>0.001021</td>
<td>0.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>0.000329</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>0.000000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Separate testing on each slope coefficient

<table>
<thead>
<tr>
<th>Gender (Male)</th>
<th>g1 vs g2</th>
<th>g2 vs g3</th>
<th>g1 vs g3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.114</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>Place of residence (Marche region)</td>
<td>0.202</td>
<td>0.131</td>
<td>0.024</td>
</tr>
<tr>
<td>Place of residence (outside Marche)</td>
<td>0.051</td>
<td>0.990</td>
<td>0.081</td>
</tr>
<tr>
<td>Courses attendance (regular)</td>
<td>0.253</td>
<td>0.484</td>
<td>0.599</td>
</tr>
<tr>
<td>Foreign experience (yes)</td>
<td>0.005</td>
<td>0.646</td>
<td>0.000</td>
</tr>
<tr>
<td>Working student</td>
<td>0.609</td>
<td>0.436</td>
<td>0.969</td>
</tr>
<tr>
<td>Years to get a degree</td>
<td>0.008</td>
<td>0.115</td>
<td>0.000</td>
</tr>
<tr>
<td>Diploma mark</td>
<td>0.341</td>
<td>0.006</td>
<td>0.549</td>
</tr>
</tbody>
</table>

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

Further developments

- Explore alternatives to partition the $\theta_{best}$ vector
- Introduce cluster validation statistics
- Simulation study
- Comparison with competitive methods
A simulation study

Exploring the robustness of the method with respect to:

- the degree and type of overlapping among the groups;
- the cardinality of each group (equal or unbalanced);
- the sample size.

case of one regressor and two groups

Generation of a set of scenarios:

- Case 1: parallel group structures;
- Case 2: group structures crossing outside the considered range of the regressor;
- Case 3: group structures crossing inside the considered range of the regressor.

Comparison with competitive methods

Clusterwise linear regression

- It is a useful technique when heterogeneity is present in the data
- It identifies both the partition of the data and the relevant regression models, one for each cluster.
- It estimates simultaneously the classes and the parameters of the models which are considered different on each class
- Number of classes a-priori defined
- Not exact solution method
- Performance is sensitive to the initial partition and outliers
- Overlapping among groups

A comparison with the 'votes' dataset

The proposed approach

Best partition: 3 groups

- $\theta_{best}^1 = 0.18$
- $\theta_{best}^2 = 0.49$
- $\theta_{best}^3 = 0.79$

http://rcarbonneau.com/ClusterwiseRegressionDatasets.htm
A comparison with the ‘votes’ dataset

The proposed approach
Best partition: 3 groups
- $\theta_{1}^{\text{best}} = 0.18$
- $\theta_{2}^{\text{best}} = 0.49$
- $\theta_{3}^{\text{best}} = 0.79$

Clusterwise linear regression
- a-priori definition of the number of classes
- No exact solution method
- Performance is sensitive to the initial partition and outliers
- Overlapping among groups

C. Davino (University of Naples)  
Handling heterogeneity in QR  
Paris, April 2019  
97 / 110

A comparison with the ‘votes’ dataset

Comparison with alternative methods

Clustering linear regression
- a-priori definition of the number of classes
- No exact solution method
- Performance is sensitive to the initial partition and outliers
- Overlapping among groups

Research questions to be explored
- How to compare results?
- What are other alternative methods?
- ....
- ....

C. Davino (University of Naples)  
Handling heterogeneity in QR  
Paris, April 2019  
98 / 110
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

Concluding remarks: motivation

Motivation (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.

Concluding remarks: motivation

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

Caution
QR offers information on the whole conditional distribution of the response variable, allowing us to discern effects that would otherwise be judged equivalent using only conditional expectation. Nonetheless, the QR ability to statistically detect more effects can not be considered a panacea for investigating relationships between variables: in fact, the improved ability to detect a multitude of effects forces the investigator to clearly articulate what is important to the process being studied and why.
Main references

Main references


