

Double-Penalized Quantile Regression in Partially Linear Models

Yunlu Jiang

Department of Statistics, College of Economics, Jinan University, Guangzhou, China Email: jiangyl@jnu.edu.cn

Received 11 March 2015; accepted 22 April 2015; published 23 April 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

Abstract

In this paper, we propose the double-penalized quantile regression estimators in partially linear models. An iterative algorithm is proposed for solving the proposed optimization problem. Some numerical examples illustrate that the finite sample performances of proposed method perform better than the least squares based method with regard to the non-causal selection rate (NSR) and the median of model error (MME) when the error distribution is heavy-tail. Finally, we apply the proposed methodology to analyze the ragweed pollen level dataset.

Keywords

Quantile Regression, Partially Linear Model, Heavy-Tailed Distribution

1. Introduction

Since semiparametric regression models combine both parametric and nonparametric components, they are much more flexible than the linear regression model, and are easier interpretation of the effect of each variable than completely nonparametric regressions. Therefore, semiparametric regression models are very popular models in practical applications. In this paper, we consider a partially linear model

$$Y = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + g(\boldsymbol{Z}) + \boldsymbol{\epsilon},\tag{1}$$

where β is a *p*-dimensional unknown parameter vector with its true value β_0 , $g(\cdot)$ is a twice-differentiable unknown smooth function, ϵ is independent of X and is a random error satisfying $E(\epsilon | X, Z) = 0$. Since [1] first applied the partially linear model to study the relationship between weather and electricity sales, this model had received a considerable amount of research in the past several decades.

In practice, many potential explanatory variables should be involved in this model, but the number of impor-

tant ones is usually relatively small. Therefore, selection of important explanatory variables is often one of the most important goals in the real data analysis. In this paper, we are interested in automatic selection, and estimation for parametric components, and treat $g(\cdot)$ as a nuisance effect. There are many authors developed several approaches in the literature. For example, in the kernel smoothing framework, [2] first extended the penalized least squares criterion to partially linear models. [3] introduced a class of sieve estimators using a penalized least squares technique for semiparametric regression models. [4] studied variable selection for semiparametric regression models. [5] considered variable selection for partially linear models when the covariates were measured with additive errors. [6] combined the ideas of profiling and adaptive Elastic-Net [7] to select the important variables in X. In the framework of spline smoothing, [8] achieved sparsity in the linear part by using the SCAD-penalty [9] for partially linear models for high dimensional data, but the nonparametric function was estimated by the polynomial regression splines. [10] applied a shrinkage penalty on parametric components to obtain the significant variables and used the smoothing spline to estimate the nonparametric component.

It is very important to note that many of those methods are closely related to the classical least squares method. It is well known that the least squares method is not robust and can produce large bias when there are outliers in the dataset. Therefore, the outliers can give rise to serious problems for the least squares based methods in variable selection. In this article, we propose the double-penalized quantile regression estimators. Based on the quantile regression loss function (check function), we apply a shrinkage penalty for parametric parts to yield the significant variables, and use the smoothing spline to estimate the nonparametric component. Simulation studies illustrate that the proposed method can achieve a consistent variable selection when there are outliers in the dataset or the error term follows a heavy-tailed distribution.

The rest of this paper is organized as follows. In Section 2, we first introduce the double-penalized quantile regression estimators in a partially linear regression model, and then propose an iterative algorithm to solve the proposed optimization problem. In Section 3, simulation studies are conducted to compare the finite-sample performance of the existing and proposed methods. In Section 4, we apply the proposed method to analyze a real data analysis. Finally, we conclude with a few remarks in Section 5.

2. Methodology and Main Results

2.1. Double-Penalized Quantile Regression Estimators

Suppose that $(Y_i, X_i^T, Z_i)^T$, $i = 1, \dots, n$ satisfy a following partially linear regression model,

$$Y_{i} = \boldsymbol{X}_{i}^{\mathrm{T}}\boldsymbol{\beta} + g\left(\boldsymbol{Z}_{i}\right) + \epsilon_{i}, i = 1, \cdots, n.$$

$$\tag{2}$$

Without loss of generality, we assume that $Z_i \in [0,1]$, and g(t) is in the Sobolev space V, where V is defined by

$$V = \left\{ g(t) : g, g' \text{ are absolutely continuous, and } \int_0^1 \left\{ g''(t) \right\}^2 dt < \infty \right\}.$$

To simultaneously achieve the selection of important variables and the estimation of the nonparametric function $g(\cdot)$, [10] proposed a double-penalized least squares (DPLS) estimators by minimizing

$$L_{n}(\boldsymbol{\beta},g(\cdot)) = \frac{1}{2}\sum_{i=1}^{n} \{Y_{i} - \boldsymbol{X}_{i}^{\mathrm{T}}\boldsymbol{\beta} - g(\boldsymbol{Z}_{i})\}^{2} + \frac{n\lambda_{1}}{2}\int_{0}^{1} \{g''(t)\}^{2} dt + n\sum_{j=1}^{p} p_{\lambda_{2}}(|\beta_{j}|).$$

where $p_{\lambda_2}(|\beta_j|)$ is nonnegative and nondecreasing functions in $|\beta_j|$. Under some regular conditions, [10] proved that the proposed estimators could be as efficient as the oracle estimator.

To our knowledge, the ordinary least squares (OLS) estimator is not robust. If there are outliers in the dataset or the error follows a heavy-tailed distribution, it can product the large bias. In contrast to the least squares method, quantile regression introduced by [11] serves as a robust alternative since the asymptotic properties of quantile regression estimator do not depend on the variance of the error. In the following, we introduce a double-penalized quantile regression (DPQR) in partially linear models. For $0 < \tau < 1$, the DPQR estimators can be obtained by minimizing the following function,

$$Q_n\left(\boldsymbol{\beta}, g\left(\cdot\right)\right) = \sum_{i=1}^n \rho_\tau \left\{Y_i - \boldsymbol{X}_i^{\mathrm{T}} \boldsymbol{\beta} - g\left(\boldsymbol{Z}_i\right)\right\} + \lambda_1 \int_0^1 \left\{g''(t)\right\}^2 \mathrm{d}t + \sum_{j=1}^p p_{\lambda_2}\left(\left|\boldsymbol{\beta}_j\right|\right),\tag{3}$$

where $\rho_{\tau}(t) = \tau t - tI(t < 0)$.

Let $G = (g(Z_1), \dots, g(Z_n))^T$ and the order statistics of a random sample $\{Z_1, \dots, Z_n\}$ be $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ and $h_i = Z_{(i+1)} - Z_{(i)}$ for $i = 1, \dots, n-1$. According to [12], we have

$$\int_0^1 \left\{ g''(t) \right\}^2 \mathrm{d}t = \boldsymbol{G}^\mathrm{T} \boldsymbol{K} \boldsymbol{G},$$

where K is an $n \times n$ matrix given by $K = QR^{-1}Q^{T}$, Q is the $n \times (n-2)$ matrix of second differences, with entries

$$q_{j-1,j} = h_{j-1}^{-1}, q_{jj} = -h_{j-1}^{-1} - h_j^{-1}, q_{j+1,j} = h_j^{-1}, \text{ for } j = 2, \dots, n-1,$$
$$q_{ij} = 0 \text{ for } |i-j| \ge 2.$$

R is a symmetric tridiagonal matrix of order (n-2) with elements r_{ii}

$$r_{ii} = \frac{1}{3} (h_{i-1} + h_i) \text{ for } i = 2, \dots, n-1,$$

$$r_{i,i+1} = r_{i+1,i} = \frac{1}{6} h_i \text{ for } i = 2, \dots, n-2.$$

Therefore, Equation (3) can be rewrote as

$$Q_{n}\left(\boldsymbol{\beta},g\left(\cdot\right)\right) = \sum_{i=1}^{n} \rho_{\tau}\left\{Y_{i} - \boldsymbol{X}_{i}^{\mathrm{T}}\boldsymbol{\beta} - g\left(Z_{i}\right)\right\} + \lambda_{1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{G} + \sum_{j=1}^{p} p_{\lambda_{2}}\left(\left|\boldsymbol{\beta}_{j}\right|\right).$$
(4)

2.2. Algorithm

To solve the optimization problem (4), we propose the following iterative algorithm. The estimation procedures are stated as follows:

Step 1 Given $\hat{\boldsymbol{G}}_n = (\hat{g}(Z_1), \dots, \hat{g}(Z_n))^T$, obtaining the estimator $\hat{\boldsymbol{\beta}}_n$ by minimizing the following objective function,

$$\sum_{i=1}^{n} \rho_{\tau} \left\{ Y_{i} - \boldsymbol{X}_{i}^{\mathrm{T}} \boldsymbol{\beta} - \hat{g}\left(\boldsymbol{Z}_{i}\right) \right\} + \sum_{j=1}^{p} p_{\lambda_{2}}\left(\left| \boldsymbol{\beta}_{j} \right| \right).$$

Step 2 Given $\hat{\beta}_n$, obtaining \hat{G}_n by solving

$$\min_{\boldsymbol{G}}\left[\sum_{i=1}^{n} \rho_{\tau}\left\{Y_{i}-\boldsymbol{X}_{i}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{n}-g\left(\boldsymbol{Z}_{i}\right)\right\}+\lambda_{\mathrm{I}} \boldsymbol{G}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{G}\right]$$

Step 3 Repeat Step 1 and Step 2 until convergence.

Remark 1 In the above algorithm, we first obtain the initial estimators by minimizing the following objective function

$$\sum_{i=1}^{n} \left\{ Y_{i} - \boldsymbol{X}_{i}^{\mathrm{T}} \boldsymbol{\beta} - g\left(\boldsymbol{Z}_{i}\right) \right\}^{2} + \lambda_{1} \boldsymbol{G}^{\mathrm{T}} \boldsymbol{K} \boldsymbol{G}.$$

$$\tag{5}$$

Let $B_{\lambda_1} = (I + \lambda_1 K)^{-1}$. Given $\boldsymbol{\beta}$ and λ_1 , we obtain $\hat{\boldsymbol{G}}_n$ by (5), $\hat{\boldsymbol{C}} = \boldsymbol{P} \cdot (\boldsymbol{V} - \boldsymbol{V}^T \boldsymbol{\beta})$

$$\hat{\boldsymbol{G}}_{n} = \boldsymbol{B}_{\lambda_{1}} \left(\boldsymbol{Y} - \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta} \right).$$

where $\boldsymbol{Y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$, and $\boldsymbol{X} = (\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)^{\mathrm{T}}$. We plug $\hat{\boldsymbol{G}}_n$ into (5), we have

$$\hat{\boldsymbol{\beta}}_{n} = \left[\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{B}_{\lambda_{1}}\right)^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{B}_{\lambda_{1}}\right)\boldsymbol{X}\right]^{-1}\boldsymbol{X}^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{B}_{\lambda_{1}}\right)^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{B}_{\lambda_{1}}\right)\boldsymbol{Y}.$$

Finally, the estimator \hat{G}_n of nonparametric component is obtained as follows:

$$\hat{\boldsymbol{G}}_{n} = \boldsymbol{B}_{\lambda_{1}} \left(\boldsymbol{Y} - \boldsymbol{X}^{\mathrm{T}} \hat{\boldsymbol{\beta}}_{n} \right).$$

Remark 2 Since the check function $\rho_r(\cdot)$ is not smooth, we use the majorization-minimization (MM) algorithm introduced by [13] to optimize **Step 1** and **Step 2**.

Advocated in [14], the check function $\rho_{\tau}(t)$ can be approximated by its perturbation for some small $\epsilon > 0$,

$$\rho_{\tau}^{\epsilon}(t) = \rho_{\tau}(t) - \frac{\epsilon}{2} \log(\epsilon + |t|).$$

Furthermore, $\rho_{\tau}^{\epsilon}(t)$ can be majorized at $t^{(k)}$ by the following surrogate function given in [14],

$$\tilde{\rho}_{\tau}^{\epsilon}\left(t\left|t^{(k)}\right)=\frac{1}{4}\left[\frac{t^{2}}{\epsilon+\left|t^{(k)}\right|}+\left(4\tau-2\right)t+c\right].$$

The penalty functions can be approximated by the local quadratic approximation advocated in [15]

$$p_{\lambda_2}\left(\left|\beta_j\right|\right) \approx p_{\lambda_2}\left(\left|\beta_j^{(0)}\right|\right) + \frac{1}{2} \left\{p_{\lambda_2}'\left(\left|\beta_j^{(0)}\right|\right) / \beta_j^{(0)}\right\} \left(\beta_j^2 - \beta_j^{(0)2}\right), \text{ for } \beta_j \approx \beta_j^{(0)}.$$

The minimization problem in **Step 1** and **Step 2** is a quadratic function after above these approximations, and can be solved in closed form. In our implementation, we set $\epsilon = 10^{-6}$.

3. Simulation Study

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed estimators. We simulate 100 data sets from the following model (6) with sample sizes n = 80,100,150.

$$Y_i = X_i^{\mathrm{T}} \boldsymbol{\beta}_0 + g\left(Z_i\right) + \epsilon_i, \ i = 1, \cdots, 100.$$
(6)

In this simulation, we choose $\boldsymbol{\beta}_0 = (1, 2, 2, 2, 0, 0, 0)^T$, $g(z) = \sin(z)$, X_i 's follows a 8-dimensional standard normal distribution, and the error term follows the following two distributions: standard normal distribution (N(0,1)), and standard Cauchy distribution. We consider $\tau = 0.25, 0.5, 0.75$. Although the choice of penalized parameters λ_1 and λ_2 is very important, we take $\lambda_1 = \lambda_2 = \log n/n$ in this paper. Meanwhile, we take the penality function $p_{\lambda_2}(|\beta_j|) = \lambda_2 |\beta_j| / |\tilde{\beta}_{nj}|^2$, $j = 1, \dots, p$, where $\boldsymbol{\beta}_n = (\tilde{\beta}_{n1}, \dots, \tilde{\beta}_{np})^T$ is a \sqrt{n} -consistent estimator to $\boldsymbol{\beta}_0$. For example, we can use least squares estimator.

We compare our proposed estimators (DPQR) with the DPLS estimators and Oracle estimator based on the quantile regression. In order to measure the finite-sample performance, for the parameteric component, we calculate the non-causal selection rate (NSR) [9], the positive selection rate (PSR) [16] as well as the median of the model error (MME) advocated by [9], where the model error is defined as follows:

$$ME = \left(\hat{\boldsymbol{\beta}}_{n} - \boldsymbol{\beta}_{0}\right)^{\mathrm{T}} E\left[\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\right] \left(\hat{\boldsymbol{\beta}}_{n} - \boldsymbol{\beta}_{0}\right).$$

The simulation results are reported in Table 1 and Table 2. From Table 1, we can see that all these methods obtain the same PSR and NSR when the error term follows the standard normal distribution, but the DPLS estimator yields smaller the MME than the DPQR estimator and Oracle estimator. Whereas, when the error follows a Cauchy distribution, we find from Table 2 that the MME of our proposed method is smaller than the DPLS method. In variable selection, the PSR is around 1 for all three methods. However, what distinguishes DPQR from DPLS is NSR. Indeed, the NSR of the DPQR estimator is as close 1 as that of the oracle estimator, while the NSR of the DPLS estimator is about 30%. This illustrates that our proposed method leads to a consistent variable selection to errors with heavy tails.

4. Real Data Application

In this section, we illustrate our proposed double-penalized quantile regression method through application to

Y. L. Jiang

~ .

Cable 1. Simulation results under normal error.								
п	τ	Method	PSR	NSR	MME			
		DPQR	1.0000	1.0000	0.1175			
	0.25	DPLS	1.0000	1.0000	0.0489			
		Oracle	1.0000	1.0000	0.1231			
		DPQR	1.0000	1.0000	0.2753			
80	0.50	DPLS	1.0000	1.0000	0.0440			
		Oracle	1.0000	1.0000	0.2263			
		DPQR	1.0000	1.0000	0.2908			
	0.75	DPLS	1.0000	1.0000	0.0452			
		Oracle	1.0000	13K 13K 13K 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0 1.0000 1.0000 0	0.3961			
		DPQR	1.0000	1.0000	0.0860			
	0.25	DPLS	1.0000	1.0000	0.0378			
		Oracle	1.0000	1.0000	0.0818			
		DPQR	1.0000	1.0000	0.2462			
100	0.50	DPLS	1.0000	1.0000	0.0385			
		Oracle	1.0000	1.0000	0.2792			
		DPQR	1.0000	1.0000	0.2838			
	0.75	DPLS	1.0000	1.0000	0.0357			
		Oracle	1.0000	1.0000	0.2343			
		DPQR	1.0000	1.0000	0.0607			
	0.25	DPLS	1.0000	1.0000	0.0233			
		Oracle	1.0000	1.0000	0.0665			
		DPQR	1.0000	1.0000	0.1990			
150	0.50	DPLS	1.0000	1.0000	0.0232			
		Oracle	1.0000	1.0000	0.1952			
		DPQR	1.0000	1.0000	0.2584			
	0.75	DPLS	1.0000	1.0000	0.0279			
		Oracle	1.0000	1.0000	0.2403			

1 able 2. Simulation results under cauchy error.								
n	τ	Method	PSR	NSR	MME			
		DPQR	1.0000	1.0000	0.2201			
	0.25	DPLS	0.9900	0.3575	8.4638			
		Oracle	1.0000	1.0000	0.2377			
	0.50	DPQR	1.0000	1.0000	0.3513			
80		DPLS	0.9925	0.3375	16.318			
		Oracle	1.0000	1.0000	0.3724			
	0.75	DPQR	1.0000	1.0000	0.4126			
		DPLS	0.9800	0.2900	12.773			
		Oracle	1.0000	1.0000	0.4961			
		DPQR	1.0000	1.0000	0.1671			
	0.25	DPLS	0.9975	0.2950	11.938			
		Oracle	1.0000	1.0000	0.1520			
		DPQR	1.0000	1.0000	0.3007			
100	0.50	DPLS	0.9925	0.3075	8.7854			
		Oracle	1.0000	1.0000	0.2822			
		DPQR	1.0000	1.0000	0.4061			
	0.75	DPLS	0.9950	0.2525	16.399			
		Oracle	1.0000	1.0000	0.4292			
		DPQR	1.0000	1.0000	0.1134			
	0.25	DPLS	0.9975	0.2150	18.382			
		Oracle	1.0000	1.0000	0.1281			
		DPQR	1.0000	1.0000	0.2996			
150	0.50	DPLS	0.9925	0.2675	11.863			
		Oracle	1.0000	1.0000	0.1885			
		DPQR	1.0000	1.0000	0.3409			
	0.75	DPLS	0.9975	0.2150	18.382			
		Oracle	1.0000	1.0000	0.3621			



Fi	gure 1	. (a)) Histogram of	v and (b) v against dav
		• (••)	, increased and of	,	, , and and a court

Table 3. Estimated regression coefficients from the ragweed pollen level data.									
Estimator	<i>X</i> ₁	<i>X</i> ₂	<i>X</i> ₃	x_2^2	x_{3}^{2}	$x_1 x_2$	$X_1 X_3$	<i>x</i> ₂ <i>x</i> ₃	
DPLS	1.5567	0.1774	0.1640	0.0000	0.0000	0.0000	0.0000	0.0000	
DPQR	1.8086	-0.1948	-0.1368	0.0037	0.0069	0.0000	0.0000	0.0000	

the Ragweed Pollen Level data, which are collected in Kalamazoo, Michigan during the 1993 ragweed season. This dataset consists of 87 observations, and contains the following four variables: ragweed (the daily ragweed pollen level (grains/m³)), rain x_1 (indicator of significant rain of the following day: 1 = at least 3 hours of steady or brief but intense rain, 0 = otherwise), temperature x_2 (temperature of following day (degrees Fahrenheit)), wind speed x_3 (wind speed forecast for following day (knots)), and day (day number in the current ragweed pollen season). The ragweed is the response variable, and the rest are the explanatory variables.

The goal is to understand the effect of the explanatory variables on ragweed, and to obtain accurate models to predict the ragweed. According to [17], we take $y = \sqrt{ragweed}$. Histogram of y in Figure 1(a) indicates that the response is rather skewed. Therefore, there are outliers in the response or the error follows a heavy-tailed distribution. In addition, we plot y against day in Figure 1(b). From Figure 1(b), we can find that there is a strong nonlinear relationship between y and the day number. As a consequence, a semiparametric regression model with a nonparametric baseline g (day) is very reasonable. In this paper, we add some quadratic and interaction terms, and consider a more complex semiparametric regression model.

In the following, we apply the DPLS method and DPQR method to fit the semiparametric regression model. For the DPQR method, we take $\tau = 0.5$. The results are summarized in **Table 3**. From **Table 3**, we find that there also exists a nonlinear relationship between y and temperature and between y and wind speed by the DPQR method.

5. Discussion

In this paper, we introduced a double-penalized quantile regression method in partially linear models. The merits of our proposed methodology were illustrated via simulation studies and a real data analysis. According to numerical simulations, our proposed method could achieve a consistent variable selection when there were outliers in the dataset or the error followed a heavy-tailed distribution.

Acknowledgements

Jiang's research is partially supported by the National Natural Science Foundation of China (No.11301221).

References

- Engle, R.F., Granger, C.W., Rice, J. and Weiss, A. (1986) Semiparametric Estimates of the Relation between Weather and Electricity Sales. *Journal of the American Statistical Association*, 81, 310-320. http://dx.doi.org/10.1080/01621459.1986.10478274
- [2] Fan, J.Q. and Li, R.Z. (2004) New Estimation and Model Selection Procedures for Semiparametric Modeling in Longitudinal Data Analysis. *Journal of the American Statistical Association*, **99**, 710-723. http://dx.doi.org/10.1198/016214504000001060
- Bunea, F. (2004) Consistent Covariate Selection and Post Model Selection Inference in Semi-Parametric Regression. *The Annals of Statistics*, 32, 898-927. http://dx.doi.org/10.1214/009053604000000247
- [4] Li, R.Z. and Liang, H. (2008) Variable Selection in Semiparametric Regression Modeling. *The Annals of Statistics*, 36, 261-286. http://dx.doi.org/10.1214/00905360700000604
- [5] Liang, H. and Li., R.Z. (2009) Variable Selection for Partially Linear Models with Measurement Errors. *Journal of the American Statistical Association*, **104**, 234-248. <u>http://dx.doi.org/10.1198/jasa.2009.0127</u>
- [6] Chen, B.C., Yu, Y., Zou, H. and Liang, H. (2012) Profiled Adaptive Elastic-Net Procedure for Partially Linear Models with High-Dimensional Covariates. *Journal of Statistical Planning and Inference*, 142, 1733-1745. http://dx.doi.org/10.1016/j.jspi.2012.02.035
- [7] Zou, H. and Zhang, H.H. (2009) On the Adaptive Elastic-Net with a Diverging Number of Parameters. *The Annals of Statistics*, 37, 1733-1751. <u>http://dx.doi.org/10.1214/08-AOS625</u>
- [8] Xie, H.L. and Huang, J. (2009) SCAD-Penalized Regression in High-Dimensional Partially Linear Models. *The Annals of Statistics*, 37, 673-696. <u>http://dx.doi.org/10.1214/07-AOS580</u>
- [9] Fan, J.Q. and Li, R.Z. (2001) Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association*, 96, 1348-1360. <u>http://dx.doi.org/10.1198/016214501753382273</u>
- [10] Ni, X., Zhang, H.H. and Zhang, D.W. (2009) Automatic Model Selection for Partially Linear Models. *Journal of multivariate Analysis*, 100, 2100-2111. <u>http://dx.doi.org/10.1016/j.jmva.2009.06.009</u>
- Koenker, R. (2005) Quantile Regression. Cambridge University Press, Cambridge. <u>http://dx.doi.org/10.1017/CBO9780511754098</u>
- [12] Green, P.J. and Silverman, B.W. (1994) Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach. Chapman & Hall, London.
- [13] Hunter, D.R. and Lange, K. (2004) A Tutorial on MM Algorithms. *The American Statistician*, 58, 30-37. http://dx.doi.org/10.1198/0003130042836
- [14] Hunter, D.R. and Lange, K. (2000) Quantile Regression via an MM Algorithm. Journal of Computational and Graphical Statistics, 9, 60-77. <u>http://dx.doi.org/10.2307/1390613</u>
- [15] Zou, H. and Li, R.Z. (2008) One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models. *The Annals of Statistics*, 36, 1509-1533. <u>http://dx.doi.org/10.1214/00905360700000802</u>
- [16] Chen, J.H. and Chen, Z.H. (2008) Extended Bayesian Information Criteria for Model Selection with Large Model Spaces. *Biometrika*, 95, 759-771. <u>http://dx.doi.org/10.1093/biomet/asn034</u>
- [17] Ruppert, D., Wand, M.P. and Carroll, R. J. (2003) Semiparametric Regression. Cambridge University Press, Cambridge. <u>http://dx.doi.org/10.1017/CBO9780511755453</u>