Stochastic Search Variable Selection in Quantile Regression Based on Empirical Likelihood

Ruibin Xi
Peking University
第10届全国概率统计会议
Joint work with Yunxiao Li and Yiming Hu

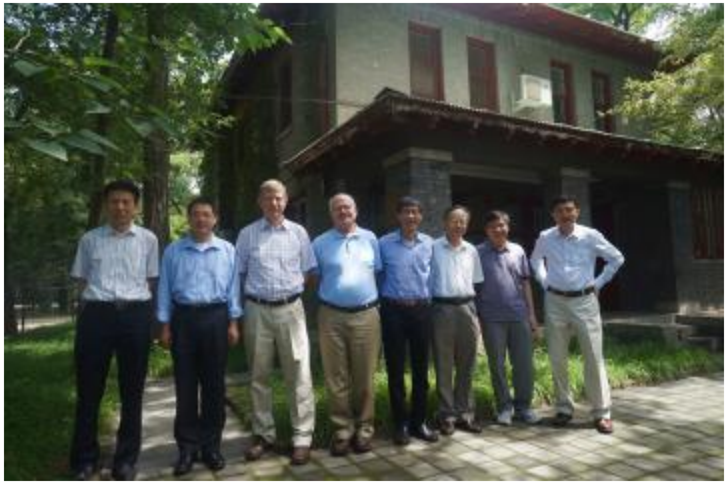
PEKING UNIVERSITY

北京大学数学科学学院

School of Mathematical Sciences, Peking University

北京大学统计科学中心

Center for Statistical Science, Peking University



Outline

- Introduction
  - Quantile regression
  - Empirical likelihood
- Bayesian model selection in quantile regression based on empirical likelihood
  - Asymptotic property
  - Gibbs sampler
  - Simulation
  - Real Data analysis
- Discussion

Quantile regression

- In linear model setup

  response = signal + i.i.d. error

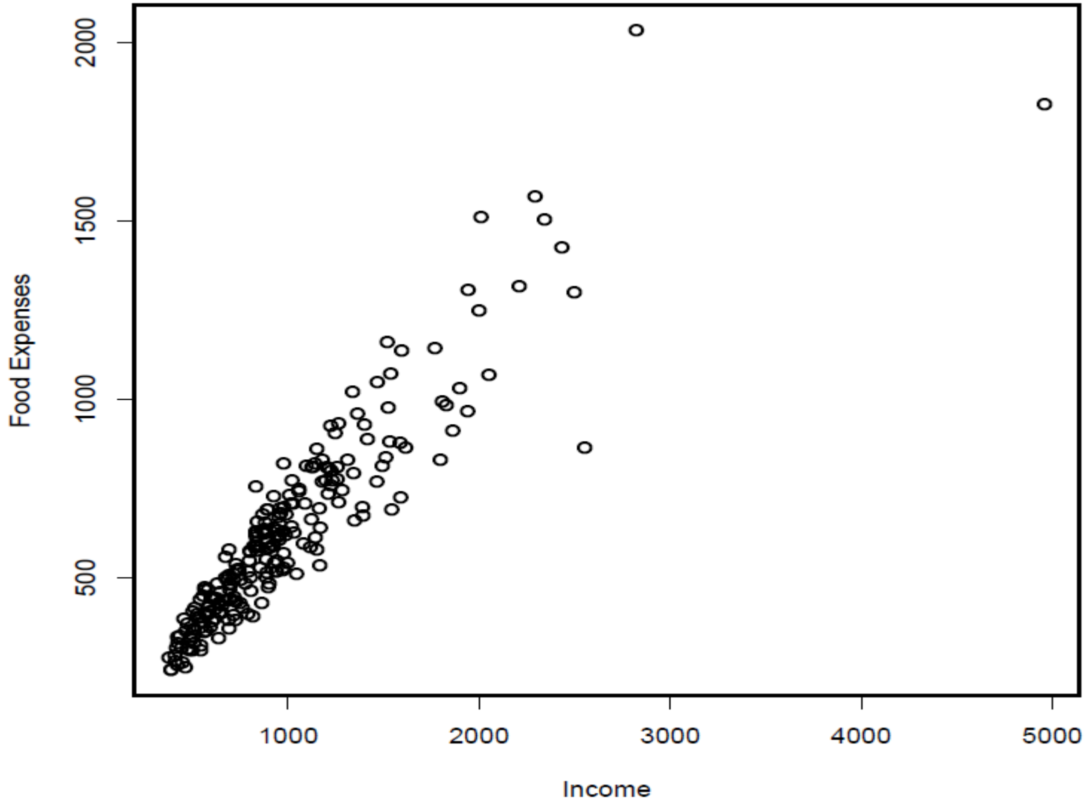
  OLS for parameter estimating

Quantile regression

- In linear model setup

  response = signal + i.i.d. error

  OLS for parameter estimating

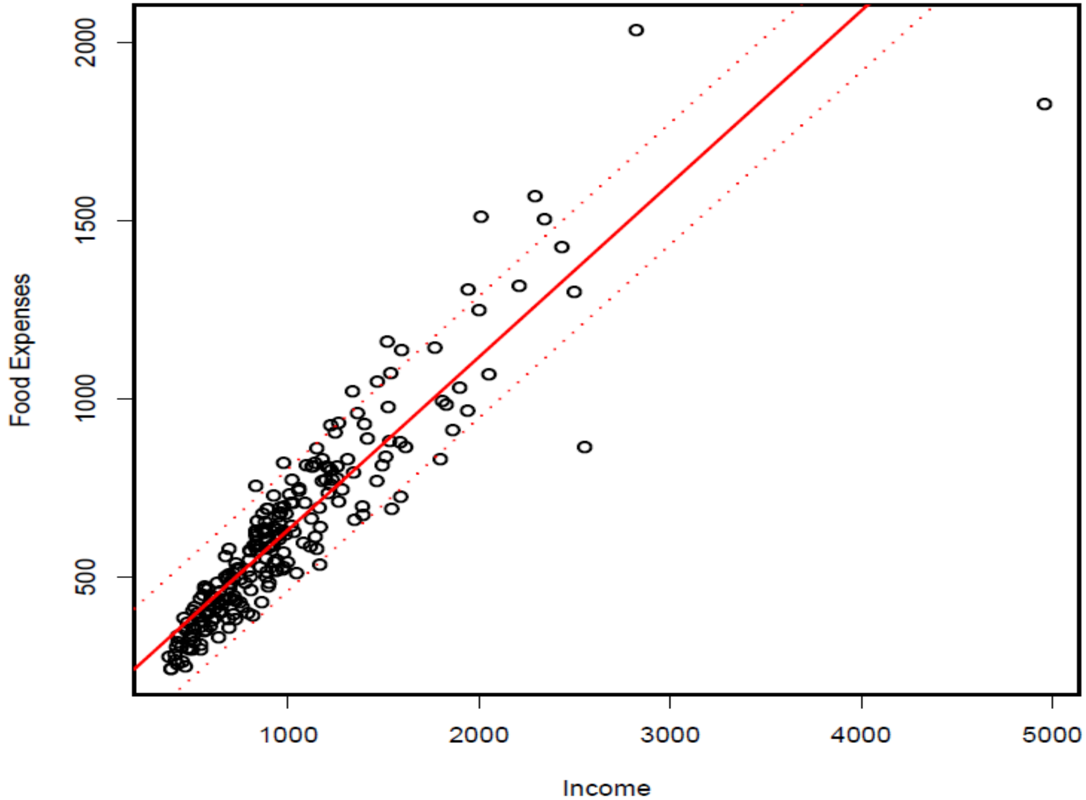


Quantile regression

- In linear model setup

  response = signal + i.i.d. error

  OLS for parameter estimating

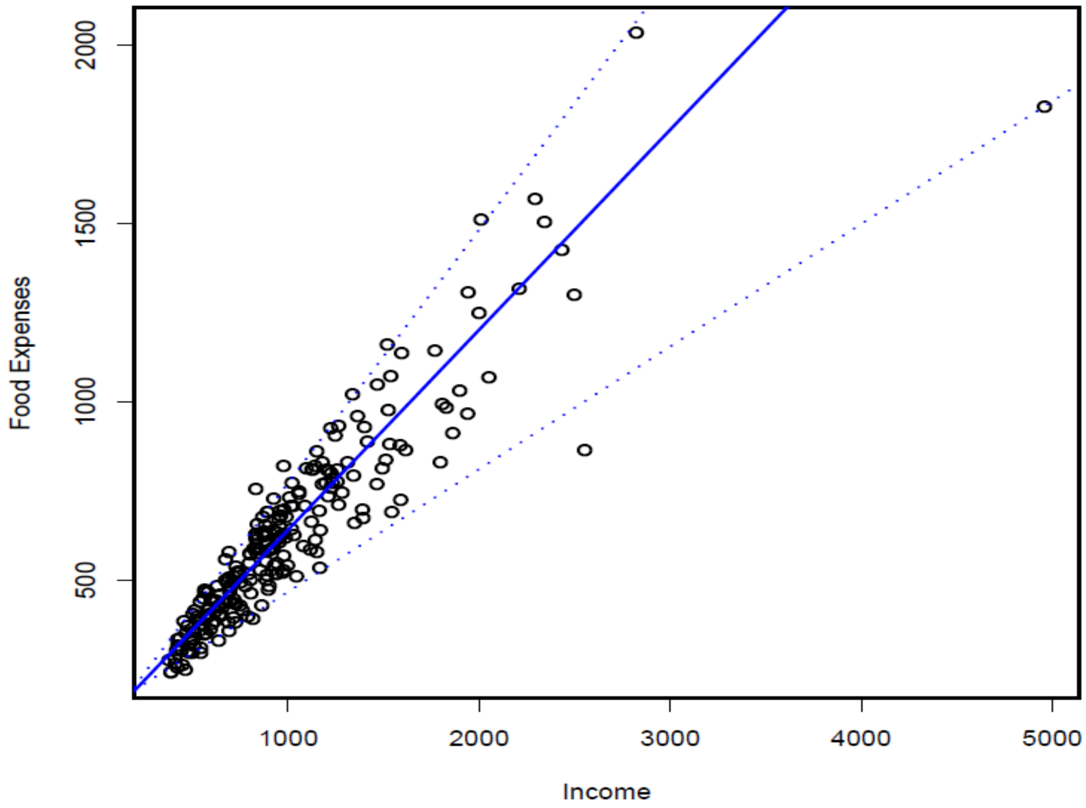


Quantile regression

- In linear model setup

  response = signal + i.i.d. error

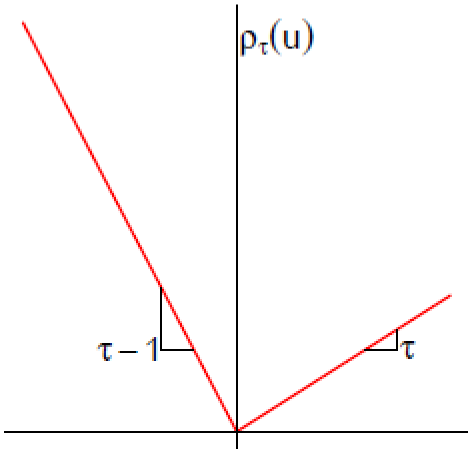
  OLS for parameter estimating



The Check function

- We define a loss function

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



- Note that if τ=0.5, $\rho_\tau(u) = |u|/2$
- Quantiles solve a simple optimization problem

$$\hat{\alpha}(\tau) = \text{argmin}\, \mathbb{E}\, \rho_\tau(Y - \alpha)$$

Quantile regression

- The usual linear regression solves

$$\min_{b \in \mathcal{R}^p} \sum_{i=1}^{n} (y_i - x_i^T b)^2$$

- Quantile regression solves (Koenker and Bassett 1978; Koenker 2005)

$$\min_{b \in \mathcal{R}^p} \sum_{i=1}^{n} \rho_\tau (y_i - x_i^T b)$$

Bayesian Analysis of quantile regression

- Bayesian methods of quantile regression
  - Skewed Laplace distribution (Yu and Moyeed 2001, Li et al. 2010)
  - Dirichlet process (Kottas and Krnjajic 2009)
  - Empirical likelihood (Yang and He 2012; Kim and Yang 2011)
- Advantage of Bayesian analysis
  - Easily incorporate prior information
  - Exact inference when sample size is small

Skewed Laplace distribution-based Bayesian analysis

- The skewed Laplace distribution

$$f(u|\sigma) = \tau(1-\tau)\sigma \exp\{-\sigma\rho_\tau(u)\}$$

- $$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + u_i,\ i = 1, \ldots, n$$

  u_i i.i.d. skewed-Laplace

- The joint likelihood

$$f(\boldsymbol{y}|\boldsymbol{X}) = \tau^n(1-\tau)^n\sigma^n \exp\{-\sigma \sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i^T\beta)\}$$

- Maximizing the likelihood is equivalent to minimizing

$$\min_{b \in \mathcal{R}^p} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T b)$$

Bayesian model selection

- Bayesian Lasso

$$\pi(\beta_k|\sigma,\lambda) = \sigma\lambda\exp\{-\sigma\lambda|\beta_k|\}$$

- Bayesian elastic net

$$\pi(\beta_k \mid \eta_1, \eta_2) = C(\eta_1, \eta_2)\frac{\eta_1}{2}\exp\{-\eta_1|\beta_k| - \eta_2\beta_k^2\}$$

- Bayesian group lasso

$$\pi(\boldsymbol{\beta}_g|\ \eta) = C_{d_g}\sqrt{\det(\mathbf{K}_g)}\eta^{d_g}\exp\left(-\eta\|\boldsymbol{\beta}_g\|_{\mathbf{K}_g}\right)$$

Empirical likelihood (EL)

- First introduced by Owen (1988)
  - constructing confidence interval for the mean

- Linear model (Owen 1991), general estimating equation (Qin and Lawless 1994)

- Given an estimating equation $\sum_{i=1}^{n} g(x_i, \boldsymbol{\theta}) = 0,$
  the EL is defined as

$$L(\boldsymbol{\theta}) = \sup\{\prod_{i=1}^{n} p_i \mid \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i g(x_i, \boldsymbol{\theta}) = 0,\ \text{and } 0 \le p_i \le 1\}.$$

Empirical likelihood

- Asymptotic properties
  - Wilk's theorem: the EL ratio $\xrightarrow{d} \chi_p^2$
  - The maximum EL estimator (MELE) is asymptotically normally distributed.

- Note: $g(x_i, \theta)$ usually need to be sufficiently smooth in θ for technical reasons

EL for quantile regression

➢ Taking directional derivative about β, the quantile regression estimates solves

$$\Sigma_{i=1}^{n}\phi_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i \approx 0,$$
$$\phi_\tau(t) = \tau - I_{[t<0]}$$

The EL for quantile regression is

$$L(\boldsymbol{\beta}) = \sup\{\prod_{i=1}^{n} p_i | \Sigma_{i=1}^{n} p_i\phi_\tau(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i = 0, \Sigma_{i=1}^{n} p_i = 1, 0 \le p_i \le 1\}$$

Model selection in EL settings

- Maximizing the penalized EL (Tang and Leng 2010)

$$\log(L(\boldsymbol{\theta})) - n\sum_{j=1}^{p} p_{\lambda}(|\theta_j|),$$

- The penalty can be Lasso (Tibshirani), elastic net (Zou and Hastie 2005), SCAD (Fan and Li 2001) ….

- Difficulty:
  - Computationally expensive, especially for quantile regression
  - Choice of the tuning parameter

Bayesian Model selection in EL

- Put a "spike and slab" prior on β_i

$$\theta_i I_{\{\beta_i=0\}} + (1-\theta_i) I_{\{\beta_i \neq 0\}} N(0, \sigma^2)$$

- The hierarchical model is

$$\begin{aligned}
\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta} &\sim L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = \sup\{\prod_{i=1}^{n} p_i \mid \Sigma_{i=1}^{n} p_i \phi_\tau(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta}) \boldsymbol{x}_i = 0, \Sigma_{i=1}^{n} p_i = 1, 0 \leq p_i \leq 1\} \\
\beta_i|\theta_i, \sigma &\sim \theta_i I_{\{\beta_i=0\}} + (1-\theta_i) I_{\{\beta_i \neq 0\}} N(0, \sigma^2), \quad i = 1, ..., p \\
\theta_i &\sim U(0,1), \quad i = 1, ..., p \\
1/\sigma^2 &\sim \Gamma(a, b) \quad a > 0, b > 0.
\end{aligned}$$

Asymptotic property

Theorem 1 *Under some regularity conditions, we have*

- *if* $\beta_j = 0$, *then* $P(\beta_j = 0 | \boldsymbol{X}, \boldsymbol{Y}) \to 1$ *in probability.*
- *if* $\beta_j \neq 0$, *the posterior distribution of* β_j *is approximately normally distributed.*

Proof:

$$L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = \exp\{-\frac{n}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})^T V_{12}^T V_{11}^{-1} V_{12}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}) + O_p(n^{-1/2})\}$$

Parameter estimation (1)

- Maximize the posterior likelihood?

$$f(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{Y}; \alpha, \eta) \propto L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) \prod_{i=1}^{p} \pi(\beta_i|\theta_i, \sigma^2) I_{(0,1)}(\theta_i) \Gamma(\frac{1}{\sigma^2}; a, b)$$

- Use Gibbs sampler

$$f(\sigma^{-2}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y}) \propto \Gamma(a + h/2, b + \frac{1}{2}\sum_{j \in H} \beta_j^2)$$

$$f(\theta_j|\boldsymbol{\beta}, \boldsymbol{\theta}_{-j}, \sigma^{-2}, \boldsymbol{X}, \boldsymbol{Y}) \propto \mathrm{Beta}(1 + I(\beta_j = 1), 1 + I(\beta_j \neq 1))$$

$$f(\beta_j|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}_{-j}) \propto L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y})\pi(\beta_j|\theta_j, \sigma^2).$$

where $H = \{j : \beta_j \neq 0\}$ $h = \#H$

Parameter estimation (1)

- Maximize the posterior likelihood?

$$f(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{X}, \boldsymbol{Y}; \alpha, \eta) \propto L(\boldsymbol{\beta} | \boldsymbol{X}, \boldsymbol{Y}) \prod_{i=1}^{p} \pi(\beta_i | \theta_i, \sigma^2) I_{(0,1)}(\theta_i) \Gamma(\frac{1}{\sigma^2}; a, b)$$

- Use Gibbs sampler

$$\begin{aligned}
f(\sigma^{-2} | \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{Y}) &\propto \Gamma(a + h/2, b + \frac{1}{2} \sum_{j \in H} \beta_j^2) \\
f(\theta_j | \boldsymbol{\beta}, \boldsymbol{\theta}_{-j}, \sigma^{-2}, \boldsymbol{X}, \boldsymbol{Y}) &\propto \mathrm{Beta}(1 + I(\beta_j = 1), 1 + I(\beta_j \neq 1))
\end{aligned}$$

$$f(\beta_j | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}_{-j}) \propto L(\boldsymbol{\beta} | \boldsymbol{X}, \boldsymbol{Y}) \pi(\beta_j | \theta_j, \sigma^2).$$

where $H = \{j : \beta_j \neq 0\}$ $h = \#H$

A mixture of the point mass at zero and a continuous distribution. Hard to sample

Parameter estimation (2)

- Use a Metropalis-Hastings (M-H) step to sample from

$$f(\beta_j|\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2, \beta_{-j})$$

- The M-H algorithm
  - Target: sample from $\pi(x)$

    choose a proposing distribution $q(x, y)$

    - Generate y from $q(x^{(j)}, \cdot)$ and u from $\mathcal{U}(0, 1)$
    - If $u \leq \alpha(x^{(j)}, y)$
      —set $x^{(j+1)} = y$.
    - Else
      —set $x^{(j+1)} = x^{(j)}$.

where

$$\alpha(x, y) = \min\left[\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right], \qquad \text{if } \pi(x)q(x, y) > 0$$
$$= 1, \qquad \text{otherwise.}$$

Parameter estimation (3)

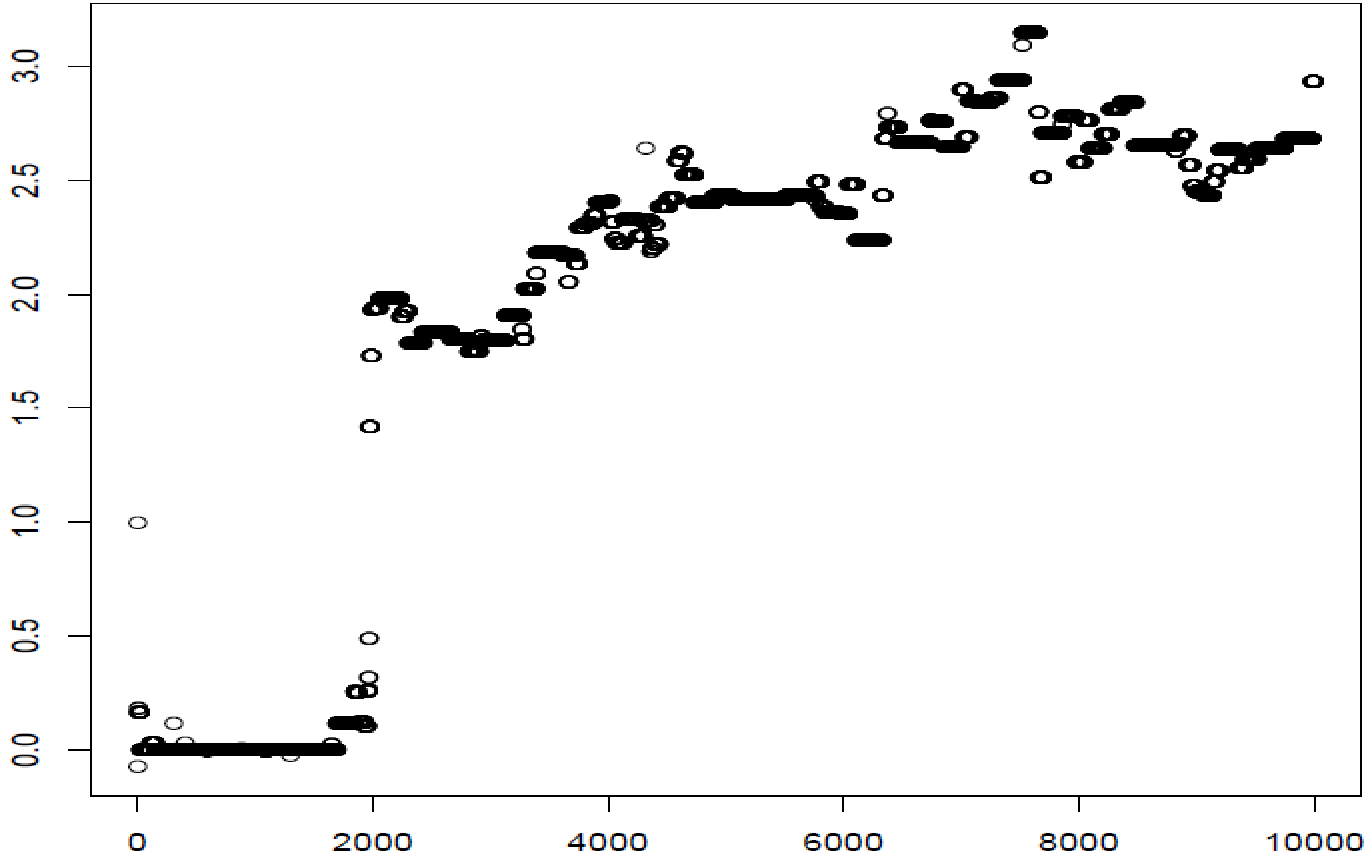
- How to choose the proposing distribution?
  - Random walk (Tierney 1994, Roberts et al. 1997):

    Given $\beta_i^{(t)}$ at the t-th step, the proposing distribution is

    $q(\beta - \beta_i^{(t)})$

    what q?
  - Kim and Yong (2011) proposed using a pre-specified distribution

Parameter estimation (3)

- 

Parameter estimation (4)

- If the EL $L(\beta)$ were smooth, the likelihood function can be approximated by $l(\beta_j) = \log(L(\beta_j, \boldsymbol{\beta}_{-j}))$

$$l(\beta_j) \approx l(\bar{\beta}_j) + \frac{1}{2} l''(\bar{\beta}_j)(\beta_j - \bar{\beta}_j)^2$$

where $l(\cdot)$ is maximized at $\bar{\beta}_j$

- Since the likelihood function convex, $v_j^{-2} = -l''(\bar{\beta}_j) > 0$ approximately

$$f(\beta_j | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}_{-j}) \quad \propto \quad \exp\{-\frac{1}{2} v_j^{-2} (\beta_j - \bar{\beta}_j)^2\} \pi(\beta_j | \theta_j, \sigma^2)$$

Parameter estimation (4)

- If the EL $L(\boldsymbol{\beta})$ were smooth, the likelihood function can be approximated by $l(\beta_j) = \log(L(\beta_j, \boldsymbol{\beta}_{-j}))$

$$l(\beta_j) \approx l(\bar{\beta}_j) + \frac{1}{2} l''(\bar{\beta}_j)(\beta_j - \bar{\beta}_j)^2$$

where $l(\cdot)$ is maximized at $\bar{\beta}_j$

- Since the likelihood function convex, $v_j^{-2} = -l''(\bar{\beta}_j) > 0$ approximately

$$f(\beta_j | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\beta}_{-j}) \quad \propto \quad \exp\{-\frac{1}{2} v_j^{-2} (\beta_j - \bar{\beta}_j)^2\} \pi(\beta_j | \theta_j, \sigma^2)$$

A mixture of the point mass at zero and a normal distribution. Easy to sample

Parameter estimation (5)

- $L(\beta)$ is not differentiable, we take $\bar{\beta}_j$ as the value that minimizes

$$\sum_{i=1}^{n} \rho_\tau(\tilde{y}_i - x_{ij}\beta_j).$$

where $\tilde{y}_i = y_i - \sum_{l \neq j} x_{il}\beta_l$

- Take v_j^{-2} as the bootstrap variance of $\bar{\beta}_j$

- The proposing distribution is chosen as

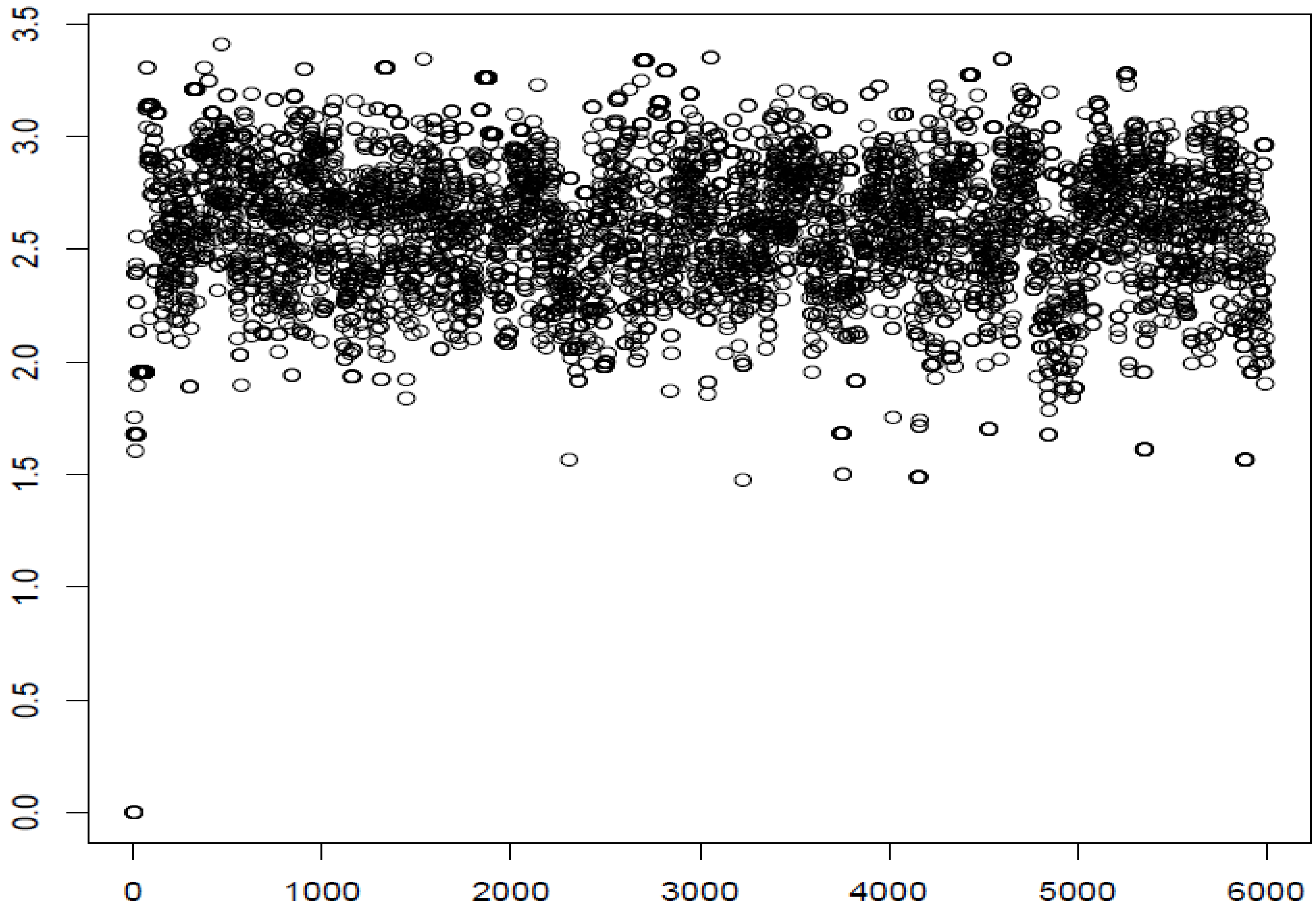
$$\exp\{-\frac{1}{2}v_j^{-2}(\beta_j - \bar{\beta}_j)^2\}\pi(\beta_j|\theta_j, \sigma^2)$$

Parameter estimation (5)

➢ mi … e that

wh

➢



Bayesian quantile regression weighted at multiple quantiles (1)

Consider the model with homogeneous errors

$$Y_i = \mu_1 + x_i\beta + u_i$$

with the τ_1th qunatile of u_i being zero

The asymptotic variance of β is inversely proportional to $f(\xi_{\tau_1})$ (*f* is the density of *u*)

If $f(\xi_{\tau_2}) > f(\xi_{\tau_1})$ for the τ_2th quantile ξ_{τ_2}

$$Y_i = \mu_2 + x_i^T\beta + w_i$$

$$\mu_2 = \mu_1 + \xi_{\tau_2} - \xi_{\tau_1}$$

$$w_i = u_i - (\xi_{\tau_2} - \xi_{\tau_1})$$

Bayesian quantile regression weighted at multiple quantiles (2)

We may consider minimizing

$$\sum_{i=1}^{n}[\rho_{\tau_1}(y_i-\mu_1-x_i\beta)+\rho_{\tau_2}(y_i-\mu_2-x_i\beta)]$$

More generally, given a set of quantile points $\tau_k \in (0,1)$ we may minimize

$$\sum_{k=1}^{m} a_k \sum_{i=1}^{n} \rho_{\tau_k}(y_i-\mu_k-\boldsymbol{x}_i^T\boldsymbol{\beta})$$

The corresponding EL is

$$L(\boldsymbol{\beta},\boldsymbol{\mu})=\sup\{\prod_{i=1}^{n} p_i | \sum_{k=1}^{m} a_k \sum_{i=1}^{n} p_i \phi_{\tau_k}(y_i-\mu_k-\boldsymbol{x}_i^T\boldsymbol{\beta})\boldsymbol{x}_i=\boldsymbol{0},$$

$$\sum_{i=1}^{n} p_i \phi_{\tau_k}(y_i-\mu_k-\boldsymbol{x}_i^T\boldsymbol{\beta})=0, \forall 1\le k\le m, \sum_{i=1}^{n} p_i=1, 0\le p_i\le 1\}$$

Bayesian quantile regression weighted at multiple quantiles (3)

Similarly define the corresponding Bayesian model and get the following asymptotic property

Theorem 2 *Under some regularity conditions, we have*

- *if* $\beta_j = 0$, *then* $P(\beta_j = 0 | \boldsymbol{X}, \boldsymbol{Y}) \to 1$ *in probability.*
- *if* $\beta_j \neq 0$, *the posterior distribution of* β_j *is approximately normally distributed.*

Simulation Study

In the simulations, we compared

- LASSO
- QR: quantile regression
- qrLasso: QR with Lasso (Li and Zhu 2008)
- bqrLasso: Bayesian regularized QR with Lasso (Li et al. 2010)
- BEQR: Bayesian EL-based QR
- BEQR.W: Bayesian EL-based QR weighted at multiple quantiles

Simulation Study-homogeneous errors

Simulation setup

$$y_i = \boldsymbol{x}_i^T \boldsymbol{\beta}_0 + u_i,\ i = 1, ..., n \qquad \boldsymbol{\beta}_0 = (3, 1.5, 0, 0, 2, 0, 0, 0)$$

where $u_i's$ have the τth quantile equal to 0.
The error distributions

$N(\mu, \sigma^2)$, with $\mu = 0, \sigma^2 = 9$.

$\text{Laplace}(\mu, b)$, with $\mu = 0, b = 3$.

$0.6N(\mu_1 - a, \sigma^2) + 0.4N(\mu_2 - a, \sigma^2)$

$\mu_1 = 2, \mu_2 = -2, \sigma^2 = 9$, and $a = 0.4$

$0.6\text{Laplace}(\mu_1 - a, b) + 0.4\text{Laplace}(\mu_2 - a, b)$

Simulation Study-homogeneous errors

quantile	Method	Error Distribution			
		normal	Laplace	normal mixture	Laplace mixture
$\theta = 0.9$	Lasso	0.43	0.79	0.61	0.80
	qrLasso	0.72	1.16	0.29	0.44
	bqrLasso	0.73	1.21	0.29	0.44
	BEQR	0.53	1.17	**0.20**	0.35
	BEQR.W	**0.41**	**0.65**	0.25	**0.29**
$\theta = 0.5$	Lasso	0.40	0.57	0.80	0.85
	qrLasso	0.53	0.54	0.41	0.46
	bqrLasso	0.51	0.51	0.44	0.48
	BEQR	**0.37**	**0.37**	0.65	0.67
	BEQR.W	0.39	0.42	**0.29**	**0.33**
$\theta = 0.1$	Lasso	0.43	0.78	0.66	0.45
	qrLasso	0.71	1.14	0.66	0.80
	bqrLasso	0.73	1.18	0.71	0.84
	BEQR	0.56	1.03	0.63	0.63
	BEQR.W	**0.41**	**0.56**	**0.36**	**0.41**

Simulation Study-homogeneous errors

θ/mean	Method	Error Distribution			
		normal TP/FP	Laplace TP/FP	normal mixture TP/FP	Laplace mixture TP/FP
mean	Lasso	3.00/2.30	3.00/2.16	3.00/2.20	3.00/2.12
$\theta = (0.9, 0.5, 0.1)$	BEQR.W	3.00/0.18	2.99/0.10	3.00/0.13	3.00/0.10
$\theta = 0.9$	qrLasso	3.00/2.78	2.94/2.83	3.00/2.66	3.00/2.74
	bqrLasso	3.00/0.86	2.80/0.68	3.00/0.42	3.00/0.42
	BEQR	2.96/0.19	2.63/0.15	3.00/0.08	2.99/0.12
$\theta = 0.5$	qrLasso	3.00/2.01	3.00/1.80	3.00/1.64	3.00/1.57
	bqrLasso	3.00/0.20	3.00/0.06	3.00/0.19	3.00/0.21
	BEQR	2.98/0.12	2.99/0.08	3.00/0.16	3.00/0.11
$\theta = 0.1$	qrLasso	3.00/2.78	2.92/2.83	3.00/2.66	2.99/2.74
	bqrLasso	2.99/0.86	2.80/0.68	3.00/0.42	2.98/0.42
	BEQR	2.97/0.19	2.68/0.15	2.97/0.08	2.95/0.12

Simulation Study-heterogeneous errors

Consider the model

$$y_i = \beta_{10} x_{i1} + \sum_{j=2}^{8} \beta_{j0} x_{ij} + x_{i1} \epsilon_i$$

ϵ_i are generated as in the i.i.d. case

$$\beta_0 = (\beta_{10}, \cdots, \beta_{80}) = (3, 1.5, 0, 0, 2, 0, 0, 0)$$

Simulation Study-heterogeneous errors

quantile	Method	Error Distribution			
		normal	Laplace	normal mixture	Laplace mixture
$\theta = 0.90$	Lasso	2.05	2.70	1.53	1.81
	qrLasso	0.95	1.32	0.45	0.54
	bqrLass	0.37	0.62	0.15	0.24
	BEQR	**0.28**	**0.46**	**0.10**	**0.16**
$\theta = 0.50$	Lasso	0.41	0.55	0.94	1.04
	qrLasso	0.31	0.32	0.58	0.67
	bqrLass	0.26	0.23	0.18	0.23
	BEQR	**0.20**	**0.16**	**0.14**	**0.19**
$\theta = 0.10$	Lasso	1.95	2.65	1.83	1.92
	qrLasso	0.55	1.03	0.50	0.54
	bqrLass	**0.35**	**0.62**	0.32	**0.40**
	BEQR	**0.35**	0.64	**0.30**	**0.40**

An application

- microRNA: small non-coding RNA binds to 3-UTR region of mRNAs

- Contradicting opinion about microRNA
  - Canalization effect (reduce gene expression variance)

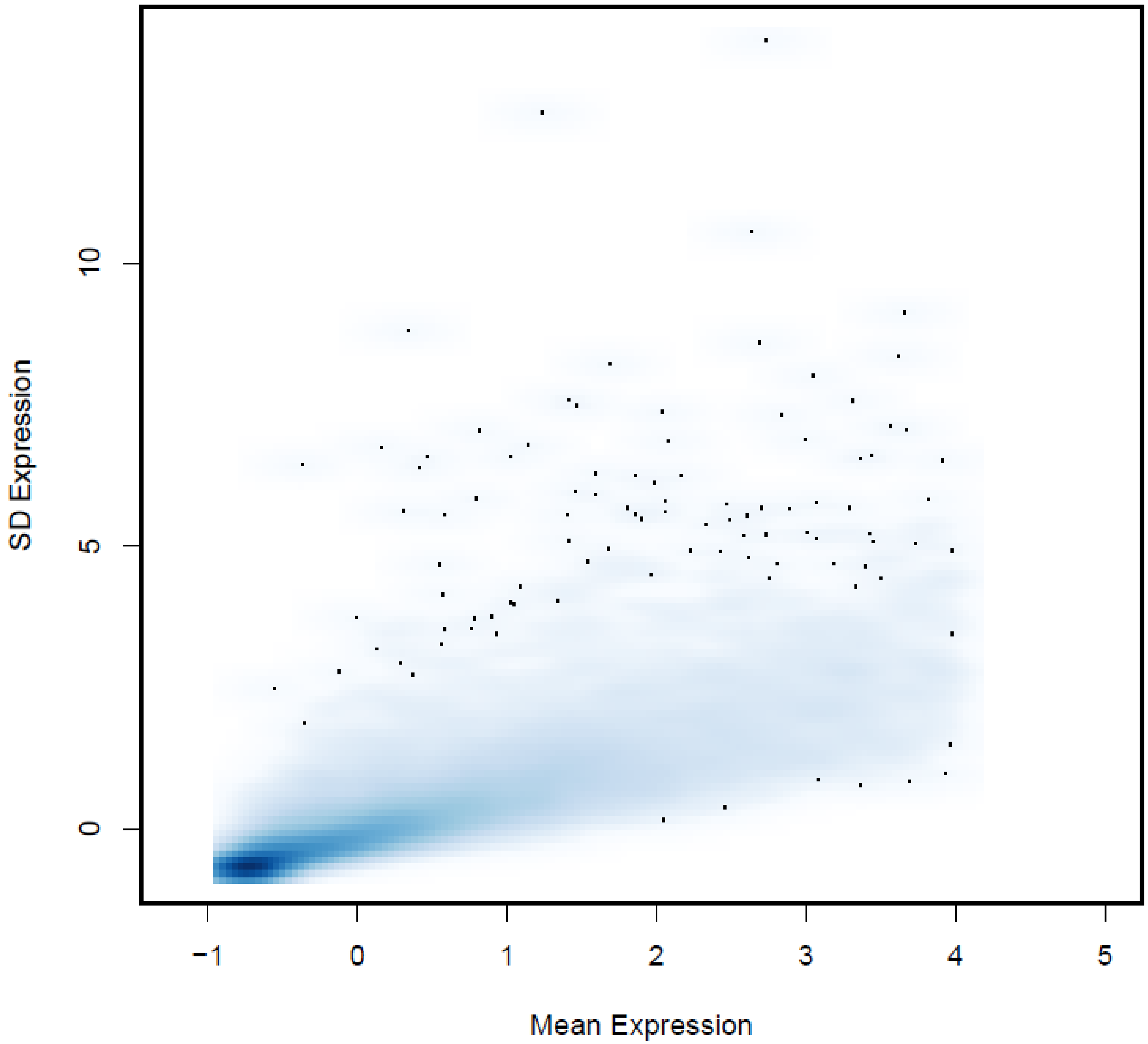
    Hornstein and Shomron, Nature Genetics, 2006

    Wu et al. Genome Research 2009
  - Increase gene expression variation

    Lu and Clark, Genome Research 2012

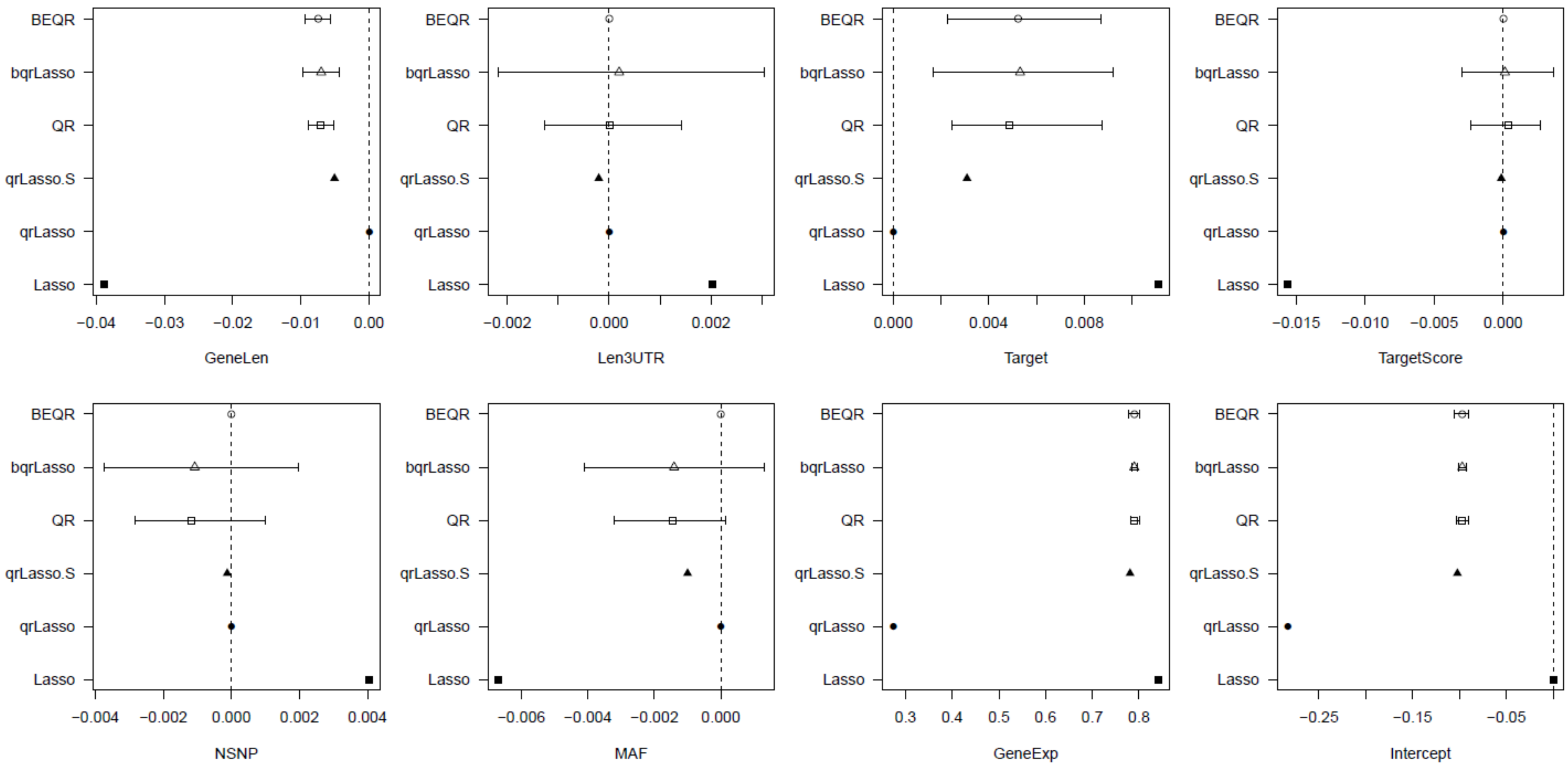
- mi TR
regio

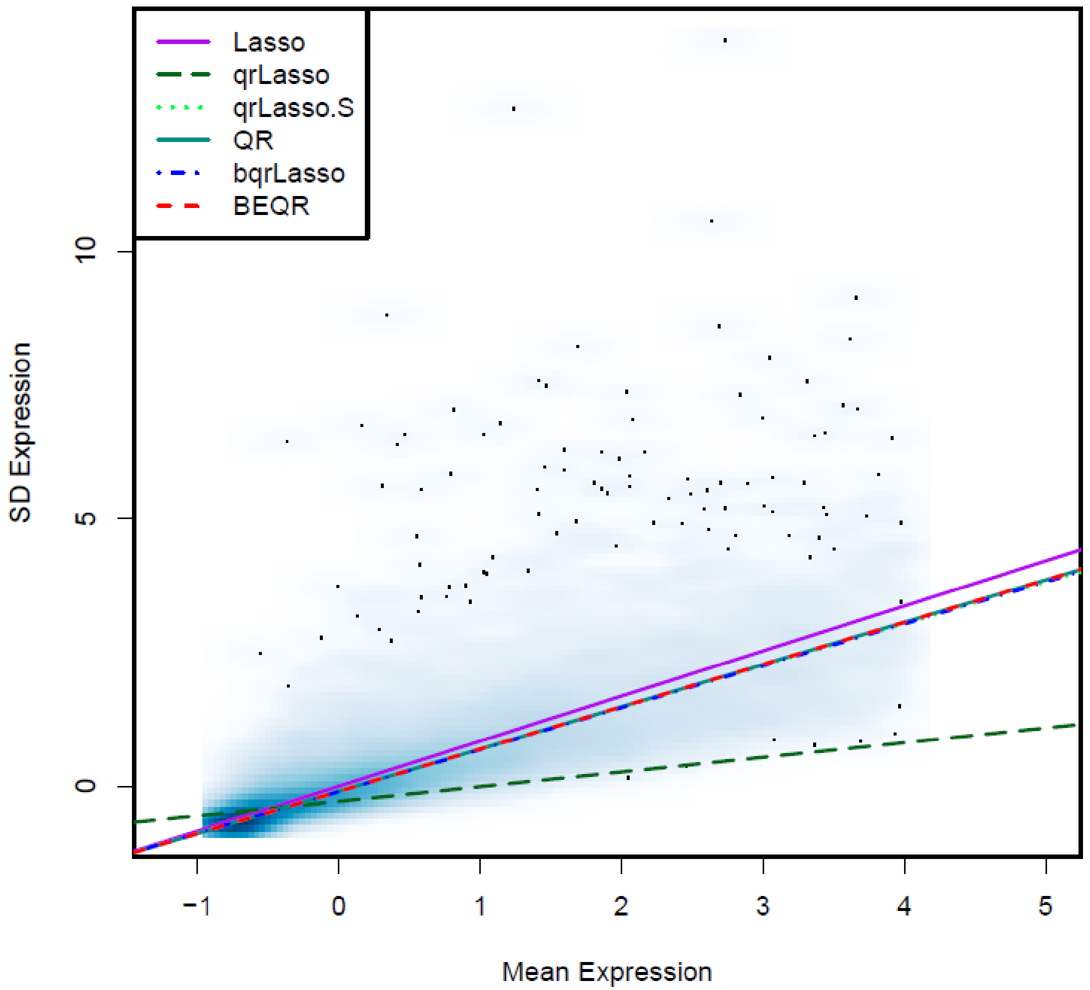
- Co
  - ce)
  -



An application

- Data: RNAseq data from 70 individuals
- ~ 20000 genes
- Y: expression variation for each gene
- Covariates:
  - Mean Expression
  - # of microRNA targets, target Score of 3`-UTR
  - # of SNP on 3`-UTR, Gene Length, length of 3`-UTR


BEQR
bqrLasso
QR
qrLasso.S
qrLasso
Lasso
−0.04 −0.03 −0.02 −0.01 0.00
GeneLen
−0.002 0.000 0.002
Len3UTR
0.000 0.004 0.008
Target
−0.015 −0.010 −0.005 0.000
TargetScore
−0.004 −0.002 0.000 0.002 0.004
NSNP
−0.006 −0.004 −0.002 0.000
MAF
0.3 0.4 0.5 0.6 0.7 0.8
GeneExp
−0.25 −0.15 −0.05
Intercept


Lasso
qrLasso
qrLasso.S
QR
bqrLasso
BEQR
SD Expression
Mean Expression
-1
0
1
2
3
4
5
0
5
10

Method	Error Distribution				
	$\tau = 0.1$	$\tau = 0.3$	$\tau = 0.5$	$\tau = 0.7$	$\tau = 0.9$
Lasso	0.08109	0.12607	0.13156	0.16285	0.11509
qrLasso	0.03792	0.12233	0.20715	0.22757	0.13412
qrLasso.S	0.03750	0.09058	0.12150	0.13073	0.10171
QR	0.03749	0.09056	0.12144	0.12743	0.09053
bqrLasso	0.03750	0.09057	0.12143	0.12742	0.09053
BEQR	0.03750	0.09059	0.12146	0.12743	0.09062

Conclusion

- Developed an EL based Baeysian model selection method in quantile regression

- Asymptotic property

- Simulation study shows that BEQR and BEQR.W performs better in general

- Disadvantage: cannot handle p >= n.

Postdoc position on bioinformatics available!

ruibinxi@math.pku.edu.cn