Semiparametric Expectile Regression for High-dimensional Heavy-tailed and Heterogeneous Data

Guan'ao Yan

School of Mathematical Sciences, Zhejiang University
Joint work with Jun Zhao and Yi Zhang

July 10th, 2019
at Dalian, China

Outline

Background

High dimensional data have been increasingly frequent in nowadays scientific areas like genomics, economics, finance and so on. The past two decades have witnessed the rapid development of high dimensional statistical analysis.

- **Homogeneity Assumption** for Data Structure

  Most of the existing work usually assume homogeneity on the data structure, like assuming a sequence of i.i.d. errors $\{\epsilon_i\}_{i=1}^n$ in regression framework(see, Bühlmann and Van De Geer (2011)).

Background

Plenty of work shows that high dimensional data display an opposite feature "**heterogeneity**", due to multi-sources data collection technology and error accumulation in data preprocessing.

- **Heteroscedasticity** in High Dimensional Data

  Plenty of works showed that Heteroscedasticity is of necessity to be considered for high dimensional questions, see Daye et al. (2012), Wang et al. (2012),Gu and Zou (2016) and Zhao et al. (2018). Variance of error terms are not identical, like $\mathbb{V}ar(\epsilon_i) = \mathrm{f}(\mathbf{x}_i, \epsilon_i)$, for some specific structure function f.

Motivation

Expectile Regression

Expectile regression was originally proposed in risk management area, proposed by Aigner et al. (1976) and Newey and Powell (1987). It uses the asymmetric squared loss function $\phi_\alpha(\cdot)$ defined below,

$$\phi_\alpha(r) = |\alpha - \mathbb{I}(r < 0)|r^2 = \begin{cases} \alpha r^2, & r \geq 0, \\ (1-\alpha)r^2, & r < 0. \end{cases} \tag{1}$$

And the α-th$(0 < \alpha < 1)$ expectile of random variable y is denoted by

$$m_\alpha(y) = \arg\min_{m \in \mathbb{R}} \mathbb{E}\phi_\alpha(y - m).$$

Motivation

- Sherwood and Wang (2007) used a partially linear additive quantile regression to analyse the heteroscedastic data.
- **Why Expectile Regression?**

  Expectile regression shares good properties:
  - different weights according to the positive and negative error respectively
  - differentiable loss function
  - less vulnerable for crossing problems than quantile regression

  Due to these good properties, expectile regression has great potential to analyze heterogeneity in semiparametric framework.

Motivation

- **Heavy Tailed Error Term**

  Most of existing literatures assume that regression errors follow the **gaussian or sub-gaussian distribution**, which is a relatively stringent assumption. More and more evidence is against it, especially in genetics and finance, where regression errors **do not have the tail of exponentially decreasing rate** (Fan et al. (2017)) or even worse is **heavy-tailed with only finite moments** (Zhao et al. (2018)).

Model Setting

Suppose that we have a high-dimensional data sample $\{Y_i, \mathbf{x}_i, \mathbf{z}_i\}$, $i = 1, \ldots, n$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, $i = 1, \ldots, n$ are independent and identically distributed p-dimensional covariates along with the common mean 0 and $\mathbf{z}_i = (z_{i1}, \ldots, z_{id})$ are d-dimensional covariates. Consider the data are from the following partially linear additive model,

$$Y_i = \mu_0 + \sum_{k=1}^{p} \beta_k^* x_{ik} + \sum_{j=1}^{d} g_{0j}(z_{ij}) + \epsilon_i = \mathbf{x}_i' \boldsymbol{\beta}^* + g_0(\mathbf{z}_i) + \epsilon_i, \quad (2)$$

where $g_0(\mathbf{z}_i) = \mu_0 + \sum_{j=1}^{d} g_{0j}(z_{ij})$.

Model Setting

- **Error Term**

  In consideration of data heterogeneity, we adopt the so called **variance heterogeneity** from Rigby and Stasinopoulos (1996) in their 'mean and dispersion additive model'. For example, error term can take this form

$$\epsilon_i = \sigma(\mathbf{x}_i, \mathbf{z}_i)\eta_i.$$

  In addition to heterogeneity, $\{\epsilon_i\}_{i=1}^n$ are independent and assumed to satisfy $m_\alpha(\epsilon_i|\mathbf{x}_i, \mathbf{z}_i) = 0$ for some specific $0 < \alpha < 1$.

  Thus, $m_\alpha(Y_i|\mathbf{x}_i, \mathbf{z}_i) = \mathbf{x}_i'\boldsymbol{\beta}^* + g_0(\mathbf{z}_i)$. Also, $\boldsymbol{\beta}^*$ and $g_0(\cdot)$ actually minimize the conditional $\mathbb{E}[\phi_\alpha(Y_i - \mathbf{x}'\boldsymbol{\beta} - g(\mathbf{z}))]$

$$(\boldsymbol{\beta}^*, g_0(\mathbf{z})) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, g\in\mathcal{G}}{\arg\min}\ \mathbb{E}[\phi_\alpha(Y_i - \mathbf{x}'\boldsymbol{\beta} - g(\mathbf{z}))]. \tag{3}$$

Model Setting

Nonparametric Part

In this article, the dimensionality of the nonparametric covariates $\mathbf{z}$, d, is **fixed**.

B-spline Method The nonparametric components $g_{0j}(\cdot), j = 1, \ldots, d$ in this model are approximated by a linear combination of B-spline basis functions.

Notations:

$\boldsymbol{\pi}(t) = (b_1(t), \ldots, b_{k_n+l+1}(t))'$: normalized B-spline basis functions

$\boldsymbol{\Pi}(\mathbf{z}_i) = (1, \boldsymbol{\pi}(z_{i1})', \ldots, \boldsymbol{\pi}(z_{id})')'$

Then there exists $\boldsymbol{\xi}_0 = (\xi_{00}, \boldsymbol{\xi}_{01}, \ldots, \boldsymbol{\xi}_{0d}) \in \mathbb{R}^{D_n}$, where $D_n = d(k_n + l + 1) + 1$, such that $\sup_{\mathbf{z}_i} |\boldsymbol{\Pi}(\mathbf{z}_i)'\boldsymbol{\xi}_0 - g_0(\mathbf{z}_i)| = O(k_n^{-r})$, see Stone (1985), Schumaker (2007)

Model Setting

Linear Part

The dimensionality of $\mathbf{x}$, $p = p(n)$, follows **high-dimensional setting**, and is much larger than n. We assume the true parameter $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)$ is **sparse**.
Let $A = \{j : \beta_j^* \neq 0, 1 \leq j \leq p\}$ be the active index set and its cardinality $q = q(n) = |A|$. Sparsity means that $q < n$ and all the left $(p - q)$ coefficients are exactly zero.
Notations:
$\boldsymbol{\beta}^* = ((\boldsymbol{\beta}_A^*)', \mathbf{0}')'$: $\boldsymbol{\beta}_A^* \in \mathbb{R}^q$ and $\mathbf{0}$ denotes a $(p - q)$ dimensional vector of zero
$\mathbf{X}_j$: the jth column of $\mathbf{X}$
$\mathbf{X}_A$: the submatrix of $\mathbf{X}$ that consists of its first q columns and denote by $\mathbf{X}_{A_i}$ the ith row of $\mathbf{X}_A$.

Model Setting

- **Regularized Framework**

Under sparsity assumption, we consider a regularized framework using a **general folded concave** penalty function $P_\lambda(t)$, for examples, the SCAD or MCP penalty:

- **SCAD**. The SCAD penalty is defined as for $\theta > 0$,

$$\begin{aligned} P_\lambda(\theta) &= \lambda\theta\mathbb{I}(\theta \le \lambda) + \frac{a\lambda\theta - (\theta^2+\lambda^2)/2}{a-1}\mathbb{I}(\lambda \le \theta \le a\lambda) \\ &\quad + \frac{(a+1)\lambda^2}{2}\mathbb{I}(\theta > a\lambda). \end{aligned} \tag{4}$$

- **MCP**. The MCP penalty has the following form:

$$P_\lambda(\theta) = \mathrm{sgn}(\theta)\lambda \int_0^{|\theta|} (1 - \frac{z}{\lambda b})dz, \tag{5}$$

where $b > 0$ is a fixed parameter and $\mathrm{sgn}(\cdot)$ is the sign function.

Estimators

The proposed estimators are obtained by solving the following optimization problem,

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p, \boldsymbol{\xi}\in\mathbb{R}^{D_n}}{\arg\min} \; L(\boldsymbol{\beta}, \boldsymbol{\xi}), \tag{6}$$

where $L(\boldsymbol{\beta}, \boldsymbol{\xi})$, the penalized expectile loss function for our model, is

$$L(\boldsymbol{\beta}, \boldsymbol{\xi}) = \frac{1}{n}\sum_{i=1}^{n} \phi_\alpha(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \boldsymbol{\Pi}(\mathbf{z}_i)'\boldsymbol{\xi}) + \sum_{j=1}^{p} P_\lambda(|\beta_j|). \tag{7}$$

Estimators

Denote by $\hat{\boldsymbol{\xi}} = (\hat{\xi}_0, \hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_d)$, then the estimator of $g_0(\mathbf{z}_i)$ is

$$\hat{g}(\mathbf{z}_i) = \hat{\mu} + \sum_{j=1}^{d} \hat{g}_j(z_{ij}),$$

where

$$\hat{\mu} = \hat{\xi}_0 + n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{d} \boldsymbol{\pi}(z_{ij})' \hat{\boldsymbol{\xi}}_j,$$

$$\hat{g}_j(z_{ij}) = \boldsymbol{\pi}(z_{ij})' \hat{\boldsymbol{\xi}}_j - n^{-1} \sum_{i=1}^{n} \boldsymbol{\pi}(z_{ij})' \hat{\boldsymbol{\xi}}_j.$$

Technical Conditions

Condition 1. There exists a positive constant C such that $\mathbb{E}\left(\epsilon_i^{2k}|x_i, \mathbf{z}_i\right) < C < \infty$ for all i and some $k \geq 1$.

Condition 2. There exists positive constants M_1 and M_2 such that $|x_{ij}| \leq M_1,\ \forall 1 \leq i \leq n,\ 1 \leq j \leq p_n$ and $\mathbb{E}\left(\delta_{ij}^4\right) \leq M_2,\ \forall 1 \leq i \leq n,\ 1 \leq j \leq q_n$. There exist finite positive constants C_1 and C_2 such that with probability one

$$C_1 \leq \lambda_{max}\left(n^{-1}X_A X_A'\right) \leq C_2, \quad C_1 \leq \lambda_{max}\left(n^{-1}\Delta_n \Delta_n'\right) \leq C_2.$$

Condition 3. For $r = m + v > 1.5$, $g_0(\cdot) \in \mathcal{G} \cap \mathcal{H}_r$. The dimension of the spline basis k_n satisfies $k_n \approx n^{1/(2r+1)}$

Condition 4. $q_n = O(n^{C_3})$ for some $C_3 < \frac{1}{3}$.

Oracle Study

The **oracle estimator** is defined as $(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\xi}}^*)$ with $\hat{\boldsymbol{\beta}}^* = (\hat{\boldsymbol{\beta}}_A^{*\prime}, \mathbf{0}'_{p-q})'$, where

$$(\hat{\boldsymbol{\beta}}_A^*, \hat{\boldsymbol{\xi}}^*) = \underset{\boldsymbol{\beta}\in\mathbb{R}^q, \boldsymbol{\xi}\in\mathbb{R}^{D_n}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \phi_\alpha(y_i - \mathbf{x}'_{A_i}\boldsymbol{\beta} - \boldsymbol{\Pi}(\mathbf{z}_i)'\boldsymbol{\xi}). \tag{8}$$

Theorem 1

Assume conditions 1-4 hold. Then the oracle estimator obtained by the optimization problem (8) satisfies

$$\| \hat{\boldsymbol{\beta}}_A^* - \boldsymbol{\beta}_A \| = O_p(\sqrt{n^{-1}q_n}), \tag{9}$$

$$n^{-1}\sum_{i=1}^{n}(\hat{g}(\mathbf{z}_i) - g_0(\mathbf{z}_i))^2 = O_p(n^{-1}(q_n + k_n)). \tag{10}$$

Asymptotic Results for Proposed Estimator

Condition 5. There exist positive constants C_4 and C_5 such that $2C_3 < C_4 < 1$ and

$$n^{(1-C_4)/2} \min_{1 \le j \le q_n} |\beta_j^*| \ge C_5,$$

Define $\mathcal{E}(\lambda)$ be the set of local minima of $L(\boldsymbol{\beta}, \boldsymbol{\xi})$ with the tuning parameter λ. The following theorem builds up the relationship between the oracle estimator and the penalized nonconvex optimization problem: with probability tending to one, the oracle estimator $(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\xi}}^*)$ is a local minimizer of $L(\boldsymbol{\beta}, \boldsymbol{\xi})$.

Asymptotic Results for Proposed Estimator

Guan'ao Yan

School of Mathematical Sciences, Zhejiang University
Joint work with Jun Zhao and Yi Zhang

Theorem 2

Assume Conditions 1-5 are satisfied. The tuning parameter $\lambda = o\left(n^{-(1-C_4)/2}\right)$, $q_n = o(n\lambda^2)$, $k_n = o(n\lambda^2)$ *and* $p = o\left((n\lambda^2)^k\right)$. *We have that with probability tending to one, the oracle estimator* $(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\xi}}^*)$ *lies in the set* $\mathcal{E}(\lambda)$ *consisting of local minima of* $L(\boldsymbol{\beta}, \boldsymbol{\xi})$, *i.e.,*

$$\mathbb{P}((\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\xi}}^*) \in \mathcal{E}(\lambda)) \to 1, \text{as } n \to \infty. \tag{11}$$

Simulation

- Data Generation

The response variable y is generated from the following sparse model,

$$y = x_6\beta_6 + x_{12}\beta_{12} + x_{15}\beta_{15} + x_{20}\beta_{20} + \sin(2\pi z_1) + z_2^3 + \epsilon, \quad (12)$$

where $\beta_j = 1$ for $j = 6, 12, 15$ and 20 and ϵ is independent of the covariates $\mathbf{x}$.

Firstly, the quasi-covariates $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_{p+2})'$ is generated from the multivariate normal distribution $N_{p+2}(\mathbf{0}, \Sigma)$ where $\Sigma = (\sigma_{ij})_{(p+2)\times(p+2)}$, $\sigma_{ij} = 0.5^{|i-j|}$ for $i, j = 1, \ldots, p+2$. Then we set $x_1 = \sqrt{12}\Phi(\tilde{x}_1)$ where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and $\sqrt{12}$ scales x_1 to have standard deviation 1. Furthermore, let $z_1 = \Phi(\tilde{x}_{25})$ and $z_2 = \Phi(\tilde{x}_{26})$, $x_i = \tilde{\mathbf{x}}_i$ for $i = 2, \ldots, 24$ and $x_i = \tilde{x}_{i+2}$ for $i = 25, \ldots, p$.

Simulation

1 **Case 1:** Homogeneous Error Term

When ϵ is homogeneous, we investigate the performances of our proposed method in coefficient estimation, nonparametric approximation accuracy and model selection.
Error term follows two kinds of distributions:

- Standard normal distribution $N(0,1)$;
- Standard t-distribution with degrees of freedom 5 (t_5), where $\mathbb{E}\epsilon^{4+\delta}$ exists for $\delta \in (0,1)$.

2 **Case 2:** Heterogeneous Error Term

When ϵ has heteroscedasticity and here we assume the heteroscedastic structure $\epsilon = 0.70x_1\varsigma$.
ς is independent of x_1 and has the following 2 kinds of distribution:

- Standard normal distribution $N(0, 1)$;
- Standard t-distribution with degrees of freedom 5 (t_5), where $\mathbb{E}\epsilon^{4+\delta}$ exists for $\delta \in (0, 1)$.

Simulation

We set sample size $n = 300$ and covariate dimension $p = 400$ or 600. In addition to our E-SCAD method, for comparing purpose, E-LASSO and Oracle are included in the work, too. We repeat the simulation procedure 100 times and evaluate the performance in term of the following criteria:

- AE: the average absolute estimation error defined by $\sum_i^p |\hat{\beta}_j - \beta_j^*|$.
- SE: the average square estimation error defined by $\sqrt{\sum_i^p |\hat{\beta}_j - \beta_j^*|^2}$.
- ADE: the average of the average absolute deviation (ADE) of the fit of the nonlinear part defined by $\frac{1}{n}\sum_{i=1}^n |\hat{g}(\mathbf{z}_i) - g_0(\mathbf{z}_i)|$
- Size: the average number of nonzero regression coefficients $\hat{\beta}_j \neq 0$ for $j = 1, \ldots, p$. In the heteroscedastic case, given the role of x_1, the true size of our data generation model supposes to be 5.
- F: the frequency that $x_6, x_{12}, x_{15}, x_{20}$ are selected during the 100 repetitions.

Simulation

Results for Case 1

Table 1: Simulation results for homogeneous erros when $n = 300$, $p = 400$.

	Criteria	$N(0,1)$ E-SCAD	$N(0,1)$ E-Lasso	$N(0,1)$ Oracle	t_5 E-SCAD	t_5 E-Lasso	t_5 Oracle
$\alpha = 0.10$	AE	0.24(0.09)	1.01(0.24)	0.23(0.08)	0.43(0.19)	1.47(0.58)	0.35(0.14)
	SE	0.14(0.05)	0.37(0.08)	0.13(0.05)	0.22(0.09)	0.53(0.13)	0.20(0.08)
	ADE	0.19(0.04)	0.17(0.04)	0.16(0.04)	0.23(0.06)	0.24(0.06)	0.23(0.06)
	Size	4.67(0.83)	17.60(4.23)	-	6.25(2.15)	17.14(6.16)	-
	F	100	100	-	100	100	-
$\alpha = 0.50$	AE	0.21(0.08)	0.86(0.19)	0.19(0.08)	0.34(0.14)	1.13(0.31)	0.25(0.09)
	SE	0.11(0.04)	0.30(0.06)	011(0.04)	0.16(0.05)	0.38(0.09)	0.15(0.05)
	ADE	0.25(0.03)	0.26(0.04)	0.13(0.03)	0.27(0.03)	0.27(0.04)	0.27(0.04)
	SIZE	5.14(0.97)	19.26(3.79)	-	7.56(2.36)	20.29(5.54)	-
	F	100	100	-	100	100	-
$\alpha = 0.90$	AE	0.22(0.08)	0.98(0.23)	0.21(0.08)	0.41(0.17)	1.44(0.42)	0.36(0.12)
	SE	0.13(0.05)	0.36(0.07)	0.13(0.05)	0.22(0.09)	0.53(0.11)	0.21(0.08)
	ADE	0.42(0.03)	0.41(0.05)	0.43(0.05)	0.41(0.04)	0.41(0.06)	0.43(0.06)
	SIZE	4.14(0.51)	16.25(4.42)	-	5.32(1.68)	16.63(4.97)	-
	F	100	100	-	100	100	-

Simulation

Results for Case 2

Table 2: Simulation results for heteroscedastic errors when $n = 300$, $p = 400$.

	Criteria	*Hetero*_$N(0, 1)$			*Hetero*_t_5		
		E-SCAD	E-Lasso	Oracle	E-SCAD	E-Lasso	Oracle
$\alpha = 0.10$	AE	0.76(0.27)	1.94(0.42)	0.83(0.17)	1.55(1.08)	3.01(1.22)	1.13(0.33)
	SE	0.47(0.20)	0.59(0.11)	0.56(0.11)	0.71(0.33)	0.83(0.21)	0.72(0.18)
	ADE	0.54(0.10)	0.67(0.10)	0.26(0.08)	0.73(0.16)	0.73(0.15)	0.36(0.13)
	Size	6.79(1.73)	24.05(5.51)	-	11.97(5.17)	28.25(8.82)	-
	F,F1	100, 87	100, 97	-	99, 89	100, 94	-
$\alpha = 0.50$	AE	0.31(0.14)	1.26(0.30)	0.28(0.11)	0.47(0.16)	1.49(0.23)	0.41(0.13)
	SE	0.18(0.08)	0.41(0.09)	0.16(0.06)	0.25(0.10)	0.50(0.09)	0.22(0.07)
	ADE	0.38(0.24)	0.37(0.24)	0.18(0.05)	0.44(0.18)	0.43(0.18)	0.32(0.12)
	Size	4.64(0.78)	21.78(4.86)	-	6.49(1.42)	20.43(3.01)	-
	F,F1	100, 0	100, 8	-	100, 0	100, 8	-
$\alpha = 0.90$	AE	0.74(0.27)	1.76(0.36)	0.82(0.16)	1.31(1.00)	2.94(1.23)	1.12(0.31)
	SE	0.47(0.20)	0.59(0.09)	0.56(0.11)	0.64(0.32)	0.84(0.19)	0.71(0.18)
	ADE	0.49(0.14)	0.76(0.22)	0.25(0.09)	0.43(0.18)	0.68(0.13)	0.38(0.13)
	Size	6.21(1.39)	19.54(4.91)	-	9.34(4.78)	26.06(9.19)	-
	F,F1	100, 88	100, 96	-	99, 80	100, 87	-

Real Data

Low infant birth weight has always been a comprehensive quantitative trait, as it affects directly the post-neonatal mortality, infant and childhood morbidity, as well as its life-long body condition. With at total of **65 observations** contained in this genetic set, the gene expression profiles were collected from mother's placenta and assayed by Illumina Expression Beadchip v3 for the **24,526 genes transcripts** and then normalized in a quantile method. As the **infant's birth weight** was recorded along with **the age of mother**, **gestational age**, **parity**, **maternal blood cotinine level** and **mother's BMI index**, we consider using this data set to depict the infant birth weight's determinants.

- $n = 65$
- **Linear Covariates:** Normalized genetic data, parity, gestational age, maternal blood, cotinine level and BMI;
- **Nonlinear Covariates:** Age of mother

Real Data

Table 3: Numeric results at three expectile levels

	Criteria	All Data		Random Partition	
		E-SCAD	E-LASSO	E-SCAD	E-LASSO
$\alpha = 0.1$	L_1	0.66	0.67	0.90(0.21)	0.74(0.17)
	L_2	0.12	0.11	0.30(0.07)	0.26(0.06)
	$\hat{\mathcal{A}}_\alpha$	7.00	8.00	5.72(1.91)	8.19(2.74)
	$\hat{\mathcal{A}}_\alpha \bigcap \hat{\mathcal{A}}_{0.5}$	1	1	3.86 (1.6433)	1.16(0.39)
$\alpha = 0.3$	L_1	0.60	0.53	0.81(0.17)	0.59(0.13)
	L_2	0.10	0.09	0.27(0.06)	0.19(0.04)
	$\hat{\mathcal{A}}_\alpha$	9.00	19.00	9.00(2.72)	13.94(3.03)
	$\hat{\mathcal{A}}_\alpha \bigcap \hat{\mathcal{A}}_{0.5}$	3	3	2.27(1.13)	3.25(1.18)
$\alpha = 0.5$	L_1	0.38	0.34	0.89(0.20)	0.41(0.08)
	L_2	0.06	0.05	0.30(0.06)	0.13(0.02)
	$\hat{\mathcal{A}}_\alpha$	14.00	20.00	4.72(1.83)	20.25(2.67)
	$\hat{\mathcal{A}}_\alpha \bigcap \hat{\mathcal{A}}_{0.5}$	-	-	-	-

Real Data

Table 4: Top 6 Frequent Covariates Selected at Three Expectile Weight Levels among 100 Partitions

E-SCAD $\alpha = 0.1$		E-SCAD $\alpha = 0.3$		E-SCAD $\alpha = 0.5$	
Variables	Frequency	Variables	Frequency	Variables	Frequency
PTPN3	34	GPR50	46	PTPN3	33
FXR1	40	FXR1	49	GPR50	40
GPR50	43	EPHA3	50	FXR1	41
LEO1	43	LEO1	59	LEO1	44
SLCO1A2	63	LOC388886	65	SLCO1A2	65
Gestational age	79	Gestational age	97	Gestational age	83

Real Data

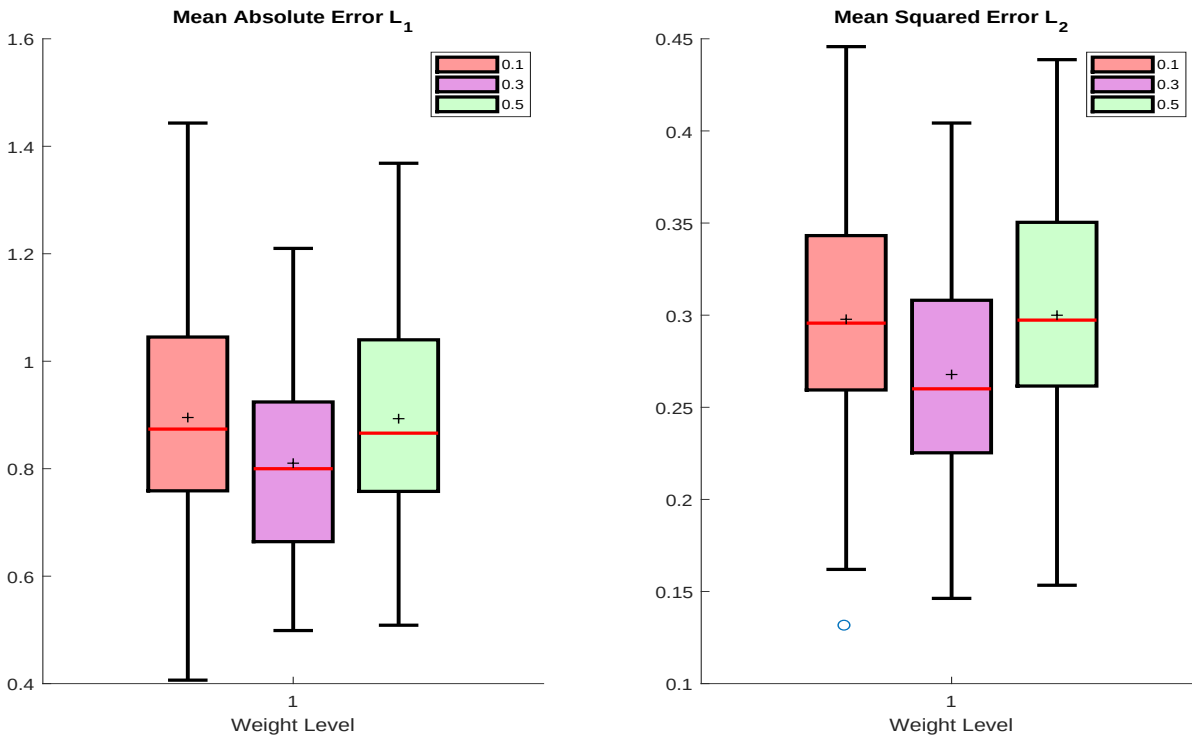

Figure 1: Boxplots of Prediction Errors.

Real Data

Conclusion:

- In Table 3, the selected genes and corresponding cardinalities are different for different weight levels, which means that different levels of birth weight are influenced by different genes, an indication of heterogeneity in the data.
- Furthermore, an interesting observation arises that the scenarios $\alpha = 0.1$ and $\alpha = 0.5$ perform similarily while the scenario $\alpha = 0.3$ displays some different characteristics.

Thanks

Thanks for your attention!