Robust LASSO and Efficient Quantile Regression through Regularization of Case-Specific Parameters

Yoonsuh Jung

Department of Statistics
The Ohio State University

Jan 21, 2010

Introduction: Brief Overview

1. Motivation
   - To reduce the effect of outlying or influential cases
2. Method
   - Consider a modeling procedure under the frame of loss function minimization
   - Modify the current modeling procedure by adding case-specific parameters
3. Conjectured results
   - Decrease the potential impacts of outliers
   - Attain robustness and/or efficiency

Introduction: Brief Overview

1. Motivation
   - To reduce the effect of outlying or influential cases
2. Method
   - Consider a modeling procedure under the frame of loss function minimization
   - Modify the current modeling procedure by adding case-specific parameters
3. Conjectured results
   - Decrease the potential impacts of outliers
   - Attain robustness and/or efficiency

Introduction: Brief Overview

1. Motivation
   - To reduce the effect of outlying or influential cases
2. Method
   - Consider a modeling procedure under the frame of loss function minimization
   - Modify the current modeling procedure by adding case-specific parameters
3. Conjectured results
   - Decrease the potential impacts of outliers
   - Attain robustness and/or efficiency

Introduction: Case-specific Parameter and Regularization

1. Case-specific parameter in the linear model

$$y_i = \beta_0 + x_i^\top \beta + \gamma_i + \epsilon_i$$

Or in matrix notation,

$$\begin{aligned} Y &= X\beta + \mathrm{I}\gamma + \epsilon \\ &= (X \quad \mathrm{I}) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \epsilon \end{aligned}$$

2. Dimension: Ill posed problem
   - Regularization method is needed

Introduction: Case-specific Parameter and Regularization

1. Case-specific parameter in the linear model

$$y_i = \beta_0 + x_i^\top \beta + \gamma_i + \epsilon_i$$

Or in matrix notation,

$$\begin{aligned} Y &= X\boldsymbol{\beta} + \mathrm{I}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ &= (X \quad \mathrm{I}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} + \boldsymbol{\epsilon} \end{aligned}$$

2. Dimension: Ill posed problem
   - Regularization method is needed

Introduction: Case-specific Parameter and Regularization

1. Case-specific parameter in the linear model

$$y_i = \beta_0 + x_i^\top \beta + \gamma_i + \epsilon_i$$

Or in matrix notation,

$$\begin{aligned} Y &= X\boldsymbol{\beta} + \mathrm{I}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ &= (X \quad \mathrm{I}) \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix} + \boldsymbol{\epsilon} \end{aligned}$$

2. Dimension: Ill posed problem
   - Regularization method is needed

General Modeling Procedure : Through the objective function minimization

1. Standard regularization of location model

$$L(\beta) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta) + \lambda_\beta J_1(\beta)$$

2. Add case-specific parameters

$$L(\beta, \gamma) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta - \gamma_i) + \lambda_\beta J_1(\beta) + \lambda_\gamma J_2(\gamma)$$

3. With $\hat{\beta}$, find the minimizer, $\hat{\gamma}$, of

$$L(\hat{\beta}, \gamma) = \sum_{i=1}^{n} g(r_i - \gamma_i) + \lambda_\gamma J_2(\gamma), \quad \text{where } r_i = y_i - x_i^\top \hat{\beta}$$

4. With $\hat{\gamma}$, iterate step 2 and 3 until convergence.

General Modeling Procedure : Through the objective function minimization

1. Standard regularization of location model

$$L(\beta) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta) + \lambda_\beta J_1(\beta)$$

2. Add case-specific parameters

$$L(\beta, \gamma) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta - \gamma_i) + \lambda_\beta J_1(\beta) + \lambda_\gamma J_2(\gamma)$$

3. With $\hat{\beta}$, find the minimizer, $\hat{\gamma}$, of

$$L(\hat{\beta}, \gamma) = \sum_{i=1}^{n} g(r_i - \gamma_i) + \lambda_\gamma J_2(\gamma), \quad \text{where } r_i = y_i - x_i^\top \hat{\beta}$$

4. With $\hat{\gamma}$, iterate step 2 and 3 until convergence.

General Modeling Procedure : Through the objective function minimization

1. Standard regularization of location model

$$L(\beta) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta) + \lambda_\beta J_1(\beta)$$

2. Add case-specific parameters

$$L(\beta, \gamma) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta - \gamma_i) + \lambda_\beta J_1(\beta) + \lambda_\gamma J_2(\gamma)$$

3. With $\hat{\beta}$, find the minimizer, $\hat{\gamma}$, of

$$L(\hat{\beta}, \gamma) = \sum_{i=1}^{n} g(r_i - \gamma_i) + \lambda_\gamma J_2(\gamma), \quad \text{where } r_i = y_i - x_i^\top \hat{\beta}$$

4. With $\hat{\gamma}$, iterate step 2 and 3 until convergence.

General Modeling Procedure : Through the objective function minimization

1. Standard regularization of location model

$$L(\beta) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta) + \lambda_\beta J_1(\beta)$$

2. Add case-specific parameters

$$L(\beta, \gamma) = \sum_{i=1}^{n} g(y_i - x_i^\top \beta - \gamma_i) + \lambda_\beta J_1(\beta) + \lambda_\gamma J_2(\gamma)$$

3. With $\hat{\beta}$, find the minimizer, $\hat{\gamma}$, of

$$L(\hat{\beta}, \gamma) = \sum_{i=1}^{n} g(r_i - \gamma_i) + \lambda_\gamma J_2(\gamma), \quad \text{where } r_i = y_i - x_i^\top \hat{\beta}$$

4. With $\hat{\gamma}$, iterate step 2 and 3 until convergence.

LASSO: Modification Procedure I

1. Standard LASSO

$$L(\beta) = \frac{1}{2}(Y - X\beta)^\top (Y - X\beta) + \lambda_\beta \sum_{j=1}^{p} |\beta_j|$$

2. Robust LASSO

$$L(\beta, \gamma) = \frac{1}{2}\{Y - (X\beta + \gamma)\}^\top \{Y - (X\beta + \gamma)\} + \lambda_\beta \sum_{j=1}^{p} |\beta_j| + \lambda_\gamma \sum_{i=1}^{n} |\gamma_i|$$

Case-specific parameters and extra penalty are included

LASSO: Modification Procedure I

1. Standard LASSO

$$L(\beta) = \frac{1}{2}(Y - X\beta)^\top (Y - X\beta) + \lambda_\beta \sum_{j=1}^{p} |\beta_j|$$

2. Robust LASSO

$$\begin{aligned} L(\beta, \gamma) &= \frac{1}{2}\{Y - (X\beta + \gamma)\}^\top \{Y - (X\beta + \gamma)\} \\ &+ \lambda_\beta \sum_{j=1}^{p} |\beta_j| + \lambda_\gamma \sum_{i=1}^{n} |\gamma_i| \end{aligned}$$

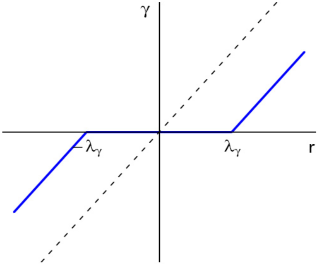
Case-specific parameters and extra penalty are included

LASSO: Modification Procedure II

1. With $\hat{\beta}$,

$$L(\hat{\beta}, \gamma) = \frac{1}{2}(r - \gamma)^\top (r - \gamma) + \lambda_{\hat{\beta}} \sum_{j=1}^{p} |\hat{\beta}_j| + \lambda_\gamma \sum_{i=1}^{n} |\gamma_i|$$

Minimizer is $\hat{\gamma} = sgn(r)(|r| - \lambda_\gamma)_+$



LASSO: Modification Procedure III

1. Now with $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} \frac{1}{2}(y_i - x_i^\top \beta)^2 + \lambda_\beta \sum_{j=1}^p |\beta_j| & \text{for} |y_i - x_i^\top \beta| < \lambda_\gamma \\ \frac{1}{2}\lambda_\gamma^2 + \lambda_\gamma(|y_i - x_i^\top \beta| - \lambda_\gamma) + \lambda_\beta \sum_{j=1}^p |\beta_j| & \text{otherwise} \end{cases}$$

2. Coincide with Huberized LASSO (Rosset & Zhu, 2004)
3. Conjecture : Achieve some Robustness

LASSO: Modification Procedure III

1. Now with $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} \frac{1}{2}(y_i - x_i^\top \beta)^2 + \lambda_\beta \sum_{j=1}^{p} |\beta_j| & \text{for} |y_i - x_i^\top \beta| < \lambda_\gamma \\ \frac{1}{2}\lambda_\gamma^2 + \lambda_\gamma(|y_i - x_i^\top \beta| - \lambda_\gamma) + \lambda_\beta \sum_{j=1}^{p} |\beta_j| & \text{otherwise} \end{cases}$$

2. Coincide with Huberized LASSO (Rosset & Zhu, 2004)
3. Conjecture : Achieve some Robustness

LASSO: Modification Procedure III

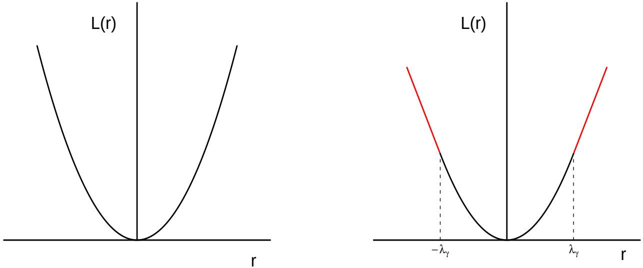
1. Now with $\hat{\gamma}, L(\beta, \hat{\gamma})$ for i^{th} observation becomes

$$\begin{cases} \frac{1}{2}(y_i - x_i^\top \beta)^2 + \lambda_\beta \sum_{j=1}^{p} |\beta_j| & \text{for} |y_i - x_i^\top \beta| < \lambda_\gamma \\ \frac{1}{2}\lambda_\gamma^2 + \lambda_\gamma(|y_i - x_i^\top \beta| - \lambda_\gamma) + \lambda_\beta \sum_{j=1}^{p} |\beta_j| & \text{otherwise} \end{cases}$$

2. Coincide with Huberized LASSO (Rosset & Zhu, 2004)
3. Conjecture : Achieve some Robustness

Graphical Summary

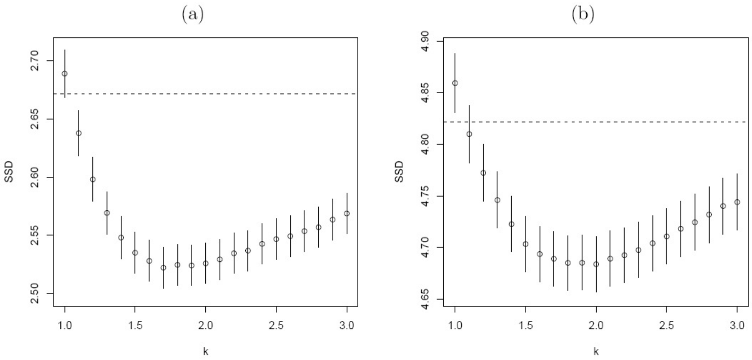
Loss Functions: Standard LASSO and Robust LASSO



Application to Language Data (Baayen, 2007)

1. Total sample size: about 4500 cases
2. Variables: (small) 20 variables (large) 9 additional variables
3. simulated sample size: 400 cases
4. 5000 replicates
5. Calculate sum of squared deviations (SSD) from Baayens fitted values

Application to Language Data (Baayen, 2007)



Sum of squared deviations (SSD) from Baayens fits in the simulation study. The horizontal line is the mean SSD for the LASSO while the points represent the mean of SSDs for the robust LASSO. The vertical lines give approximate 95% confidence intervals for the mean SSDs. Panel (a) presents results for the small set of covariates and panel (b) presents results for the large set of covariates.

More on λ_β and λ_γ

1. λ_β is chosen by Robust Cp (Ronchetti and Staudte, 1994)
2. Set $\lambda_\gamma = k \cdot \sigma$
3. Estimate σ with robust $\hat{\sigma}$ such as MAD
4. (Huber; 1981) suggests $k \in [1, 2]$

More on λ_β and λ_γ

1. λ_β is chosen by Robust Cp (Ronchetti and Staudte, 1994)
2. Set $\lambda_\gamma = k \cdot \sigma$
3. Estimate σ with robust $\hat{\sigma}$ such as MAD
4. (Huber; 1981) suggests $k \in [1, 2]$

More on λ_β and λ_γ

1. λ_β is chosen by Robust Cp (Ronchetti and Staudte, 1994)
2. Set $\lambda_\gamma = k \cdot \sigma$
3. Estimate σ with robust $\hat{\sigma}$ such as MAD
4. (Huber; 1981) suggests $k \in [1, 2]$

More on λ_β and λ_γ

1. λ_β is chosen by Robust Cp (Ronchetti and Staudte, 1994)
2. Set $\lambda_\gamma = k \cdot \sigma$
3. Estimate σ with robust $\hat{\sigma}$ such as MAD
4. (Huber; 1981) suggests $k \in [1, 2]$

Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^\top \beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^\top$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta)$

Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^\top \beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^\top$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta)$

Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^\top \beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^\top$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta)$

Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^\top \beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^\top$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta)$

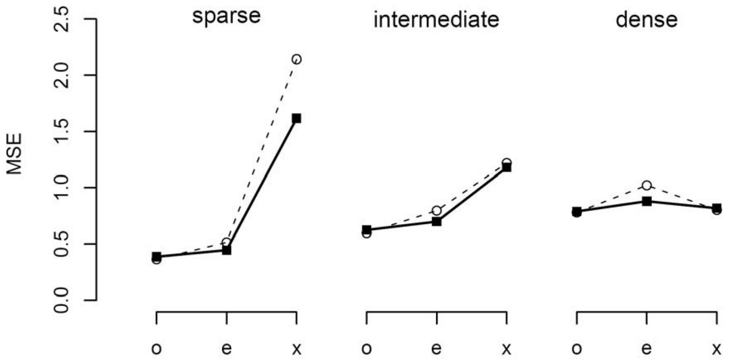
Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^{\top}\beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^{\top}$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^{\top} \Sigma (\hat{\beta}^i - \beta)$

Simulation Setting (Tibshirani, 1996)

1. Standard linear model $y = x^\top \beta + \epsilon$ was assumed
2. Generated $x = (x_1, \ldots, x_8)^\top$ from $N(0, \Sigma)$, where $\rho_{i,j} = (0.5)^{|i-j|}$.
3. Three scenarios were considered
   - a Sparse: $\beta = (5, 0, 0, 0, 0, 0, 0, 0)$
   - b Intermediate: $\beta = (3, 1.5, 0, 0, 2, 0, 0, 0)$
   - c dense: $\beta_j = 0.85$ for all $j = 1, \ldots, 8$
4. $\epsilon_i \sim N(0, 3^2)$
5. Contamination
   - No Contamination
   - first 5% ϵ_i were tripled
   - first 5% of x_1 were tripled
6. $MSE = 100^{-1} \sum_{i=1}^{100} (\hat{\beta}^i - \beta)^\top \Sigma (\hat{\beta}^i - \beta)$

Simulation Results



In each scenario, o, e, and x indicate clean data, data with contaminated measurement errors, and data with mismeasured first covariate. The dotted lines are for LASSO while the solid lines are for Robust LASSO. The points are the average *MSE* from 100 data sets

Median Regression: Modification Procedure I

1. Median Regression

$$L_\lambda(\beta) = \sum_{i=1}^{n} |y_i - x_i^\top \beta|$$

2. Modified Median Regression

$$L(\beta, \gamma) = \sum_{i=1}^{n} |y_i - x_i^\top \beta - \gamma_i| + \frac{\lambda_\gamma}{2} \sum_{i=1}^{n} \gamma_i^2$$

Case-specific parameters and extra penalty are included

Median Regression: Modification Procedure I

1. Median Regression

$$L_\lambda(\beta) = \sum_{i=1}^{n} |y_i - x_i^\top \beta|$$

2. Modified Median Regression

$$L(\beta, \gamma) = \sum_{i=1}^{n} |y_i - x_i^\top \beta - \gamma_i| + \frac{\lambda_\gamma}{2} \sum_{i=1}^{n} \gamma_i^2$$

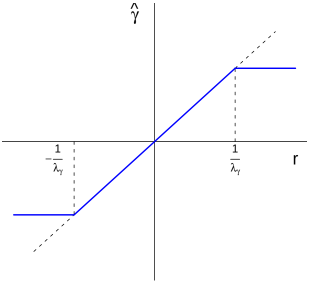
Case-specific parameters and extra penalty are included

Median Regression: Modification Procedure II

1. With $\hat{\beta}$,

   $L(\hat{\beta}, \gamma) = \sum_{i=1}^{n} |r_i - \gamma_i| + \frac{\lambda_\gamma}{2} \sum_{i=1}^{n} \gamma_i^2$

2. Minimizer, $\hat{\gamma} = sgn(r) \frac{1}{\lambda_\gamma} I\left(|r| > \frac{1}{\lambda_\gamma}\right) + rI\left(|r| \le \frac{1}{\lambda_\gamma}\right)$



Median Regression: Modification Procedure III

1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \le \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Median Regression: Modification Procedure III

1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \leq \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Median Regression: Modification Procedure III

1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \leq \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Median Regression: Modification Procedure III

1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \leq \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Median Regression: Modification Procedure III

1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \leq \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Median Regression: Modification Procedure III

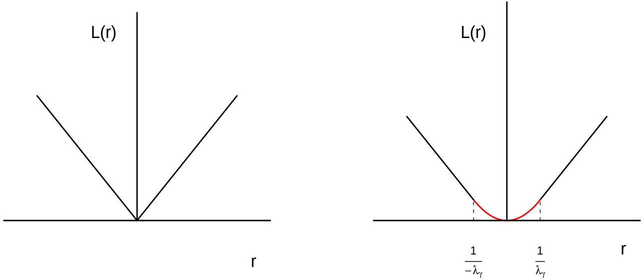
1. With $\hat{\gamma}$, $L(\beta, \hat{\gamma})$ for ith observation becomes

$$\begin{cases} |y_i - x_i^\top \beta| - \frac{1}{2\lambda_\gamma} & \text{for} |y_i - x_i^\top \beta| > \frac{1}{\lambda_\gamma} \\ \frac{\lambda_\gamma}{2}(y_i - x_i^\top \beta)^2 & \text{for} |y_i - x_i^\top \beta| \leq \frac{1}{\lambda_\gamma} \end{cases}$$

2. Again, Huber's loss function
3. Quadratic adjustment of the V shape
4. Conjecture : Achieve some Efficiency
5. Details of bending constant, λ_γ, come later
6. Natural extension to Quantile Regression

Graphical Summary

Loss Function: Median Regression and Efficient Median Regression



Quantile Regression: Introduction

1. Check loss function for estimating q^{th} regression quantile, $0 < q < 1$ (Koenker& Bassett, 1978)

$$\rho(u) = \begin{cases} qu & \text{for } u \geq 0 \\ (q-1)u & \text{for } u < 0. \end{cases}$$

2. Finding the minimizer is equivalent to finding the zero of its derivative,

$$\psi(u) = \begin{cases} q & \text{for } u \geq 0 \\ (q-1) & \text{for } u < 0. \end{cases}$$

Quantile Regression: Introduction

1. Check loss function for estimating q^{th} regression quantile, $0 < q < 1$ (Koenker& Bassett, 1978)

$$\rho(u) = \begin{cases} qu & \text{for } u \geq 0 \\ (q-1)u & \text{for } u < 0. \end{cases}$$

2. Finding the minimizer is equivalent to finding the zero of its derivative,

$$\psi(u) = \begin{cases} q & \text{for } u \geq 0 \\ (q-1) & \text{for } u < 0. \end{cases}$$

Quantile Regression: Modification Procedure I

1. Modified quantile regression

$$L(\beta, \gamma) = \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \gamma_i) + \lambda_\gamma J_2(\gamma)$$

2. Consider asymmetric $J_2(\gamma)$

$$J_2(\gamma) = \frac{q}{1-q}\gamma^2 I(\gamma \geq 0) + \frac{1-q}{q}\gamma^2 I(\gamma < 0)$$

3. With $\hat{\beta}$, the minimizer of $L(\hat{\beta}, \gamma)$ is

$$\hat{\gamma} = -\frac{q}{2\lambda_\gamma} I(r < -\frac{q}{2\lambda_\gamma}) + rI(-\frac{q}{2\lambda_\gamma} \leq r < \frac{1-q}{2\lambda_\gamma}) + \frac{1-q}{2\lambda_\gamma} I(r \geq \frac{1-q}{2\lambda_\gamma})$$

1. Modified quantile regression

$$L(\beta,\gamma) = \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \gamma_i) + \lambda_\gamma J_2(\gamma)$$

2. Consider asymmetric $J_2(\gamma)$

$$J_2(\gamma) = \frac{q}{1-q}\gamma^2 I(\gamma \geq 0) + \frac{1-q}{q}\gamma^2 I(\gamma < 0)$$

3. With $\hat{\beta}$, the minimizer of $L(\hat{\beta},\gamma)$ is

$$\hat{\gamma} = -\frac{q}{2\lambda_\gamma} I(r < -\frac{q}{2\lambda_\gamma}) + rI(-\frac{q}{2\lambda_\gamma} \leq r < \frac{1-q}{2\lambda_\gamma}) + \frac{1-q}{2\lambda_\gamma} I(r \geq \frac{1-q}{2\lambda_\gamma})$$

Quantile Regression: Modification Procedure I

1. Modified quantile regression

$$L(\beta,\gamma) = \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \gamma_i) + \lambda_\gamma J_2(\gamma)$$

2. Consider asymmetric $J_2(\gamma)$

$$J_2(\gamma) = \frac{q}{1-q}\gamma^2 I(\gamma \geq 0) + \frac{1-q}{q}\gamma^2 I(\gamma < 0)$$

3. With $\hat{\beta}$, the minimizer of $L(\hat{\beta},\gamma)$ is

$$\hat{\gamma} = -\frac{q}{2\lambda_\gamma} I(r < -\frac{q}{2\lambda_\gamma}) + rI(-\frac{q}{2\lambda_\gamma} \leq r < \frac{1-q}{2\lambda_\gamma}) + \frac{1-q}{2\lambda_\gamma} I(r \geq \frac{1-q}{2\lambda_\gamma})$$

Quantile Regression: Modification Procedure I

1. Modified quantile regression

$$L(\beta, \gamma) = \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \gamma_i) + \lambda_\gamma J_2(\gamma)$$

2. Consider asymmetric $J_2(\gamma)$

$$J_2(\gamma) = \frac{q}{1-q}\gamma^2 I(\gamma \geq 0) + \frac{1-q}{q}\gamma^2 I(\gamma < 0)$$

3. With $\hat{\beta}$, the minimizer of $L(\hat{\beta}, \gamma)$ is

$$\hat{\gamma} = -\frac{q}{2\lambda_\gamma} I(r < -\frac{q}{2\lambda_\gamma}) + rI(-\frac{q}{2\lambda_\gamma} \leq r < \frac{1-q}{2\lambda_\gamma}) + \frac{1-q}{2\lambda_\gamma} I(r \geq \frac{1-q}{2\lambda_\gamma})$$

Quantile Regression: Modification Procedure II

1. With $\hat{\gamma}$

$$L(\beta, \hat{\gamma}) = \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \hat{\gamma}_i) + \lambda_\gamma J_2(\hat{\gamma})$$
$$= \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)$$

2. $\rho^M(u)$ is given by

$$\rho_\gamma^M(u) = \begin{cases} qu - \frac{q(1-q)}{4\lambda_\gamma} & \text{if } \frac{1-q}{2\lambda_\gamma} \le u \\ \lambda_\gamma \frac{q}{1-q} u^2 & \text{if } 0 \le u < \frac{1-q}{2\lambda_\gamma} \\ \lambda_\gamma \frac{1-q}{q} u^2 & \text{if } -\frac{q}{2\lambda_\gamma} \le u < 0 \\ (q-1)u - \frac{q(1-q)}{4\lambda_\gamma} & \text{if } u < -\frac{q}{2\lambda_\gamma} \end{cases}$$

Quantile Regression: Modification Procedure II

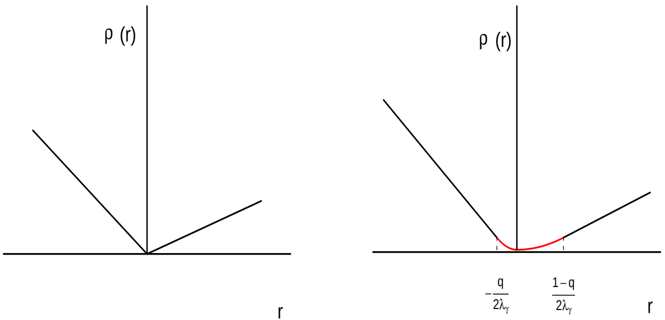
1. With $\hat{\gamma}$

$$
\begin{aligned}
L(\beta, \hat{\gamma}) &= \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta - \hat{\gamma}_i) + \lambda_\gamma J_2(\hat{\gamma}) \\
&= \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)
\end{aligned}
$$

2. $\rho^M(u)$ is given by

$$
\rho_\gamma^M(u) = \begin{cases}
qu - \frac{q(1-q)}{4\lambda_\gamma} & \text{if } \frac{1-q}{2\lambda_\gamma} \le u \\
\lambda_\gamma \frac{q}{1-q} u^2 & \text{if } 0 \le u < \frac{1-q}{2\lambda_\gamma} \\
\lambda_\gamma \frac{1-q}{q} u^2 & \text{if } -\frac{q}{2\lambda_\gamma} \le u < 0 \\
(q-1)u - \frac{q(1-q)}{4\lambda_\gamma} & \text{if } u < -\frac{q}{2\lambda_\gamma}
\end{cases}
$$

Loss Function: Standard QR and Efficient QR



Quantile Regression: Modification Procedure III

1. $\psi^M(u)$ is employed for computing purpose. *rlm*(*MASS*) procedure is used in *R*.

$$\psi^M(u) = \begin{cases} q & \text{if } \frac{1-q}{2\lambda_\gamma} \le u \\ 2\lambda_\gamma \frac{q}{1-q} u & \text{if } 0 \le u < \frac{1-q}{2\lambda_\gamma} \\ 2\lambda_\gamma \frac{1-q}{q} u & \text{if } -\frac{q}{2\lambda_\gamma} \le u < 0 \\ (q-1) & \text{if } u < -\frac{q}{2\lambda_\gamma} \end{cases}$$

2. Recall the $\psi(u)$ for standard QR

$$\psi(u) = \begin{cases} q & \text{for } u \ge 0 \\ (q-1) & \text{for } u < 0. \end{cases}$$

Quantile Regression: Modification Procedure III

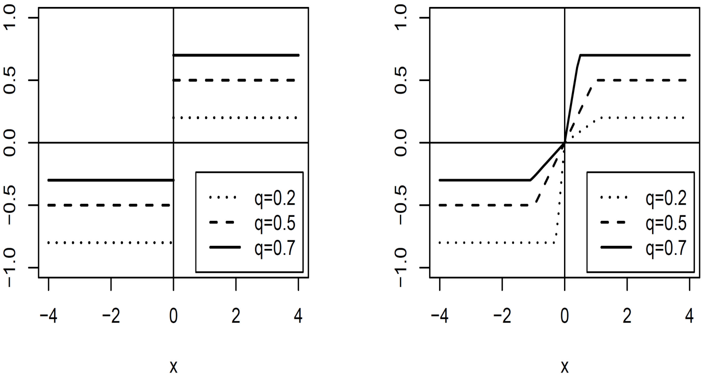
1. $\psi^M(u)$ is employed for computing purpose. $rlm(MASS)$ procedure is used in R.

$$\psi^M(u) = \begin{cases} q & \text{if } \frac{1-q}{2\lambda_\gamma} \le u \\ 2\lambda_\gamma \frac{q}{1-q} u & \text{if } 0 \le u < \frac{1-q}{2\lambda_\gamma} \\ 2\lambda_\gamma \frac{1-q}{q} u & \text{if } -\frac{q}{2\lambda_\gamma} \le u < 0 \\ (q-1) & \text{if } u < -\frac{q}{2\lambda_\gamma} \end{cases}$$

2. Recall the $\psi(u)$ for standard QR

$$\psi(u) = \begin{cases} q & \text{for } u \ge 0 \\ (q-1) & \text{for } u < 0. \end{cases}$$

ψ Function: Standard QR and Modified QR



Quantile Regression: A Rule for λ_γ I

1. Through simulation search for c which provides 'minimum' *MSE*.
2. Error distributions investigated are
   - Standard Normal
   - T distributions
   - Gamma distributions
   - Log-normal distribution
   - Exponential distribution
3. Sample sizes: $10^2, 10^{2.5}, 10^3, 10^{3.5}, 10^4$
4. quantiles: 0.1, 0.2, . . . ,0.9
5. The length of interval (of adjustment) is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$

Quantile Regression: A Rule for λ_γ I

1. Through simulation search for c which provides 'minimum' *MSE*.
2. Error distributions investigated are
   - Standard Normal
   - T distributions
   - Gamma distributions
   - Log-normal distribution
   - Exponential distribution
3. Sample sizes: $10^2, 10^{2.5}, 10^3, 10^{3.5}, 10^4$
4. quantiles: 0.1, 0.2, ... ,0.9
5. The length of interval (of adjustment) is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$

Quantile Regression: A Rule for λ_γ I

1. Through simulation search for c which provides 'minimum' MSE.
2. Error distributions investigated are
   - Standard Normal
   - T distributions
   - Gamma distributions
   - Log-normal distribution
   - Exponential distribution
3. Sample sizes: $10^2, 10^{2.5}, 10^3, 10^{3.5}, 10^4$
4. quantiles: 0.1, 0.2, . . . ,0.9
5. The length of interval (of adjustment) is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$

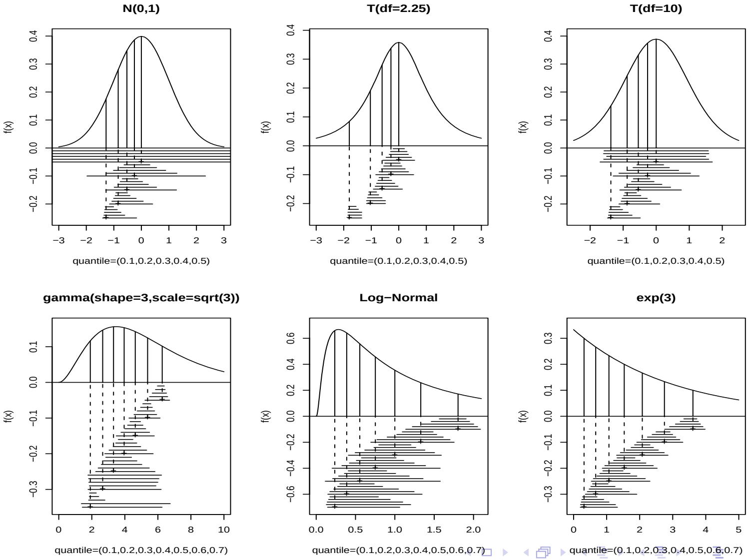
Quantile Regression: A Rule for λ_γ I

1. Through simulation search for c which provides 'minimum' *MSE*.
2. Error distributions investigated are
   - Standard Normal
   - T distributions
   - Gamma distributions
   - Log-normal distribution
   - Exponential distribution
3. Sample sizes: $10^2, 10^{2.5}, 10^3, 10^{3.5}, 10^4$
4. quantiles: 0.1, 0.2, . . . ,0.9
5. The length of interval (of adjustment) is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$

Quantile Regression: A Rule for λ_γ I

1. Through simulation search for c which provides 'minimum' *MSE*.
2. Error distributions investigated are
   - Standard Normal
   - T distributions
   - Gamma distributions
   - Log-normal distribution
   - Exponential distribution
3. Sample sizes: $10^2, 10^{2.5}, 10^3, 10^{3.5}, 10^4$
4. quantiles: 0.1, 0.2, . . . ,0.9
5. The length of interval (of adjustment) is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$

Quantile Regression: A Rule for λ_γ II



Quantile Regression: A Rule for λ_γ III

1. Again, the length of interval adjusted is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$
2. Given n, α, rule for c is,

$$\hat{c} = \begin{cases} e^{-2.118-1.097q} & \text{for } q < 0.5 \\ e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5 \end{cases}$$

   which is developed from exponential error distribution.
3. For computation : Embed the rule in the $\psi^M(u)$ then use $rlm(MASS)$ in R.
4. Prediction ability?

Quantile Regression: A Rule for λ_γ III

1. Again, the length of interval adjusted is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$
2. Given n, α, rule for c is,

$$\hat{c} = \begin{cases} e^{-2.118-1.097q} & \text{for } q < 0.5 \\ e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5 \end{cases}$$

which is developed from exponential error distribution.

3. For computation : Embed the rule in the $\psi^M(u)$ then use *rlm*(*MASS*) in *R*.
4. Prediction ability?

Quantile Regression: A Rule for λ_γ III

1. Again, the length of interval adjusted is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$
2. Given n, α, rule for c is,

$$\hat{c} = \begin{cases} e^{-2.118-1.097q} & \text{for } q < 0.5 \\ e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5 \end{cases}$$

   which is developed from exponential error distribution.

3. For computation : Embed the rule in the $\psi^M(u)$ then use *rlm*(*MASS*) in *R*.
4. Prediction ability?

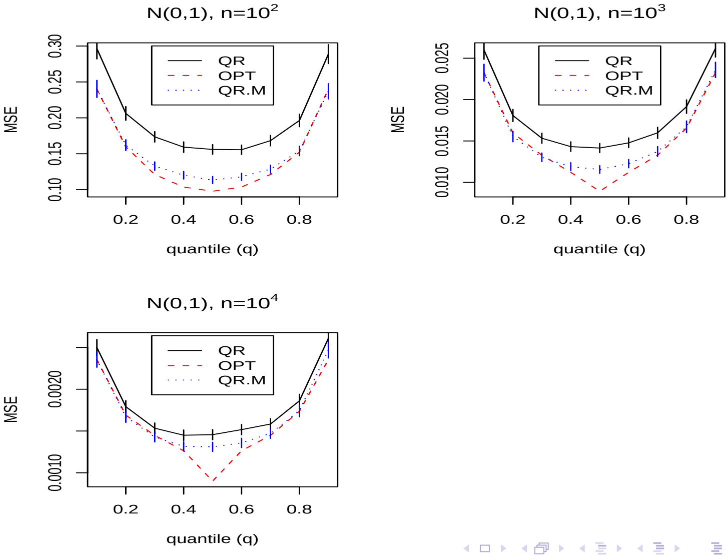
Quantile Regression: A Rule for λ_γ III

1. Again, the length of interval adjusted is $\frac{1}{2\lambda_\gamma} = \frac{\hat{\sigma}}{2c \cdot n^\alpha}$
2. Given n, α, rule for c is,

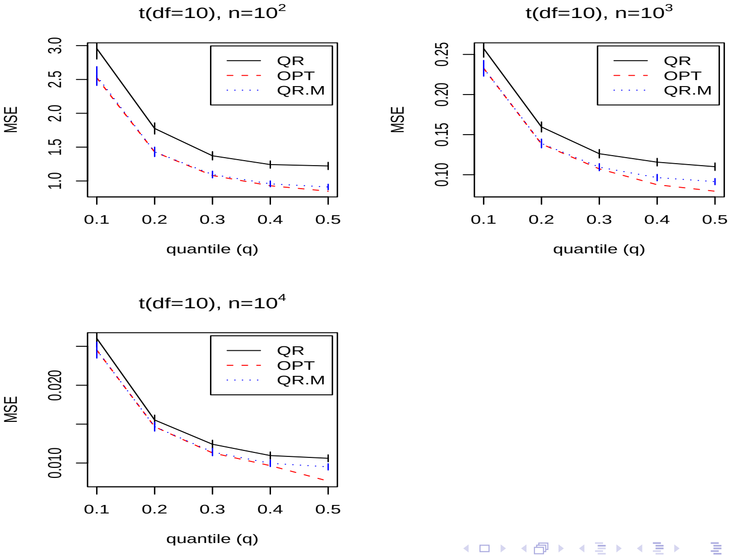
$$\hat{c} = \begin{cases} e^{-2.118-1.097q} & \text{for } q < 0.5 \\ e^{-2.118-1.097(1-q)} & \text{for } q \geq 0.5 \end{cases}$$

   which is developed from exponential error distribution.
3. For computation : Embed the rule in the $\psi^M(u)$ then use $rlm(MASS)$ in R.
4. Prediction ability?

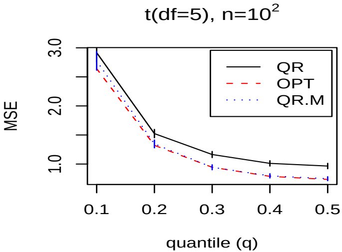
Quantile Regression: Prediction with the Rule on N(0,1) error distribution

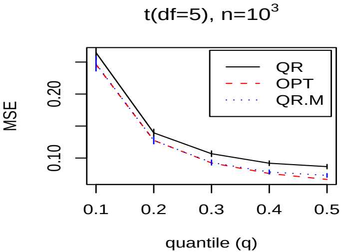


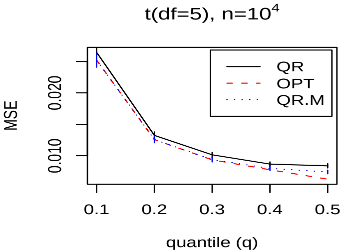
Quantile Regression: Prediction with the Rule on t(df=10) error distribution



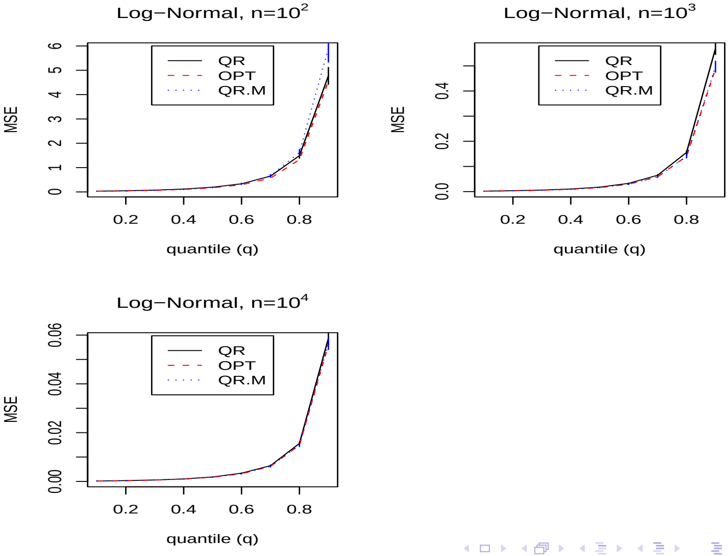
Quantile Regression: Prediction with the Rule on t(df=5) error distribution



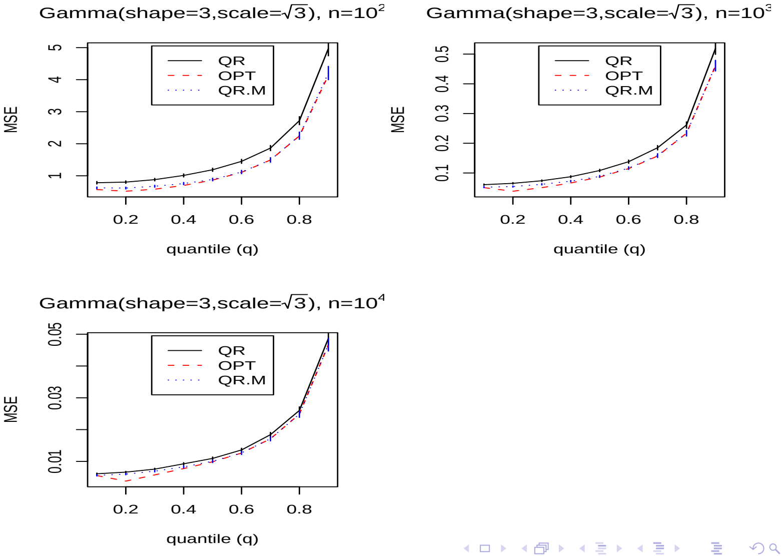




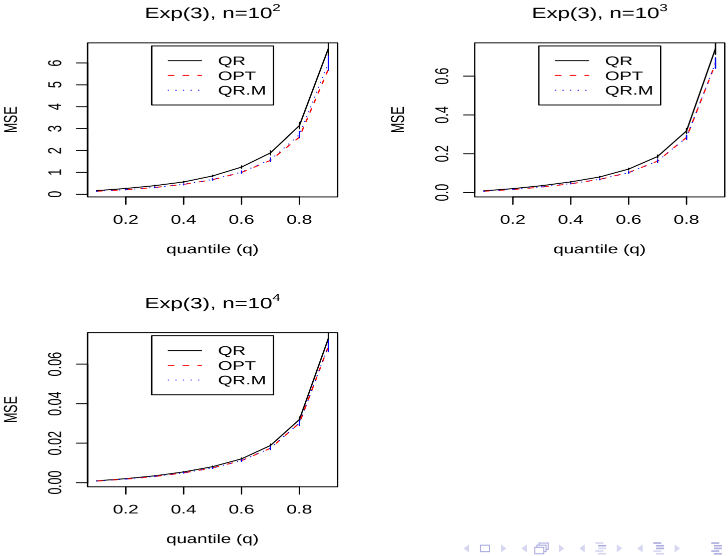
Quantile Regression: Prediction with the Rule on log-normal error distribution



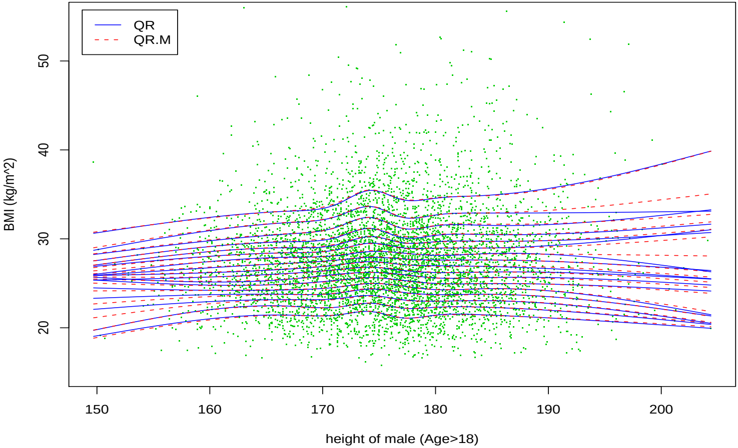
Quantile Regression: Prediction with the Rule on gamma error distribution



Quantile Regression: Prediction with the Rule on exponential error distribution

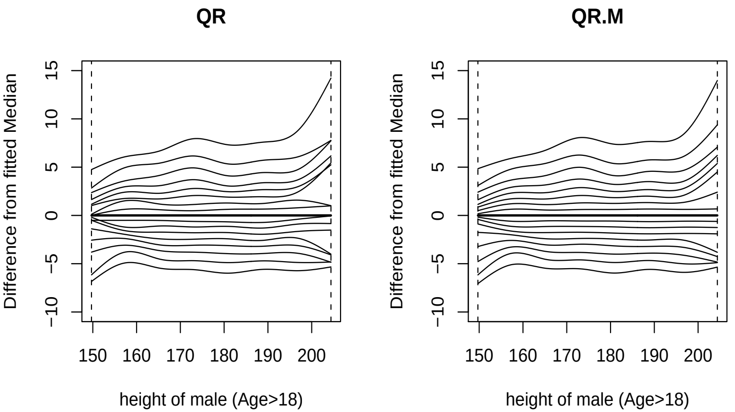


Application to NHANES Data I



Body Mass Index vs height from 5938 U.S. male (age>18)

Application to NHANES Data II



Difference between fitted median line and the other fitted quantiles.

Quantile Regression: Asymptotic Properties I

Assumptions

1. There exist continuous and differentiable densities, $f_i(\xi)$ uniformly bounded away from 0 and ∞ at ξ_i, i=1,2,....
2. $\lim_{n\to\infty} n^{-1} \sum x_i x_i^\top = D_0$, where D_0 is positive definite
3. $\lim_{n\to\infty} n^{-1} \sum f_i(\xi_i) x_i x_i^\top = D_1$, where D_1 is positive definite
4. $\max_{i=1,.,n} \parallel x_i \parallel / \sqrt{n} \to 0$

Definition: Standard QR Estimator

$$\hat{\beta}_q = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_q - \beta) \to N(0, q(1-q) D_1^{-1} D_0 D_1^{-1})$$

Quantile Regression: Asymptotic Properties I

Assumptions

1. There exist continuous and differentiable densities, $f_i(\xi)$ uniformly bounded away from 0 and ∞ at ξ_i, i=1,2,....
2. $\lim_{n\to\infty} n^{-1} \sum x_i x_i^\top = D_0$, where D_0 is positive definite
3. $\lim_{n\to\infty} n^{-1} \sum f_i(\xi_i) x_i x_i^\top = D_1$, where D_1 is positive definite
4. $\max_{i=1,.,n} \| x_i \| / \sqrt{n} \to 0$

Definition: Standard QR Estimator

$$\hat{\beta}_q = \arg\min_{\beta\in\mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_q - \beta) \to N(0, q(1-q) D_1^{-1} D_0 D_1^{-1})$$

Quantile Regression: Asymptotic Properties I

Assumptions

1. There exist continuous and differentiable densities, $f_i(\xi)$ uniformly bounded away from 0 and ∞ at ξ_i, i=1,2,....
2. $\lim_{n\to\infty} n^{-1} \sum x_i x_i^\top = D_0$, where D_0 is positive definite
3. $\lim_{n\to\infty} n^{-1} \sum f_i(\xi_i) x_i x_i^\top = D_1$, where D_1 is positive definite
4. $\max_{i=1,.,n} \parallel x_i \parallel / \sqrt{n} \to 0$

Definition: Standard QR Estimator

$$\hat{\beta}_q = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_q - \beta) \to N(0, q(1-q) D_1^{-1} D_0 D_1^{-1})$$

Quantile Regression: Asymptotic Properties II

1. Modified QR Estimator

$$\hat{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_1^{-1} D_0 D_1^{-1}), \text{ if } \alpha > 1/3$$

2. Modified QR Estimator under Location-Scale Family Model : $y_i = x_i^\top \beta + (x_i^\top \tau)\epsilon_i$, where ϵ_i are iid

$$\check{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\gamma^M((y_i - x_i^\top \beta)/x_i^\top \tau)$$

With $\sqrt{n}$-consistent estimator of τ (Koenker,Zhao: 1994)

$$\sqrt{n}(\check{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{ if } \alpha > 1/3$$

Quantile Regression: Asymptotic Properties II

1. Modified QR Estimator

$$\hat{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_{q,M} - \beta) \to N(0, q(1-q) D_1^{-1} D_0 D_1^{-1}), \text{if } \alpha > 1/3$$

2. Modified QR Estimator under Location-Scale Family Model : $y_i = x_i^\top \beta + (x_i^\top \tau)\epsilon_i$, where ϵ_i are iid

$$\check{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\gamma^M((y_i - x_i^\top \beta)/x_i^\top \tau)$$

With $\sqrt{n}$-consistent estimator of τ (Koenker,Zhao: 1994)

$$\sqrt{n}(\check{\beta}_{q,M} - \beta) \to N(0, q(1-q) D_2^{-1}), \text{if } \alpha > 1/3$$

Quantile Regression: Asymptotic Properties II

1. Modified QR Estimator

$$\hat{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_1^{-1} D_0 D_1^{-1}), \text{if } \alpha > 1/3$$

2. Modified QR Estimator under Location-Scale Family

Model : $y_i = x_i^\top \beta + (x_i^\top \tau)\epsilon_i$, where ϵ_i are iid

$$\check{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\gamma^M((y_i - x_i^\top \beta)/x_i^\top \tau)$$

With $\sqrt{n}$-consistent estimator of τ (Koenker,Zhao: 1994)

$$\sqrt{n}(\check{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{if } \alpha > 1/3$$

Quantile Regression: Asymptotic Properties II

1. Modified QR Estimator

$$\hat{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\hat{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_1^{-1}D_0D_1^{-1}), \text{ if } \alpha > 1/3$$

2. Modified QR Estimator under Location-Scale Family

Model : $y_i = x_i^\top \beta + (x_i^\top \tau)\epsilon_i$, where ϵ_i are iid

$$\check{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\gamma^M((y_i - x_i^\top \beta)/x_i^\top \tau)$$

With $\sqrt{n}$-consistent estimator of τ (Koenker,Zhao: 1994)

$$\sqrt{n}(\check{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{ if } \alpha > 1/3$$

Quantile Regression: Asymptotic Properties III

1. Modified QR Estimator under Heterogeneous data

$$\tilde{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} f_i(\xi_i)\rho_\gamma^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\tilde{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{if } \alpha > 1/3$$

2. Remark: $D_1^{-1} D_0 D_1^{-1} > D_2^{-1}$
3. Performance of $\tilde{\beta}_{q,M}$ in finite sample size?

Quantile Regression: Asymptotic Properties III

1. Modified QR Estimator under Heterogeneous data

$$\tilde{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} f_i(\xi_i)\rho_\gamma^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\tilde{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{if } \alpha > 1/3$$

2. Remark: $D_1^{-1}D_0D_1^{-1} > D_2^{-1}$
3. Performance of $\tilde{\beta}_{q,M}$ in finite sample size?

Quantile Regression: Asymptotic Properties III

1. Modified QR Estimator under Heterogeneous data

$$\tilde{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} f_i(\xi_i)\rho_\gamma^M(y_i - x_i^\top \beta)$$

$$\sqrt{n}(\tilde{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{if } \alpha > 1/3$$

2. Remark: $D_1^{-1}D_0D_1^{-1} > D_2^{-1}$
3. Performance of $\tilde{\beta}_{q,M}$ in finite sample size?

Quantile Regression: Asymptotic Properties III

1. Modified QR Estimator under Heterogeneous data

$$\tilde{\beta}_{q,M} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} f_i(\xi_i) \rho_\gamma^M (y_i - x_i^\top \beta)$$

$$\sqrt{n}(\tilde{\beta}_{q,M} - \beta) \to N(0, q(1-q)D_2^{-1}), \text{if } \alpha > 1/3$$

2. Remark: $D_1^{-1} D_0 D_1^{-1} > D_2^{-1}$
3. Performance of $\tilde{\beta}_{q,M}$ in finite sample size?

Quantile Regression: Simulation with Heterogeneous Data

1. Consider linear model $y_i = \beta_0 + \beta_1 x_i + x_i \epsilon_i$, ϵ_i's are *iid* standard normal
2. $(\beta_0, \beta_1)^\top = (1, 2)^\top$
3. x is consist of three points; x$\in \{1, 2, 3\}$
4. sample size n=300 and n=900 were made and evenly distributed at each design point (200 replicates)
5. Standard QR, Modified QR(QR.M), Weighted QR (WQR), and Weighted QR.M (WQR.M) are compared

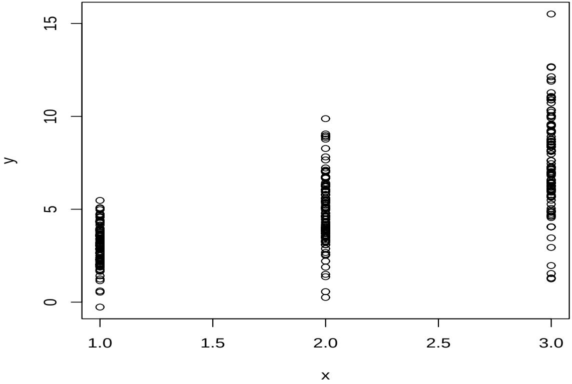

15
10
5
0
y
1.0
1.5
2.0
2.5
3.0
x

Table: mean of *MSE* and its standard error in $(\cdot)$ from 200 replications with n=300 and n=900 at q^{th} quantile, multiplied by 1000.

	q=.1	q=.2	q=.3	q=.4	q=.5
			n=300		
QR	92.31(8.85)	66.68(5.47)	56.40(4.80)	44.77(3.57)	42.53(3.33)
QR.M	92.91(8.41)	62.37(5.19)	48.38(4.06)	37.48(2.94)	33.47(2.64)
WQR	85.49(8.32)	62.10(5.53)	54.93(4.97)	42.82(3.46)	42.32(3.48)
WQR.M	84.44(8.03)	57.93(5.19)	44.84(3.92)	35.55(2.89)	31.21(2.48)
			n=900		
RQ	28.93(2.44)	22.10(2.04)	18.06(1.56)	15.50(1.43)	14.41(1.25)
RQ.M	28.89(2.40)	21.09(1.95)	15.90(1.35)	13.13(1.14)	12.22(1.07)
WQR	27.77(2.31)	21.44(1.91)	17.73(1.55)	15.20(1.44)	14.42(1.21)
WRQ.M	28.28(2.59)	20.09(1.82)	15.30(1.28)	12.58(1.10)	11.54(1.00)

Concluding Remarks

1. A new approach to treat cases
2. Regularization of case-specific parameter increases
   1. Robustness in LASSO
   2. Efficiency and Robustness in Quantile Regression
3. Broadly applicable

Future Research

1. Classification
    - Logistic Regression
    - Support Vector Machine
2. Cross Validation
    - Find more accurate loss function

Thank You!