49TH SCIENTIFIC MEETING OF THE ITALIAN STATISTICAL SOCIETY

Quantile Regression Coefficients Modeling: a Penalized Approach

Gianluca SOTTILE, Paolo FRUMENTO and Matteo BOTTAI

Contents

Contents

Background

Let be y a response variable of length n, and $\boldsymbol{x}$ a $n \times q$ matrix of covariates. The conditional quantile function could be written as:

Background

Let be y a response variable of length n, and $\boldsymbol{x}$ a $n \times q$ matrix of covariates. The conditional quantile function could be written as:

- $Q(p \mid \boldsymbol{x}) = \boldsymbol{x}^T \beta(p)$ in **quantile regression** (QR, Koenker and Bassett Jr, (1978)), where quantiles are estimated one at the time and the estimated coefficients are generally unsmooth functions of p;

Background

Let be y a response variable of length n, and $\boldsymbol{x}$ a $n \times q$ matrix of covariates. The conditional quantile function could be written as:

- $Q(p \mid \boldsymbol{x}) = \boldsymbol{x}^T\beta(p)$ in **quantile regression** (QR, Koenker and Bassett Jr, (1978)), where quantiles are estimated one at the time and the estimated coefficients are generally unsmooth functions of p;
- $Q(p \mid \boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{x}^T\beta(p \mid \boldsymbol{\theta}) = \boldsymbol{x}^T\boldsymbol{\theta}\boldsymbol{b}(p)$ in **quantile regression coefficients modeling** (QRCM, Frumento and Bottai, (2016)), where the estimated coefficients are functions of the order of the quantiles.

Contents

QRCM

A parametric approach that permits modeling the entire quantile function. Consider, for example, describing $\beta(p \mid \boldsymbol{\theta})$ by k-th degree polynomial functions:

$$\beta_j(p \mid \boldsymbol{\theta}) = \theta_{j0} + \theta_{j1} p + \ldots + \theta_{jk} p^k, \; j = 1, \ldots, q.$$

Each covariate has $(k+1)$ associated parameters, for a total of $q \times (k+1)$ model coefficients.

QRCMPEN

Issue

When q is large, estimation may become difficult and the model may be poorly identified, causing the variability to grow out of control.

QRCMPEN

Issue

When q is large, estimation may become difficult and the model may be poorly identified, causing the variability to grow out of control.

Solution

an L_1-penalty term in the QRCM framework ($\lambda\|\boldsymbol{\theta}\|_1$)
(a new LASSO-type model)

QRCMPEN

Issue

When q is large, estimation may become difficult and the model may be poorly identified, causing the variability to grow out of control.

Solution

an L_1-penalty term in the QRCM framework ($\lambda\|\boldsymbol{\theta}\|_1$)
(a new LASSO-type model)

Standard L_1-QR

It focus on model selection when estimating one quantile at a time. This is inefficient and makes it difficult to interpret the results, because some coefficients could be only significant at some quantiles.

Penalized integrated loss function (PILM)

We propose minimizing

$$\overline{L}^{(\lambda)}_{\text{PEN}}(\boldsymbol{\theta}) = \int_0^1 L(\beta(p \mid \boldsymbol{\theta})) + \lambda \sum_{j=1}^{q} \sum_{h=1}^{k} \mid \theta_{jh} \mid \mathrm{d}p,$$

where $L(\beta(p))$ is the loss function of standard quantile regression given by $L = \sum_{i=1}^{n} (p - I(y_i \leq \boldsymbol{x}_i^T \beta(p)))(y_i - \boldsymbol{x}_i^T \beta(p))$, and $\lambda \geq 0$ is the tuning parameter.

Penalized integrated loss function (PILM)

We propose minimizing

$$\overline{L}_{\text{PEN}}^{(\lambda)}(\boldsymbol{\theta}) = \int_0^1 L(\beta(p \mid \boldsymbol{\theta})) + \lambda \sum_{j=1}^{q} \sum_{h=1}^{k} \mid \theta_{jh} \mid \mathrm{d}p,$$

where $L(\beta(p))$ is the loss function of standard quantile regression given by $L = \sum_{i=1}^{n} (p - I(y_i \leq \boldsymbol{x}_i^T \beta(p)))(y_i - \boldsymbol{x}_i^T \beta(p))$, and $\lambda \geq 0$ is the tuning parameter.

Optimization and Implementation

`qrcm` R package + coordinate descent algo $\Rightarrow$ `qrcmNP` R package

Contents

AIC and BIC criteria

With a given set of real data, the true model is not known. In penalized regression, the tuning parameter λ balances the trade-off between goodness of fit and efficiency.

AIC and BIC criteria

With a given set of real data, the true model is not known. In penalized regression, the tuning parameter λ balances the trade-off between goodness of fit and efficiency.

Following the definitions proposed by Schwarz, 1978; Lee et al., 2014; Zheng and Peng, 2017, we propose to use

$$\mathrm{AIC}^{(\lambda)} = \log \overline{L}_{\mathrm{PEN}}^{(\lambda)}(\widehat{\boldsymbol{\theta}}) + 2\mathrm{df}^{(\lambda)} n^{-1}, \quad (1)$$

$$\mathrm{BIC}^{(\lambda)} = \log \overline{L}_{\mathrm{PEN}}^{(\lambda)}(\widehat{\boldsymbol{\theta}}) + \log(n)\mathrm{df}^{(\lambda)} n^{-1}. \quad (2)$$

where $\widehat{\boldsymbol{\theta}}$ is the estimator of $\boldsymbol{\theta}$ obtained by minimizing the PILM at a given value of λ, and $\mathrm{df}^{(\lambda)}$ reflects the number of nonzero coefficients.

Contents

Model selection in inspiratory capacity

Inspiratory Capacity (IC) data

A study carried out in 1988-1991 in Northern Italy

- $n = 2,201$ subjects (49% Male and 51% Female)
- $q = 9$ (age, height, body mass index (BMI), sex, current smoking status, occupational exposure, cough, wheezing and asthma)

Model selection in inspiratory capacity

Inspiratory Capacity (IC) data

A study carried out in 1988-1991 in Northern Italy

- $n = 2,201$ subjects (49% Male and 51% Female)
- $q = 9$ (age, height, body mass index (BMI), sex, current smoking status, occupational exposure, cough, wheezing and asthma)

The model basis

- Intercept: $\boldsymbol{b}(p) = [1, \log(p), \log(1-p)]^T$
- Covariates: a shifted Legendre polynomial (SLP) up to a 5^{th} degree (Abramowitz and Stegun, 1964)

Mod

Inspi

A stu

ig

The

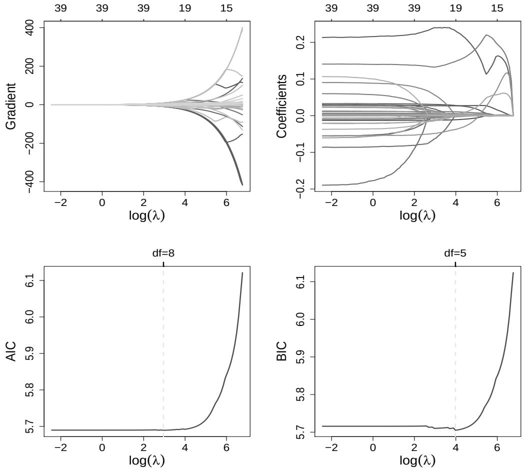

Figure 1: Gradient plot and coefficient profile plot versus $\log(\lambda)$ (top panels); AIC and BIC curves versus $\log(\lambda)$ (bottom panels), for the inspiratory capacity data.

Model selection in inspiratory capacity

Table 1: Model selection based on different criteria. We report the number of parameters, the number of selected covariates, the optimal λ value, the value of the minimized loss function, and the p-value of a Kolmogorov-Smirnov goodness-of-fit test.

Criterion	n. of parameters	n. of covariates	λ	Loss	P-value KS
AIC	31/39	10/10	20.79	293.31	.77
BIC	19/39	5/10	60.47	294.01	.53

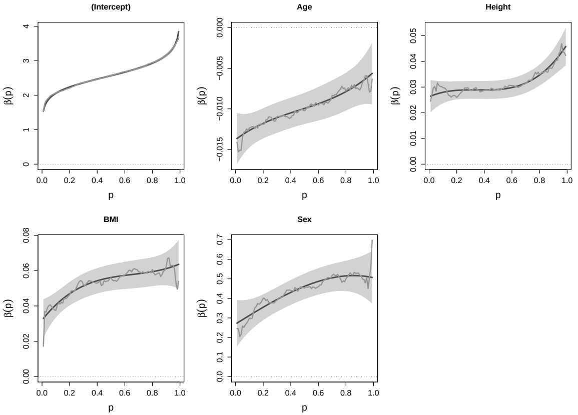

Figure 2: Unpenalized QRCM estimates of $\beta(p)$ under the model selected by BIC (see Table 1). Confidence bands are displayed as shaded areas. The broken lines connect the coefficients of ordinary quantile regression estimated at a grid of quantiles. The dashed line indicates the zero.

Contents

To sum-up

- We proposed a new LASSO-type model in the QRCM framework;

To sum-up

- We proposed a new LASSO-type model in the QRCM framework;
- We proposed two different criteria to select the optimal tuning parameter;

To sum-up

- We proposed a new LASSO-type model in the QRCM framework;
- We proposed two different criteria to select the optimal tuning parameter;
- Results on the Inspiratory Capacity data showed that our proposal is an efficient tool to recover the most informative covariates with a high probability;

To sum-up

- We proposed a new LASSO-type model in the QRCM framework;
- We proposed two different criteria to select the optimal tuning parameter;
- Results on the Inspiratory Capacity data showed that our proposal is an efficient tool to recover the most informative covariates with a high probability;
- A computationally efficient algorithm has been implemented in the `qrcmNP` package in R.

Thanks for the attention!!!

Abramowitz, M. and I.A. Stegun (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Vol. 55. Courier Corporation.

Frumento, P. and M. Bottai (2016). "Parametric modeling of quantile regression coefficient functions". In: *Biometrics* 72.1, pp. 74–84.

Koenker, R. and G. Bassett Jr (1978). "Regression quantiles". In: *Econometrica: journal of the Econometric Society*, pp. 33–50.

Lee, ER, H Noh, and BU Park (2014). "Model selection via bayesian information criterion for quantile regression models." In: *J Am Stat Assoc* 109, pp. 216–229.

Schwarz, G (1978). "Estimating the dimension of a model". In: *Ann Stat* 6, pp. 461–464.

Zheng, Q and L Peng (2017). "Consistent model identification of varying coefficient quantile regression with BIC tuning parameter selection". In: *Commun Stat-Theor M* 46.3, pp. 1031–1049.