

Conditional Quantile Regression Article Proposal

Marcelo Ruas

May 2, 2017

Overview

- ▶ Let a time series be given by

$$y_t = A(L)y_t + \beta x_t + \varepsilon_t,$$

where the distribution of ε_t is unknown.

- ▶ Quantile regression is one technique available to estimate the distribution of ε_t , by estimating a fine grid of α -quantiles at once.
- ▶ We explore different strategies of estimating the Conditional Quantile Regression focused on approaching the distribution of ε_t (or $y_{t+k|t}$).

Linear Models

Linear Models - Formulation

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (1)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \beta_{\alpha}^T x_t, \quad \forall t \in T, \forall \alpha \in A, \quad (2)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (3)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (4)$$

Linear Models - Resume

- ▶ As there are many explanatory variables for y_t , it is interesting to do a regularization process in order to select only a subset to compose the model.
- ▶ The next slides are going to cover a few different ways to achieve this using Mixed Integer Linear Programming.

Regularization by MILP - Resume

- ▶ MILP models allow only K variables to be used for each α -quantile. This means that only K coefficients $\beta_{p\alpha}$ may have nonzero values, for each α -quantile. It must be guaranteed by constraints on the optimization problem.
- ▶ We present three forms of grouping probabilities while selecting variables

One model for each α -quantile - Formulation

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (5)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (6)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (7)$$

$$-M z_{p\alpha} \leq \beta_{p\alpha} \leq M z_{p\alpha}, \quad \forall \alpha \in A, \forall p \in P, \quad (8)$$

$$\sum_{p=1}^P z_{p\alpha} \leq K, \quad \forall \alpha \in A, \quad (9)$$

$$z_{p\alpha} \in \{0, 1\}, \quad \forall \alpha \in A, \forall p \in P, \quad (10)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (11)$$

Defining groups for α -quantiles - Formulation

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (12)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (13)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (14)$$

$$-Mz_{p\alpha g} \leq \beta_{p\alpha} \leq Mz_{p\alpha g}, \quad \forall \alpha \in A, \forall p \in P, \forall g \in G \quad (15)$$

$$z_{p\alpha g} := 2 - (1 - z_{pg}) - I_{g\alpha} \quad (16)$$

$$\sum_{p=1}^P z_{pg} \leq K, \quad \forall g \in G, \quad (17)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (18)$$

$$\sum_{g \in G} I_{g\alpha} = 1, \quad \forall \alpha \in A, \quad (19)$$

$$I_{g\alpha}, z_{pg} \in \{0, 1\}, \quad \forall p \in P, \quad \forall g \in G, \quad (20)$$

Defining groups by the introduction of switching variable - Formulation

$$\min_{\beta_{0\alpha}, \beta_{\alpha}, z_{p\alpha}, \varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^-} \sum_{\alpha \in A} \sum_{t \in T} \left(\alpha \varepsilon_{t\alpha}^+ + (1 - \alpha) \varepsilon_{t\alpha}^- \right) \quad (21)$$

$$\text{s.t.} \quad \varepsilon_{t\alpha}^+ - \varepsilon_{t\alpha}^- = y_t - \beta_{0\alpha} - \sum_{p=1}^P \beta_{p\alpha} x_{t,p}, \quad \forall t \in T, \forall \alpha \in A, \quad (22)$$

$$\varepsilon_{t\alpha}^+, \varepsilon_{t\alpha}^- \geq 0, \quad \forall t \in T, \forall \alpha \in A, \quad (23)$$

$$-Mz_{p\alpha} \leq \beta_{p\alpha} \leq Mz_{p\alpha}, \quad \forall \alpha \in A, \forall p \in \{1, \dots, P\}, \quad (24)$$

$$\sum_{p=1}^P z_{p\alpha} \leq K, \quad \forall \alpha \in A, \quad (25)$$

$$z_{p\alpha} \in \{0, 1\}, \quad \forall \alpha \in A, \forall p \in \{1, \dots, P\}, \quad (26)$$

$$\beta_{0\alpha} + \beta_{\alpha}^T x_t \leq \beta_{0\alpha'} + \beta_{\alpha'}^T x_t, \quad \forall t \in T, \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha', \quad (27)$$

$$z_{p\alpha} - z_{p\alpha+1} \leq m_{p\alpha}, \quad \forall \alpha \in A', \quad \forall p \in P \quad (28)$$

$$\sum_{\alpha \in A'} r_{\alpha} \leq |G| - 1 \quad (29)$$

$$(30)$$

where $A' = A \setminus \{|A|\}$

Nonparametric model

Nonparametric model - Formulation

$$\min_{q_{\alpha t}, \delta_t^+, \delta_t^-, \xi_t}$$

s.t.

$$\sum_{\alpha \in A} \sum_{t \in T'} \left(\alpha \delta_{t\alpha}^+ + (1 - \alpha) \delta_{t\alpha}^- \right)$$

$$+ \lambda_1 \sum_{t \in T'} \gamma_{t\alpha} + \lambda_2 \sum_{t \in T'} \xi_{t\alpha}$$

$$\delta_t^+ - \delta_{t\alpha}^- = y_t - q_{t\alpha},$$

$$D_{t\alpha}^1 = \frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t},$$

$$D_{t\alpha}^2 = \frac{\left(\frac{q_{\alpha t+1} - q_{\alpha t}}{x_{t+1} - x_t} \right) - \left(\frac{q_{\alpha t} - q_{\alpha t-1}}{x_t - x_{t-1}} \right)}{x_{t+1} - 2x_t + x_{t-1}}.$$

$$\gamma_{t\alpha} \geq D_{t\alpha}^1,$$

$$\gamma_{t\alpha} \geq -D_{t\alpha}^1,$$

$$\xi_{t\alpha} \geq D_{t\alpha}^2,$$

$$\xi_{t\alpha} \geq -D_{t\alpha}^2,$$

$$\delta_{t\alpha}^+, \delta_{t\alpha}^-, \gamma_{t\alpha}, \xi_{t\alpha} \geq 0,$$

$$q_{t\alpha} \leq q_{t\alpha'},$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall \alpha \in A,$$

$$\forall t \in T', \forall (\alpha, \alpha') \in A \times A, \alpha < \alpha' \Rightarrow q_{t\alpha} \leq q_{t\alpha'}$$

Nonparametric vs. Linear Model

- ▶ The nonparametric approach is more flexible to capture heteroscedasticity.

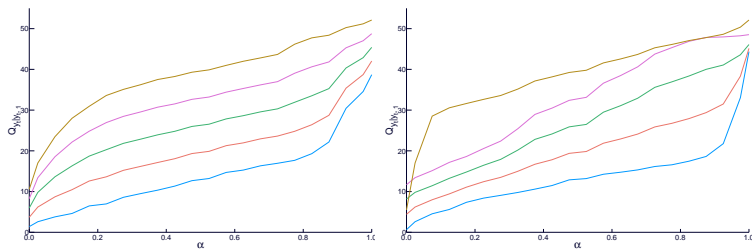


Figure 1: Estimated quantile functions, for different values of y_{t-1} . On the left using a linear model and using a nonparametric approach on the right.

Nonparametric vs. Linear Model

- This flexibility might lead to overfitting, if we don't select a proper penalty, as shown below:

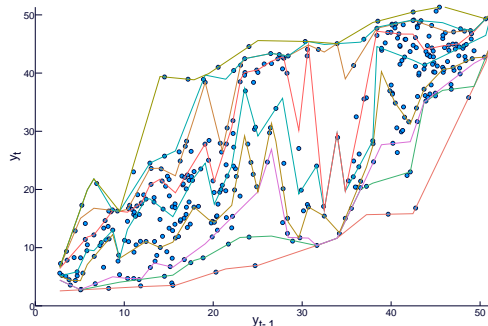


Figure 2: Example of a overfitted quantile function

Goals

Goals

Goals of this work are:

- ▶ Using multiple quantiles to estimate the empirical conditional distribution of variables in a time series
- ▶ Producing a model identification methodology
- ▶ Natural Ressources and Financial applications