Hacettepe Journal of Mathematics and Statistics
Volume 47 (3) (2018), 721 – 739

Penalized empirical likelihood based variable selection for partially linear quantile regression models with missing responses

Xinrong Tang* and Peixin Zhao†‡

Abstract

In this paper, we consider variable selection for partially linear quantile regression models with missing response at random. We first propose a profile penalized empirical likelihood based variable selection method, and show that such variable selection method is consistent and satisfies sparsity. Further more, to avoid the influence of nonparametric estimator on the variable selection for the parametric components, we also propose a double penalized empirical likelihood variable selection method. Some simulation studies and a real data application are undertaken to assess the finite sample performance of the proposed variable selection methods, and simulation results indicate that the proposed variable selection methods are workable.

Keywords: Quantile regression, Partially linear model, Variable selection, Penalized empirical likelihood.

2000 AMS Classification: 62G08, 62G20

Received : 12.06.2016 *Accepted* : 14.08.2016 *Doi* : 10.15672/HJMS.201614120416

*Department of Logistics, Chongqing Technology and Business University, Chongqing 400067, China, Email: 360223993@qq.com

†College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China, Email: zpx81@163.com

‡Corresponding Author.

1. Introduction

Quantile regression, first studied by Koenker and Bassett [1], is less sensitive to outliers, and has been proved to be a useful alternative to mean regression models. Since its flexibility and ability of describing the entire conditional distribution of the response, quantile regression has been deeply investigated in the literature and extensively applied in econometrics, social sciences and biomedical studies. For example, statistical inferences for parametric quantile regression models and pure nonparametric quantile regression models are considered by [1]-[4] and [5]-[7], respectively. In addition, because of the more flexibility of partially linear models, many authors are interested in the quantile regression modeling for the following partially linear model

$$Y_i = X_i^T \beta + \theta(U_i) + \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where $\beta = (\beta_1, \cdots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters, $\theta(\cdot)$ is an unknown smooth function, X_i and U_i are covariates, Y_i is the response, and ε_i is the model error with $P(\varepsilon_i \leq 0 | X_i, U_i) = \tau$.

For model (1.1), Lee [8] proposes an estimation method of the parametric component, and proves the proposed estimator is semiparametric efficient. Sun [9] improves the estimation procedure proposed by Lee [8], and proposes a new semiparametric efficient estimation method. He and Liang [10] proposes a quantile regression estimation procedure for model (1.1) when some covariates are measured with errors. Chen and Khan [11] proposes a quantile regression estimation method for model (1.1) with censored data. In addition, based on empirical likelihood method, Lv and Li [12] considered the confidence interval construction for model (1.1) with missing response. However, the variable selection for such partially linear quantile regression models with missing responses seems not to be studied in the above references. Taking this issue into account, in this paper, we consider the variable selection for such partially linear quantile regression models with missing responses.

Variable selection is an important topic in high-dimensional statistical modeling. Many variable selection methods have been developed in the literature, including the sequential approach, penalized likelihood approach and information-theoretic approach. But most of these variable selection methods are computationally expensive. In addition, although the penalized likelihood based variable selection method is computational efficiency and stability, in many situations a well-defined likelihood function is not easy to construct. The empirical likelihood based variable selection procedure, which is constructed based on a set of estimating equations, can overcome this problem, and is more robust.

Recently, the penalized empirical likelihood based variable selection methods have been considered by some authors. For example, Tang and Leng [13] propose a penalized empirical likelihood method for parameter estimation and variable selection problems with diverging numbers of parameters. Ren and Zhang [14] considered the variable selection for moment restriction model based on the penalized empirical likelihood method. Variyath et al. [15] proposed a variable selection method by combining the information criteria and empirical likelihood method. Wu et al. [16] considered the variable selection for the linear regression model with censored data by using the empirical likelihood method. More works for the empirical likelihood based variable selection studies can be found in [17]-[19], and among others.

This article also contributes to the rapidly growing literature on the penalized empirical likelihood, and proposes a class of penalized empirical likelihood based variable selection methods for partially linear quantile regression models with missing response at random. More specifically, we assume that the covariates X_i and U_i can be observed directly, and the response Y_i may be missing. That is, we have the incomplete observations $\{Y_i, \delta_i, X_i, U_i\}$, $i = 1, \cdots, n$ from model (1.1), where $\delta_i = 0$ if Y_i is missing, otherwise $\delta_i = 1$. In addition, we assume that Y_i is missing at random (MAR). Here, the MAR mechanism means that

$$P(\delta_i = 1|Y_i, X_i, U_i) = P(\delta_i = 1|X_i, U_i). \tag{1.2}$$

Under the assumption (1.2), we first propose a profile penalized empirical likelihood method to select important variables in model (1.1), and show that such variable selection method is consistent and satisfies the sparsity. Further more, to avoid the influence of nonparametric estimator on the variable selection for the parametric components, we also propose a double penalized empirical likelihood variable selection method, which is also consistent and satisfies the sparsity in theory. In addition, we carry out some simulation studies to assess the performance of the proposed variable selection method, and simulation results indicate that our proposed methods are workable.

The rest of the paper is organized as follows. In Section 2, a profile penalized empirical likelihood based variable selection method is proposed, and some asymptotic properties, such as the consistency and sparsity, of the proposed variable selection method are derived. In Section 3, a double penalized empirical likelihood based variable selection method is proposed, and some asymptotic properties of the variable selection method are derived, including the consistency and sparsity. In Section 4, some simulation studies and a real data analysis are conducted to assess the performances of the proposed variable selection procedures. The proofs of all asymptotic results are provided in Section 5.

2. Variable selection via profile penalized empirical likelihood

Note that $P(\varepsilon_i \le 0|X_i, U_i) = \tau$, then invoking model (1.1), it is easy to prove that

$$E\{\tau - I(Y_i - X_i^T\beta - \theta(U_i) \le 0)|X_i, U_i\} = 0, \tag{2.1}$$

where $I(\cdot)$ is the indicator function. To present the penalized empirical likelihood based variable selection method, we introduce a penalized auxiliary random vector as follows

$$\eta_i(\beta) = \delta_i X_i[\tau - I(Y_i - X_i^T\beta - \theta(U_i) \le 0)] - b_\lambda(\beta), \tag{2.2}$$

where $b_\lambda(\beta) = (p'_{\lambda_{11}}(|\beta_1|)\text{sgn}(\beta_1), \ldots, p'_{\lambda_{1p}}(|\beta_p|)\text{sgn}(\beta_p))^{\mathrm{T}}$, $\text{sgn}(w)$ means the sign function for w, and $p'_\lambda(w)$ is a penalty function, which can govern the sparsity of model by taking suitable tuning parameters $\lambda_{11}, \cdots, \lambda_{1p}$. The notation $\delta_i X_i$ means the scalar multiplication of matrix, that is, $\delta_i X_i = (\delta_i X_{i1}, \cdots, \delta_i X_{ip})^T$, where X_{ij} is the jth component of X_i.

Various penalty functions have been used in the variable selection literature such as the L_q penalty proposed by Frank and Friedman [20], the Lasso penalty proposed by Tibshirani [21], the SCAD penalty proposed by Fan and Li [22] and the MCP penalty proposed by Zhang [23]. In addition, it has been shown that the SCAD penalty has many perfect properties, such as the consistence and sparsity of SCAD penalty based estimation, in many situations. Then, in this paper, we suggest the penalty function is taken as the SCAD penalty function, which is defined as follows

$$p'_\lambda(w) = \lambda\{I(w \le \lambda) + \frac{(a\lambda - w)_+}{(a-1)\lambda} I(w > \lambda)\},$$

for some $a > 2$, $w > 0$ and $p_\lambda(0) = 0$, where λ is a tuning parameter which can govern sparsity of the estimation of model, and the notation $(z)_+$ means the positive part of z. In addition, Fan and Li [22] suggested using $a = 3.7$ for the SCAD penalty function.

However, note that (2.2) contains the nonparametric component $\theta(u)$, then $\eta_i(\beta)$ cannot be used directly to construct the penalized empirical likelihood ratio for selecting the important variables of parametric components. To overcome this inconvenience, we give an estimator of $\theta(u)$ by using basis functions approximation method. More specifically, let $W(u) = (B_1(u), \cdots, B_L(u))^T$ be B-spline basis functions with the order of M, where $L = K + M + 1$, and K denotes the number of interior knots. Then, $\theta(u)$ can be approximated by

$$\theta(u) \approx W(u)^T\gamma, \tag{2.3}$$

where $\gamma = (\gamma_1, \cdots, \gamma_L)^T$ is a vector of basis function coefficients. Hence, invoking (2.1) and (2.3), an estimating equation for β and $\theta(\cdot)$ can be given as follows

$$\sum_{i=1}^{n} \delta_i (X_i^T, W_i^T)^T [\tau - I(Y_i - X_i^T\beta - W_i^T\gamma \leq 0)] = 0, \tag{2.4}$$

where $W_i = W(U_i) = (B_1(U_i), \cdots, B_L(U_i))^T$.

2.1. Remark. When we estimate β and γ by using (2.4), the number of interior knots K should be chosen. Here we suggest estimating K by minimizing the following cross-validation score

$$CV(K) = \sum_{i=1}^{n} \delta_i \rho_\tau (Y_i - X_i^T\hat{\beta}^{[i]} - W_i^T\hat{\gamma}^{[i]}), \tag{2.5}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile loss function, and $\hat{\beta}^{[i]}$ and $\hat{\gamma}^{[i]}$ are estimators for β and γ, respectively, based on (2.4) after deleting the ith subject.

Let $\hat{\gamma}$ be the solution of (2.4), then the estimator of $\theta(u)$ can be given by

$$\hat{\theta}(u) = W(u)^T\hat{\gamma}. \tag{2.6}$$

Hence, a modified auxiliary random vector can be given by

$$\hat{\eta}_i(\beta) = \delta_i X_i[\tau - I(Y_i - X_i^T\beta - \hat{\theta}(U_i) \leq 0)] - b_\lambda(\beta). \tag{2.7}$$

Then, a profile penalized empirical log-likelihood ratio function for β can be defined as

$$R(\beta) = -2\max\left\{ \sum_{i=1}^{n} \log(np_i) \,\middle|\, p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i\hat{\eta}_i(\beta) = 0 \right\}. \tag{2.8}$$

Next, we represent some asymptotic properties of the maximum empirical likelihood estimator $\hat{\beta}$, which is the solution by maximizing $\{-R(\beta)\}$. We first give some notations. Let β_0 be the true value of β with $\beta_{k0} \neq 0$ for $k \leq d$ and $\beta_{k0} = 0$ for $k > d$. The following theorem states the existence of a consistent solution $\hat{\beta}$.

2.2. Theorem. *Suppose that conditions $C1-C7$ in Section 5 hold. Then, the maximum empirical likelihood estimator $\hat{\beta}$ is consistent, that is,*

$$\hat{\beta} = \beta_0 + O_p(n^{-1/2}).$$

Furthermore, under some conditions, the following theorem shows that such consistent estimator must possess the sparsity property.

2.3. Theorem. *Suppose that conditions $C1-C7$ in Section 5 hold. Then, with probability tending to 1, $\hat{\beta}$ must satisfy*

$$\hat{\beta}_k = 0, \qquad k = d+1, \cdots, p.$$

2.4. Remark. Theorem 2.3 indicates that, with probability tending to 1, some components of the maximum empirical likelihood estimator $\hat{\beta}$ are set to be zeros. Then, the corresponding covariates will be removed from the final model. Hence, the proposed penalized empirical likelihood procedure can be used for variable selection.

To implement this variable selection method, the tuning parameters a and $\lambda_{11}, \cdots, \lambda_{1p}$ in the penalty functions should be chosen. Similar to [22], we take $a = 3.7$. In addition, to reduce the computation task, we define adaptive tuning parameters for $\lambda_{11}, \cdots, \lambda_{1p}$, as

$$\lambda_{1k} = \frac{\lambda}{\hat{\beta}_k^{(0)}}, \qquad k = 1, \cdots, p, \tag{2.9}$$

where $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, \cdots, \hat{\beta}_p^{(0)})^T$ is the naive estimator of β based on (2.4), and λ is obtained by minimizing the following BIC criteria

$$\mathrm{BIC}(\lambda) = \log(\mathrm{RSS}(\lambda)/n) + d(\lambda)\log(n)/n,$$

where $\mathrm{RRS}(\lambda) = \sum_{i=1}^n \rho_\tau(Y_i - X_i^T\hat{\beta} - \hat{\theta}(U_i))$, and $d(\lambda) = \mathrm{tr}\{\mathbf{X}[\mathbf{X}^T\mathbf{X} + n\Sigma_\lambda(\hat{\beta})]^{-1}\mathbf{X}^T\}$ is the effective number of parameters, where $\mathbf{X} = (X_1, \ldots, X_n)^T$ and $\Sigma_\lambda(\hat{\beta}) = \mathrm{diag}\{p'_{\lambda_{11}}(|\hat{\beta}_1|)/|\hat{\beta}_1|, \ldots, p'_{\lambda_{1p}}(|\hat{\beta}_p|)/|\hat{\beta}_p|\}$.

3. Variable selection via double penalized empirical likelihood

From the definition of $R(\beta)$ in Section 2, we can see that a consistent estimator $\hat{\theta}(u)$ should be given. However, if the missing rate of data is very large, then we cannot give a workable consistent estimator of $\theta(u)$, which will influence the variable selection for the parametric components. In addition, note that if the number of interior knots K is large, the vector of basis function coefficients γ may be also sparse. Hence, this prompts us to give a penalty term on γ and β simultaneously, and propose a double penalized empirical likelihood based variable selection procedure.

To present the double penalized empirical likelihood based variable selection method, we introduce a double penalized auxiliary random vector as follows

$$\tilde{\eta}_i(\beta, \gamma) = \delta_i(X_i^T, W_i^T)^T[\tau - I(Y_i - X_i^T\beta - W_i^T\gamma \leq 0)] - b_\lambda(\beta, \gamma), \tag{3.1}$$

where $b_\lambda(\beta, \gamma) = (p'_{\lambda_{11}}(|\beta_1|)\mathrm{sgn}(\beta_1), \ldots, p'_{\lambda_{1p}}(|\beta_p|)\mathrm{sgn}(\beta_p), p'_{\lambda_{21}}(|\gamma_1|)\mathrm{sgn}(\gamma_1), \ldots,$ $p'_{\lambda_{2L}}(|\gamma_L|)\mathrm{sgn}(\gamma_L))^{\mathrm{T}}$ is the penalty vector for parametric components and basis function coefficients, and $\lambda_{11}, \cdots, \lambda_{1p}, \lambda_{21}, \cdots, \lambda_{2L}$ are tuning parameters. Then, a double penalized empirical log-likelihood ratio function can be defined as

$$\tilde{R}(\beta, \gamma) = -2\max\left\{\sum_{i=1}^n \log(np_i) \middle| p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i\tilde{\eta}_i(\beta, \gamma) = 0\right\}. \tag{3.2}$$

Let $\tilde{\beta}$ and $\tilde{\gamma}$ be the solution by maximizing $\{-\tilde{R}(\beta, \gamma)\}$. Then, $\tilde{\beta}$ is the penalized maximum empirical likelihood estimator of β, and the penalized maximum empirical likelihood estimator of $\theta(u)$ can be given by $\tilde{\theta}(u) = W(u)^T\tilde{\gamma}$. Next, we study some asymptotic properties of the resulting penalized least squares estimators $\tilde{\beta}$ and $\tilde{\gamma}$. Similar to Section 2, we write the true regression coefficients of β and γ as β_0 and γ_0, respectively. Without loss of generality, we also assume that $\beta_{k0} = 0,\ k = d+1, \cdots, p$, and $\beta_{k0},\ k = 1, \cdots, d$ are all nonzero components of β_0. Furthermore, we assume that $\gamma_{l0} = 0,\ l = s+1, \cdots, L$, and $\gamma_{l0},\ l = 1, \cdots, s$ are all nonzero components of γ_0. The following theorem shows that such penalized maximum empirical likelihood estimator $\tilde{\beta}$ also possesses the sparsity property, then we can use this double penalized empirical likelihood method for variable selection.

3.1. Theorem. *Suppose that conditions $C1-C7$ in Section 5 hold. Then, with probability tending to 1, $\tilde{\beta}$ satisfies*

$$\tilde{\beta}_k = 0, \qquad k = d+1, \cdots, p.$$

In addition, the following theorem shows that the penalized maximum empirical likelihood estimator $\tilde{\theta}(u)$ is consistent, and achieves the optimal nonparametric convergence rate (see Schumaker [24]).

3.2. Theorem. *Suppose that conditions $C1 - C7$ in Section 5 hold. Then, the double penalized empirical likelihood based nonparametric estimator $\tilde{\theta}(u)$ satisfies*

$$\|\tilde{\theta}(u) - \theta_0(u)\| = O_p(n^{\frac{-r}{2r+1}}).$$

In addition, similar to Section 2, we suggest that the tuning parameter a is also taken as 3.7, and define the adaptive tuning parameters for λ_{1k} and λ_{2l} as

$$\lambda_{1k} = \frac{\lambda}{\hat{\beta}_k^{(0)}}, \quad \lambda_{2l} = \frac{\lambda}{\hat{\gamma}_l^{(0)}}, \quad k = 1, \cdots, p, \quad l = 1, \cdots, L, \tag{3.3}$$

where $\hat{\beta}^{(0)} = (\hat{\beta}_1^{(0)}, \cdots, \hat{\beta}_p^{(0)})^T$ and $\hat{\gamma}^{(0)} = (\hat{\gamma}_1^{(0)}, \cdots, \hat{\gamma}_L^{(0)})^T$ are naive estimators of β and γ, respectively, obtained by (2.4). In addition, the number of interior knots K and the tuning parameter λ used in (3.3) are obtained by minimizing the following BIC criteria

$$\mathrm{BIC}(\lambda, K) = \log(\mathrm{RSS}(\lambda, K)/n) + d(\lambda, K)\log(n)/n,$$

where $\mathrm{RRS}(\lambda, K) = \sum_{i=1}^n \rho_\tau(Y_i - X_i^T\tilde{\beta} - W_i^T\tilde{\gamma})$, and $d(\lambda, K)$ is the effective number of parameters.

4. Numerical results

In this section, we assess the finite sample performance of the procedure by presenting several simulation experiments, and consider a real data application for further illustration.

4.1. Simulation studies. In this section, we conduct some simulations to study the finite sample performance of the proposed estimation method. We first evaluate the finite sample performance of the proposed quantile regression based profile penalized empirical likelihood (Q-PEL) variable selection procedure in terms of model complexity and model selection accuracy comparing with the mean regression based penalized empirical likelihood (M-PEL) variable selection procedure which proposed by [14]. In this simulation, data are generated from the following model

$$Y = X^T\beta + \theta(U) + \varepsilon, \tag{4.1}$$

where $U \sim U(0,1)$, X follows the $10-$dimensional normal distribution with zero mean and covariance between the sth and tth elements being $\rho^{|s-t|}$ with $\rho = 0.5$. Furthermore, the nonparametric component is taken $\theta(u) = 0.8u(1-u)$, and the parametric component is taken $\beta = (3.5, 2.5, 1.5, 0.5, 0, 0, 0, 0, 0, 0)^T$. The response Y is generated according to the model, and the model error ε is generated according to $\varepsilon = e - b_\tau$, where b_τ is the τth quantile of e. In the following simulation, the quantile τ is taken as 0.25, 0.5 and 0.75, respectively, and e follows the normal distribution $N(0, 0.5)$ (symmetrical error distribution) and chi-square distribution $\chi^2(1)$ (unsymmetrical error distribution), respectively. In addition, we take the selection probability as $P(\delta = 1|X = x, U = u) = 0.8 + 0.2(u - 0.5)$, and the corresponding missing rate is approximately 0.2. Throughout our simulation, we use the cubic B-splines for basis functions approximation, and the number of interior knots K is obtained by (2.5). In addition, we generate $n = 200$, 400 and 600 respectively, and repeat 1000 simulation runs for each case.

Table 1. Variable selection results for the parametric component by different variable selection methods when $\tau = 0.25$.

		Q-PEL				M-PEL			
Err. Dist.	n	C	I	FSR	GMAD	C	I	FSR	GMAD
Normal	200	5.325	0.001	0.057	0.235	5.336	0.001	0.051	0.203
	400	5.787	0	0.011	0.134	5.782	0	0.014	0.135
	600	5.925	0	0.007	0.112	5.919	0	0.009	0.119
Chi-square	200	5.346	0	0.052	0.209	3.212	0	0.393	0.567
	400	5.785	0	0.017	0.134	3.489	0	0.335	0.528
	600	5.918	0	0.005	0.122	3.619	0	0.292	0.502

The performance of estimator $\hat{\beta}$ will be assessed by using the generalized mean absolute deviation (GMAD), which is defined as follows

$$\text{GMAD} = \frac{1}{n}\sum_{i=1}^{n}\left|X_i^T\hat{\beta} - X_i^T\beta_0\right|.$$

In addition, the performance of model complexity will be assessed by using the average false selection rate (FSR), which is defined as FSR = IN/TN, where "IN" is the average number of the true zero coefficients incorrectly set to nonzero, and "TN" is the average total number set to nonzero. In fact, FSR represents the proportion of falsely selected unimportant variables among the total variables selected in the variable selection procedure.

The simulation results for the average number of zero coefficients, with 1000 simulation runs, are reported in Tables 1-3, where Table 1 presents the results for the case of $\tau = 0.25$, Table 2 presents the results for the case of $\tau = 0.5$, and Table 3 presents the results for the case of $\tau = 0.75$. In Tables 1-3, the column labeled "C" gives the average number of coefficients, of the six true zeros, correctly set to zero, and the column labeled "I" gives the average number of the four true nonzeros incorrectly set to zero. In addition, Tables 1-3 also present the average false selection rate (FSR) and the average generalized mean absolute deviation (GMAD) based on 1000 simulation runs.

From Tables 1-3, we can make the following observations:

(i) For any given quantile τ, the performance of the Q-PEL method becomes more and more better in terms of model error and model complexity as n increases. In addition, for given n, the performance of Q-PEL does not depend sensitively on the error distributions, which implies that the proposed Q-PEL variable selection method is robust.

(ii) For any given quantile τ, when the error distribution is a symmetrical distribution (normal distribution), the results based on Q-PEL are similar to that based on M-PEL. However, for the unsymmetrical error distribution (chi-square distribution), the Q-PEL method outperforms the M-PEL method. The latter method is less discriminative, and cannot eliminate some unimportant variables. This is mainly because the mean of chi-square distribution is not zero, which may affect the variable selection of M-PEL method.

(iii) For the given error distribution and sample size, the results based on the Q-PEL method are very similar for all considered quantiles. In addition, the estimation procedure of M-PEL is independent of all quantiles.

Table 2. Variable selection results for the parametric component by different variable selection methods when $\tau = 0.5$.

		Q-PEL				M-PEL			
Err. Dist.	n	C	I	FSR	GMAD	C	I	FSR	GMAD
Normal	200	5.331	0.002	0.054	0.205	5.336	0.001	0.051	0.203
	400	5.781	0	0.015	0.138	5.782	0	0.014	0.135
	600	5.929	0	0.008	0.110	5.919	0	0.009	0.119
Chi-square	200	5.344	0.001	0.053	0.204	3.212	0	0.393	0.567
	400	5.786	0	0.013	0.131	3.489	0	0.335	0.528
	600	5.928	0	0.008	0.112	3.619	0	0.292	0.502

Table 3. Variable selection results for the parametric component by different variable selection methods when $\tau = 0.75$.

		Q-PEL				M-PEL			
Err. Dist.	n	C	I	FSR	GMAD	C	I	FSR	GMAD
Normal	200	5.317	0.002	0.052	0.207	5.336	0.001	0.051	0.203
	400	5.782	0	0.014	0.136	5.782	0	0.014	0.135
	600	5.914	0	0.008	0.118	5.919	0	0.009	0.119
Chi-square	200	5.345	0.001	0.052	0.202	3.212	0	0.393	0.567
	400	5.784	0	0.014	0.135	3.489	0	0.335	0.528
	600	5.932	0	0.007	0.110	3.619	0	0.292	0.502

Next, we compare the performance of the double penalized empirical likelihood (Q-DEL) variable selection method, proposed by (3.2), with the profile penalized empirical likelihood (Q-PEL) variable selection method. In this simulation, we also generate data from model (4.1), and use the cubic B-splines for nonparametric approximation. The number of interior knots is fixed at $K = K_0$, $2K_0$, $3K_0$ and $4K_0$ respectively, where K_0 is the number of interior knots which is obtained by (3.3). In addition, the tuning parameter λ is obtained by (3.3) for given K. Here, we only present the results when $\tau = 0.5$ and the error distribution follows the chi-square distribution with one degree. Since the results for other cases are similar, then are not shown. Based on 1000 repeated simulation runs, the simulation results are reported in Table 4. From Table 4, we can see that

(i) The performances of the Q-DEL method are similar to the performances of the Q-PEL method in terms of model error and model complexity, when the number on interior knots K is chosen correctly ($K = K_0$).

(ii) If K is misspecified, the performance of Q-DEL is better than that of Q-PEL. The latter cannot eliminate some unimportant variables and gives larger models errors. This implies that the Q-PEL procedure cannot give an effective estimators for the nonparametric function when the number of interior knots is misspecified.

(iii) The performance of the Q-DEL method becomes more and more better in terms of model error and model complexity as n increases. Furthermore, for given n, the Q-DEL variable selection procedure performs similar for all cases of the

Table 4. Variable selection results for the parametric component by different variable selection methods.

n	K	Q-PEL C	Q-PEL I	Q-PEL FSR	Q-PEL GMAD	Q-DEL C	Q-DEL I	Q-DEL FSR	Q-DEL GMAD
200	K_0	5.344	0.001	0.053	0.204	5.401	0.001	0.045	0.197
	$2K_0$	5.053	0.001	0.084	0.223	5.398	0	0.051	0.201
	$3K_0$	4.564	0.002	0.102	0.268	5.395	0.001	0.051	0.201
	$4K_0$	4.239	0	0.084	0.296	5.385	0	0.052	0.202
400	K_0	5.786	0	0.013	0.131	5.783	0	0.013	0.134
	$2K_0$	5.443	0	0.057	0.162	5.782	0	0.013	0.134
	$3K_0$	5.109	0	0.081	0.217	5.778	0	0.014	0.136
	$4K_0$	4.787	0	0.093	0.253	5.774	0	0.016	0.139
600	K_0	5.928	0	0.008	0.112	5.929	0	0.008	0.109
	$2K_0$	5.642	0	0.012	0.141	5.927	0	0.008	0.109
	$3K_0$	5.206	0	0.061	0.189	5.927	0	0.008	0.111
	$4K_0$	4.918	0	0.076	0.231	5.925	0	0.007	0.112

number of interior knots, which implies that our penalty scheme on the nonparametric function is workable.

4.2. Application to a real data example. Now we illustrate the newly proposed procedure through an analysis of dataset from the Multi-Center AIDS Cohort study. The dataset contains the human immunodeficiency virus (HIV) status of 283 homosexual men who were infected with HIV during a follow-up period between 1984 and 1991. The observed variables in this data are cigarette smoking status, age at HIV infection, pre-infection CD4 percentage and post-infection CD4 percentage. More details of the related design, methods and medical implications for the Multi-Center AIDS Cohort study have been described by Kaslow et al. [25]. The objective of the study is to describe the trend of the mean CD4 percentage depletion over time and evaluate the effects of smoking, the pre-HIV infection CD4 percentage, and age at HIV infection on the mean CD4 percentage after infection. This dataset has been used by many authors to illustrate varying coefficient models (see [26, 27]), varying coefficient partially linear models (see [28, 29]) and partially linear models (see [30, 31]).

We take Y to be the individual's CD4 percentage, X_1 to be the centered preCD4 percentage, $X_2 = X_1^2$ to be the quadratic effect of the centered preCD4 percentage, X_3 to be the centered age at HIV infection, and X_4 to be quadratic effect of the centered age. Note that Huang et al. [26] indicates that, at significance level 0.05, the smoking status has not a significant impact on the mean CD4 percentage, then the possible effects of other available covariates are omitted. We consider the following partially linear model

$$Y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + X_4\beta_4 + \theta(t), \tag{4.2}$$

where t means the years over the HIV infection, and $\theta(t)$ represents the mean CD4 percentage t years after the infection.

In order to illustrate our approach, similar to Xue [32], we artificially make 20% Y values missing, and neglect the dependency of the repeated measurements for each subject. We first apply the proposed Q-PEL variable selection procedure to model (4.2). Here we take the quantile τ as $0.1, 0.2, \cdots, 0.9$, respectively. The variable selection results

Table 5. Application to AIDS data. The regularized estimators for parametric components based on the Q-PEL method with different quantiles.

	τ								
Variable	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
X_1	0.241	0.303	0.345	0.389	0.479	0.443	0.484	0.467	0.477
X_2	0	0	0	0	0	0	0	0	0
X_3	0	0	0	0	0	0	0	0	0
X_4	0	0	0	0	0	0	0	0	0

Table 6. Application to AIDS data. The regularized estimators for parametric components based on the Q-DEL method with different quantiles.

	τ								
Variable	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
X_1	0.251	0.308	0.354	0.431	0.443	0.442	0.493	0.479	0.465
X_2	0	0	0	0	0	0	0	0	0
X_3	0	0	0	0	0	0	0	0	0
X_4	0	0	0	0	0	0	0	0	0

for the parametric components are shown in Table 5, where nonzero estimators represent the selected important variables in final model, and zeros represent the eliminated unimportant variables from final model.

From Table 5, we can see that the preCD4 percentage has a significant effect on the mean CD4 percentage at all quantiles. The age, and the quadratic effects of age and preCD4 percentage have no significant impact on the mean CD4 percentage at all quantiles, which basically agrees with that was discovered by [29] based on varying coefficient partially linear models. In addition, the curve of estimated $\theta(t)$ is shown in Figure 1(a). Here we only show the results for the case of $\tau = 0.2$, 0.5 and 0.8, respectively. The results for other cases are similar, and then are not shown. From Figure 1(a), we find that the mean CD4 percentage is significantly time-varying for all considered quantile levels. We also can see that the rate of variation is very quickly at the beginning of HIV infection for all considered quantiles, and the rate of variation slows down two years after infection.

Furthermore, we also apply the proposed Q-DEL variable selection method to model (4.2). The regularized estimators for parametric components based on Q-DEL method are shown in Table 6. From Table 6, we can see that the Q-DEL method identifies one nonzero regression coefficient β_1 for parametric components. This indicates that only the preCD4 percentage has significant impact on the mean CD4 percentage, which agrees with that was discovered based on the Q-PEL method. In addition, the regularized estimator of $\theta(t)$ based on the Q-DEL method is shown in Figure 1(b). From Figure 1(b), we can see that the regularized estimator of $\theta(t)$ based on the Q-DEL method is similar to that based on the Q-PEL method.

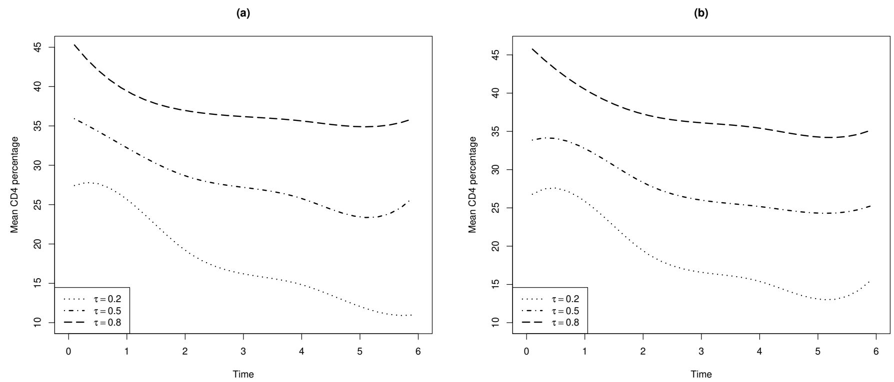

Figure 1. Application to AIDS data. The regularized estimators for the mean CD4 percentage $\theta(t)$, based on the Q-PEL (Fig(a)) and Q-DEL (Fig(b)), for the cases $\tau = 0.2$ (dotted curve), $\tau = 0.5$ (dot-dashed curve) and $\tau = 0.8$ (dashed curve).

5. Proofs of Theorems

In this section, we present the technical proofs of Theorems 2.2, 2.3, 3.1 and 3.2. For convenience and simplicity, let c denote a positive constant which may be different value at each appearance throughout this paper. To prove these asymptotic properties, the following technical conditions are needed.

C1. The nonparametric function $\theta(u)$ is rth continuously differentiable for $u \in (0,1)$, where $r > 1/2$.

C2. The number of interior knots satisfies $K = O(n^{1/(2r+1)})$. In addition, let $c_1, \cdots, c_K$ be the interior knots of $(0,1)$, and denote $c_0 = 0$, $c_{K+1} = 1$ and $h_k = c_k - c_{k-1}$, then there exists a constant c such that

$$\frac{\max h_k}{\min h_k} \le c, \qquad \max\{|h_{k+1} - h_k|\} = o(K^{-1}).$$

C3. Let $f(\cdot|X,U)$ be the conditional density function of ε given X and U, then $f(\cdot|X,U)$ has a continuous and uniformly bounded derivative.

C4. Let $\pi(x,u) = E(\delta_i|X_i = x, U_i = u)$, then $\pi(x,u)$ has continuous bounded second partial derivatives. Furthermore, we assume that $\pi(x,u) > 0$ for all x and u.

C5. Let

$$a_n = \max_{k,l}\{p'_{\lambda_{1k}}(|\beta_{k0}|), p'_{\lambda_{2l}}(|\gamma_{l0}|) : \beta_{k0} \neq 0, \gamma_{l0} \neq 0\},$$

and

$$b_n = \max_{k,l}\{|p''_{\lambda_{1k}}(|\beta_{k0}|)|, |p''_{\lambda_{2l}}(|\gamma_{l0}|)| : \beta_{k0} \neq 0, \gamma_{l0} \neq 0\},$$

then we have $n^{r/(2r+1)} a_n \to 0$ and $b_n \to 0$, as $n \to \infty$.

C6. The penalty function satisfies $\liminf_{n\to\infty} \liminf_{w\to 0+} \lambda^{-1} p'_\lambda(|w|) > 0$. Let $\lambda_{min} = \min\{\lambda_{1k}, \lambda_{2l} : k = 1, \cdots, p, l = 1, \cdots, L\}$ and $\lambda_{max} = \max\{\lambda_{1k}, \lambda_{2l} : k = 1, \cdots, p, l = 1, \cdots, L\}$. Then λ_{min} and λ_{max} satisfy $\lambda_{max} \to 0$ and $n^{r/(2r+1)}\lambda_{min} \to \infty$ as $n \to \infty$.

C7. The covariate X is a centered random vector, and is bounded in probability. In addition, we assume that the matrix $E\{\pi(X,U)XX^T\}$ is a nonsingular and finite matrix.

These conditions are commonly adopted in the nonparametric literature and variable selection methodology. Condition C1 is a smoothness condition for $\theta(u)$, which determines the rate of convergence of the spline estimator $\hat{\theta}(u) = W(u)^T\hat{\gamma}$. Condition C2 implies that $c_0, \cdots, c_{K+1}$ is a C_0-quasi-uniform sequence of partitions on $[0,1]$ (see Schumaker [24]). Conditions C3, C4 and C7 are some regularity conditions used in our estimation procedure, which are similar to those used in Lv and Li [12], Ren and Zhang [14] and Xue and Zhu [25]. Conditions C5 and C6 are assumptions for the penalty function, which are similar to those used in Ren and Zhang [14], Zhao [17] and Fan and Li [22]. The proofs of the main results rely on the following lemma.

5.1. Lemma. *Suppose that the conditions C1-C7 hold. Then we have*

$$\|\hat{\theta}(u) - \theta(u)\| = O_p\left(n^{\frac{-r}{2r+1}}\right),$$

where $\hat{\theta}(u)$ is defined in (2.6), and r is defined in condition C1.

Proof. Let $\kappa = n^{-r/(2r+1)}$, $\breve{\beta} = \beta + \kappa M_1$, $\breve{\gamma} = \gamma + \kappa M_2$ and $M = (M_1^T,\ M_2^T)^T$, we first show that, for any given $\varepsilon > 0$, there exists a large constant c such that

$$P\left\{\inf_{\|M\|=C}(\beta^T - \breve{\beta}^T, \gamma^T - \breve{\gamma}^T)\sum_{i=1}^n \tilde{\eta}_i(\breve{\beta}, \breve{\gamma}) > 0\right\} \geq 1 - \varepsilon, \tag{5.1}$$

where

$$\tilde{\eta}_i(\breve{\beta}, \breve{\gamma}) = \delta_i(X_i^T, W_i^T)^T[\tau - I(Y_i - X_i^T\breve{\beta} - W_i^T\breve{\gamma} \leq 0)].$$

invoking the definition of $\tilde{\eta}_i(\breve{\beta}, \breve{\gamma})$, a simple calculation yields

$$\begin{aligned}
\tilde{\eta}_i(\breve{\beta}, \breve{\gamma}) &= \delta_i(X_i^T, W_i^T)^T[\tau - I(Y_i - X_i^T\breve{\beta} - W_i^T\breve{\gamma} \leq 0)] \\
&= \delta_i(X_i^T, W_i^T)^T[\tau - I(X_i^T\beta + \theta(U_i) + \varepsilon_i - X_i^T\breve{\beta} - W_i^T\breve{\gamma} \leq 0)] \\
&= \delta_i(X_i^T, W_i^T)^T[\tau - I(\varepsilon_i + X_i^T(\beta - \breve{\beta}) + W_i^T(\gamma - \breve{\gamma}) + R(U_i) \leq 0)] \\
&= \delta_i f(0|X_i, U_i)(X_i^T, W_i^T)^T[X_i^T(\beta - \breve{\beta}) + W_i^T(\gamma - \breve{\gamma}) + R(U_i)] \\
&\quad + O_p(\|\beta - \breve{\beta}\|^2) + O_p(\|\gamma - \breve{\gamma}\|^2).
\end{aligned} \tag{5.2}$$

Let $\Delta(\breve{\beta}, \breve{\gamma}) = K^{-1}\left(\beta^T - \breve{\beta}^T, \gamma^T - \breve{\gamma}^T\right)\sum_{i=1}^n \tilde{\eta}_i(\breve{\beta}, \breve{\gamma})$, then a simple calculation yields

$$\begin{aligned}
\Delta(\breve{\beta}, \breve{\gamma}) &= \frac{-\kappa}{K}(M_1^T, M_2^T)\sum_{i=1}^n \delta_i f(0|X_i, U_i)(X_i^T, W_i^T)^T[X_i^T(-\kappa M_1) + W_i^T(-\kappa M_2)] \\
&\quad + \frac{-\kappa}{K}(M_1^T, M_2^T)\sum_{i=1}^n \delta_i f(0|X_i, U_i)(X_i^T, W_i^T)^T R(U_i) + O_p(nK^{-1}\kappa^2) \\
&= \frac{\kappa^2}{K}\sum_{i=1}^n \delta_i f(0|X_i, U_i)(X_i^T M_1 + W_i^T M_2)^2 \\
&\quad + \frac{-\kappa}{K}\sum_{i=1}^n \delta_i f(0|X_i, U_i)(X_i^T M_1 + W_i^T M_2)R(U_i) + O_p(1) \\
&\equiv I_1 + I_2,
\end{aligned} \tag{5.3}$$

where $R(U_i) = \theta(U_i) - W_i^T\gamma$. From conditions C1, C4 and Corollary 6.21 in [24], we get that $\|R(U_i)\| = O(K^{-r})$. Then, invoking condition C3, a simple calculation yields $I_1 = O_p(\kappa^2 nK^{-1})\|M\|^2 = O_p(\|M\|^2)$, and $I_2 = O_p(\kappa nK^{-1-r})\|M\| = O_p(\|M\|)$. Hence, by choosing a sufficiently large C, I_1 dominates I_2 uniformly in $\|M\| = C$. This implies that for any given $\varepsilon > 0$, if we choose C large enough, then

$$P\left\{\inf_{\|M\|=C} \Delta(\breve{\beta}, \breve{\gamma}) > 0\right\} \geq 1 - \varepsilon. \tag{5.4}$$

Hence (5.1) holds, and this implies, with probability at least $1 - \varepsilon$, that there exists a local minizer $\hat{\beta}$ and $\hat{\gamma}$ such that

$$\|\hat{\beta} - \beta\| = O_p(\tau) = O_p\left(n^{-r/(2r+1)}\right), \quad \|\hat{\gamma} - \gamma\| = O_p(\tau) = O_p\left(n^{-r/(2r+1)}\right). \tag{5.5}$$

In addition, note that

$$\begin{aligned}
\|\hat{\theta}(u) - \theta(u)\|^2 &= \int_0^1 \{\hat{\theta}(u) - \theta(u)\}^2 du \\
&= \int_0^1 \{B^T(u)\hat{\gamma} - B^T(u)\gamma + R(u)\}^2 du \\
&\leq 2\int_0^1 \{B^T(u)\hat{\gamma} - B^T(u)\gamma\}^2 du + 2\int_0^1 R(u)^2 du \\
&= 2(\hat{\gamma} - \gamma)^T H(\hat{\gamma} - \gamma) + 2\int_0^1 R(u)^2 du,
\end{aligned} \tag{5.6}$$

where $H = \int_0^1 B(u)B^T(u)du$. It is easy to show that $\|H\| = O_p(1)$. Then invoking (5.5), a simple calculation yields

$$(\hat{\gamma}_k - \gamma_k)^T H(\hat{\gamma}_k - \gamma_k) = O_p\left\{n^{\frac{-2r}{2r+1}}\right\}. \tag{5.7}$$

In addition, invoking $\|R(U_i)\| = O(K^{-r})$, it is easy to show that

$$\int_0^1 R(u)^2 du = O_p(n^{\frac{-2r}{2r+1}}). \tag{5.8}$$

Invoking (5.6)-(5.8), we complete the proof of this lemma. □

Proof of Theorem 2.2

Proof. With the similar argiments as in Xue and Zhu [31], we have that the solution to maximizing $\{-\hat{R}(\beta)\}$ can be given by the following penalty estimating equation

$$\sum_{i=1}^n \delta_i X_i[\tau - I(Y_i - X_i^T\beta - \hat{\theta}(U_i) \leq 0)] - nb_\lambda(\beta) = 0. \tag{5.9}$$

Let $U(\beta) = \sum_{i=1}^n \delta_i X_i[\tau - I(Y_i - X_i^T\beta - \hat{\theta}(U_i) \leq 0)]$, $U^P(\beta) = U(\beta) - nb_\lambda(\beta)$ and $\beta = \beta_0 + n^{-1/2}M$, where β_0 means the true value of β. Similar to the argument in the proof of Lemma 5.1, we want to show that for any given $\varepsilon > 0$, there exists a large constant c such that $\|M\| = c$ and

$$P\left\{\min_{\|\beta_0 - \beta\| = cn^{-1/2}} (\beta_0 - \beta)^{\mathrm{T}} U^P(\beta) > 0\right\} > 1 - \varepsilon, \tag{5.10}$$

With probability at least $1 - \varepsilon$, (5.10) implies that there exists a local solution to $U^P(\beta) = 0$ in the ball $\{\beta_0 + n^{-1/2}M : \|M\| \leq c\}$. That is, there exists a local solution $\hat{\beta}$ of $U^P(\beta) = 0$ with $\hat{\beta} = \beta_0 + O_p(n^{-1/2})$. Invoking the definition of $\hat{\eta}_i(\beta)$, and some calculations yield

$$\begin{aligned}\hat{\eta}_i(\beta) &= \delta_i X_i[\tau - I(Y_i - X_i^T\beta - \hat{\theta}(U_i) \le 0)] - b_\lambda(\beta)\\ &= \delta_i X_i[\tau - I(X_i^T\beta_0 + \theta_0(U_i) + \varepsilon_i - X_i^T\beta - \hat{\theta}(U_i) \le 0)] - b_\lambda(\beta)\\ &= \delta_i X_i[\tau - I(\varepsilon_i + X_i^T(\beta_0 - \beta) + (\theta_0(U_i) - \hat{\theta}(U_i)) \le 0)] - b_\lambda(\beta)\\ &= \delta_i f(0|X_i, U_i)X_i X_i^T(\beta_0 - \beta) + \delta_i f(0|X_i, U_i)X_i(\theta_0(U_i) - \hat{\theta}(U_i))\\ &\quad -b_\lambda(\beta) + O_p(\|\beta_0 - \beta\|^2). \end{aligned} \tag{5.11}$$

Hence we have

$$\begin{aligned}U^P(\beta) &= \sum_{i=1}^n \delta_i f(0|X_i, U_i)X_i X_i^T(\beta_0 - \beta)\\ &\quad +\sum_{i=1}^n \delta_i f(0|X_i, U_i)X_i(\theta_0(U_i) - \hat{\theta}(U_i)) - nb_\lambda(\beta) + o_p(1).\end{aligned} \tag{5.12}$$

Note that $p'_\lambda(0)\text{sgn}(0) = 0$, then condition C5 implies that $\sqrt{n}b_\lambda(\beta_0) \to 0$. Hence, it is easy to show that

$$-\sqrt{n}b_\lambda(\beta) = \sqrt{n}\{b_\lambda(\beta_0) - b_\lambda(\beta)\} + o_p(1). \tag{5.13}$$

If $\beta_{k0} \neq 0$, then $\text{sgn}(\beta_{k0}) = \text{sgn}(\beta_k)$. Hence,

$$p'_{\lambda_{1k}}(|\beta_{k0}|)\text{sgn}(\beta_{k0}) - p'_{\lambda_{1k}}(|\beta_k|)\text{sgn}(\beta_k) = \{p'_{\lambda_{1k}}(|\beta_{k0}|) - p'_{\lambda_{1k}}(|\beta_k|)\}\text{sgn}(\beta_k). \tag{5.14}$$

If $\beta_{k0} = 0$, the above equation holds naturally. Then, invoking (5.13) and (5.14), a simple calculation yields

$$-\sqrt{n}b_\lambda(\beta) = \sqrt{n}\{b_\lambda(\beta_0) - b_\lambda(\beta)\} + o_p(1) = \sqrt{n}\Lambda_\lambda(\beta^*)(\beta_0 - \beta) + o_p(1), \tag{5.15}$$

where $\Lambda_\lambda(\beta^*) = \text{diag}\{p''_{\lambda_{11}}(|\beta_1^*|)\text{sgn}(\beta_1), \ldots, p''_{\lambda_{1p}}(|\beta_p^*|)\text{sgn}(\beta_p)\}$, and β_k^* lies between β_k and β_{k0}. From (5.12) and (5.15), we can get that

$$\begin{aligned}(\beta_0 - \beta)^{\mathrm{T}}U^P(\beta) &= \sum_{i=1}^n \delta_i f(0|X_i, U_i)(\beta_0 - \beta)^{\mathrm{T}}X_i X_i^T(\beta_0 - \beta)\\ &\quad +\sum_{i=1}^n \delta_i f(0|X_i, U_i)(\beta_0 - \beta)^{\mathrm{T}}X_i(\theta_0(U_i) - \hat{\theta}(U_i))\\ &\quad +n(\beta_0 - \beta)^{\mathrm{T}}\Lambda_\lambda(\beta^*)(\beta_0 - \beta) + o_p(1)\\ &\equiv A_1 + A_2 + A_3 + o_p(1).\end{aligned} \tag{5.16}$$

Note that $E\{\pi(X, U)XX^T\}$ is a nonsingular and finite matrix, then invoking $\beta = \beta_0 + n^{-1/2}M$ and Conditions C3 and C7, it is easy to show that

$$A_1 = O_p(1)\|M\|^2. \tag{5.17}$$

Next we consider A_2. Note that X_i is the centered covariate, then by Lemma A.2 in [33], we have that

$$\max_{1\le s\le n}\left\|\sum_{i=1}^s X_i\right\| = O_p(\sqrt{n}\log n). \tag{5.18}$$

In addition, by Lemma 5.1 we have that

$$\|\hat{\theta}(u) - \theta_0(u)\| = O_p\left(n^{\frac{-r}{2r+1}}\right). \tag{5.19}$$

Then invoking (5.18) and (5.19), and using the Abel inequality, it is easy to show that

$$\begin{aligned} |A_2| &\leq \sum_{i=1}^{n} \left| \delta_i f(0|X_i,U_i)(\beta_0-\beta)^{\mathrm{T}} X_i(\theta_0(U_i)-\hat{\theta}(U_i)) \right| \\ &\leq \|\beta_0-\beta\| \max_{1\leq i\leq n} \|\delta_i f(0|X_i,U_i)[\theta_0(U_i)-\hat{\theta}(U_i)]\| \max_{1\leq s\leq n} \left\| \sum_{i=1}^{s} X_i \right\| \\ &= O_p(n^{-1/2}\|M\| \cdot n^{-r/(2r+1)} \cdot n^{1/2} \cdot \log n) = o_p(1)\|M\|. \end{aligned} \tag{5.20}$$

By condition C5, we have $\max_k p''_{\lambda_{1k}}(|\beta_k^*|) \to 0$. Hence we have

$$|A_3| = n(\beta_0-\beta)^{\mathrm{T}}\Lambda_\lambda(\beta^*)(\beta_0-\beta) = o_p(1)\|M\|^2. \tag{5.21}$$

Note that $\|M\| = c$, if we choose c large enough, A_2 and A_3 can be dominated A_1 uniformly in $\|M\| = c$. In addition, note that the sign of A_1 is positive, then by choosing a sufficiently large c, (5.10) holds. This implies, with probability at least $1-\varepsilon$, that there exists a local solution $\hat{\beta}-\beta_0 = O_p(n^{-1/2})$, which completes the proof of this theorem. □

Proof of Theorem 2.3

Proof. For this theorem, it suffices to show that for any $\varepsilon > 0$, when n is large enough, we have $P\{\hat{\beta}_k \neq 0\} < \varepsilon$, $k = d+1,\ldots,p$. Since $\hat{\beta}_k = O_p(n^{-1/2})$, when n is large enough, there exists some c such that

$$\begin{aligned} P\{\hat{\beta}_k \neq 0\} &= P\{\hat{\beta}_k \neq 0, |\hat{\beta}_k| \geq cn^{-1/2}\} + P\{\hat{\beta}_k \neq 0, |\hat{\beta}_k| < cn^{-1/2}\} \\ &< \varepsilon/2 + P\{\hat{\beta}_k \neq 0, |\hat{\beta}_k| < cn^{-1/2}\}. \end{aligned} \tag{5.22}$$

Using the kth component of (5.12), and note that $U^P(\hat{\beta}) = 0$, we can obtain that

$$\begin{aligned} \sqrt{n}p'_{\lambda_{1k}}(|\hat{\beta}_k|)\mathrm{sgn}(\hat{\beta}_k) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i f(0|X_i,U_i) X_{ik} X_i^T(\beta_0-\hat{\beta}) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i f(0|X_i,U_i) X_{ik}(\theta_0(U_i)-\hat{\theta}(U_i)) + o_p(1). \end{aligned} \tag{5.23}$$

The first two terms on the right-hand side are of order $O_p(1)$. Hence, for large n, there exists some c such that

$$P(\sqrt{n}p'_{\lambda_{1k}}(|\hat{\beta}_k|) > c) < \varepsilon/2. \tag{5.24}$$

In addition, by condition C6, we have that

$$\inf_{|\hat{\beta}_k|\leq cn^{-1/2}} \sqrt{n}p'_{\lambda_{1k}}(|\hat{\beta}_k|) = \sqrt{n}\lambda_{1k} \inf_{|\hat{\beta}_k|\leq cn^{-1/2}} \lambda_{1k}^{-1} p'_{\lambda_{1k}}(|\hat{\beta}_k|) \to \infty. \tag{5.25}$$

That is, $\hat{\beta}_k \neq 0$ and $|\hat{\beta}_k| < cn^{-1/2}$ imply that $\sqrt{n}p'_{\lambda_{1k}}(|\hat{\beta}_k|) > c$ for large n. Then, invoking (5.22), we have that

$$P\{\hat{\beta}_k \neq 0\} < \varepsilon/2 + P(\sqrt{n}p'_{\lambda_{1k}}(|\hat{\beta}_k|) > c) < \varepsilon.$$

This completes the proof of this theorem. □

5.2. Lemma. *Suppose that the conditions C1-C7 hold. Then we have*

$$\tilde{\beta} = \beta_0 + O_p\left(n^{\frac{-r}{2r+1}}\right), \qquad \tilde{\gamma} = \gamma_0 + O_p\left(n^{\frac{-r}{2r+1}}\right),$$

where $\tilde{\beta}$ and $\tilde{\gamma}$ are obtained by maximizing $\{-\tilde{R}(\beta,\gamma)\}$ which is defined in (3.2).

Proof. As in the proof of Theorem 2.2, the solution to maximizing $\{-\tilde{R}(\beta,\gamma)\}$ can be given by the following penalty estimating equation

$$\sum_{i=1}^{n}\tilde{\eta}_i(\beta,\gamma)=0.$$

The following proof is very similar to the proof of Lemma 5.1, but some details need slight modification. Here, we also use some similar denotes which are defined in the proof of Lemma 5.1. Let $\kappa=n^{-r/(2r+1)}$, $\beta=\beta_0+\kappa M_1$, $\gamma=\gamma_0+\kappa M_2$ and $M=(M_1^T,\ M_2^T)^T$, we first show that, for any given $\varepsilon>0$, there exists a large constant c such that

$$P\left\{\inf_{\|M\|=c}(\beta_0^T-\beta^T,\gamma_0^T-\gamma^T)\sum_{i=1}^{n}\tilde{\eta}_i(\beta,\gamma)>0\right\}\geq 1-\varepsilon. \tag{5.26}$$

Using the same argument as in the proof of (5.15), we have that

$$\begin{aligned}-\sqrt{n}b_\lambda(\beta,\gamma) &= \sqrt{n}\{b_\lambda(\beta_0,\gamma_0)-b_\lambda(\beta,\gamma)\}+o_p(1)\\ &= \sqrt{n}\Lambda_\lambda(\beta^*,\gamma^*)((\beta_0-\beta)^T,(\gamma_0-\gamma)^T)^T+o_p(1),\end{aligned} \tag{5.27}$$

where $\Lambda_\lambda(\beta^*,\gamma^*)=\mathrm{diag}\{p''_{\lambda_{11}}(|\beta_1^*|)\mathrm{sgn}(\beta_1),\ldots,p''_{\lambda_{1p}}(|\beta_p^*|,)\mathrm{sgn}(\beta_p),p''_{\lambda_{21}}(|\gamma_1^*|)\mathrm{sgn}(\gamma_1),\ldots,p''_{\lambda_{2L}}(|\gamma_L^*|,)\mathrm{sgn}(\gamma_L)\}$, β_k^* lies between β_k and β_{k0}, and γ_l^* lies between γ_l and γ_{l0}.

Invoking (5.27) and the definition of $\tilde{\eta}_i(\beta,\gamma)$, a simple calculation yields

$$\begin{aligned}\sum_{i=1}^{n}\tilde{\eta}_i(\beta,\gamma) &= \sum_{i=1}^{n}\delta_i(X_i^T,W_i^T)^T[\tau-I(Y_i-X_i^T\beta-W_i^T\gamma\leq 0)]-nb_\lambda(\beta,\gamma)\\ &= \sum_{i=1}^{n}\delta_i f(0|X_i,U_i)(X_i^T,W_i^T)^T[X_i^T(\beta_0-\beta)+W_i^T(\gamma_0-\gamma)+R(U_i)]\\ &\quad +n\Lambda_\lambda(\beta^*,\gamma^*)(\beta_0-\beta,\gamma_0-\gamma)+O_p(n\kappa^2).\end{aligned} \tag{5.28}$$

Let $\Delta(\beta,\gamma)=K^{-1}\left(\beta_0^T-\beta^T,\gamma_0^T-\gamma^T\right)\sum_{i=1}^{n}\tilde{\eta}_i(\beta,\gamma)$, then we can obtain that

$$\begin{aligned}\Delta(\beta,\gamma) &= \frac{-\kappa}{K}(M_1^T,M_2^T)\sum_{i=1}^{n}\delta_i f(0|X_i,U_i)(X_i^T,W_i^T)^T[X_i^T(-\kappa M_1)+W_i^T(-\kappa M_2)]\\ &\quad +\frac{1}{K}(-\kappa M_1^T,-\kappa M_2^T)\sum_{i=1}^{n}\delta_i f(0|X_i,U_i)(X_i^T,W_i^T)^T R(U_i)\\ &\quad +\frac{n}{K}(-\kappa M_1^T,-\kappa M_2^T)\Lambda_\lambda(\beta^*,\gamma^*)(-\kappa M_1^T,-\kappa M_2^T)^T+O_p(nK^{-1}\kappa^2)\\ &= \frac{\kappa^2}{K}\sum_{i=1}^{n}\delta_i f(0|X_i,U_i)(X_i^T M_1+W_i^T M_2)^2\\ &\quad +\frac{-\kappa}{K}\sum_{i=1}^{n}\delta_i f(0|X_i,U_i)(X_i^T M_1+W_i^T M_2)R(U_i)\\ &\quad +\frac{n\kappa^2}{K}(M_1^T,M_2^T)\Lambda_\lambda(\beta^*,\gamma^*)(M_1^T,M_2^T)^T+O_p(1)\\ &\equiv I_1+I_2+I_3,\end{aligned} \tag{5.29}$$

where $R(U_i)=\theta_0(U_i)-W_i^T\gamma_0$. Note that $\|R(U_i)\|=O(K^{-r})$, then some calculations yield $I_1=O_p(\kappa^2 nK^{-1})\|M\|^2=O_p(\|M\|^2)$, and $I_2=O_p(\kappa nK^{-1-r})\|M\|=O_p(\|M\|)$. In addition, by condition C5, we get that $\max_k p''_{\lambda_{1k}}(|\beta_k^*|)\to 0$ and $\max_l p''_{\lambda_{2l}}(|\gamma_l^*|)\to 0$. Then we have $I_3=o_p(\kappa^2 nK^{-1})\|M\|^2=o_p(\|M\|^2)$. Hence, by choosing a sufficiently large c, I_1 dominates I_2 and I_3 uniformly in $\|M\| = c$. This implies that for any given $\varepsilon > 0$, if we choose c large enough, then we have

$$P\left\{\inf_{\|M\|=c} \Delta(\beta,\gamma) > 0\right\} \geq 1-\varepsilon. \tag{5.30}$$

Hence (5.26) holds, and this implies, with probability at least $1-\varepsilon$, that there exists a local minizer $\tilde{\beta}$ and $\tilde{\gamma}$ such that

$$\|\tilde{\beta}-\beta\| = O_p(\tau) = O_p\left(n^{-r/(2r+1)}\right), \quad \|\tilde{\gamma}-\gamma\| = O_p(\tau) = O_p\left(n^{-r/(2r+1)}\right).$$

□

Proof of Theorem 3.1

Proof. From the proof of Theorem 2.3, we know that it is sufficient to show that, for any $\varepsilon > 0$, when n is large enough, we have $P\{\tilde{\beta}_k \neq 0\} < \varepsilon$, $k = d+1, \ldots, p$. Since $\tilde{\beta}_k = O_p(n^{-r/(2r+1)})$, when n is large enough, there exists some c such that

$$\begin{aligned} P\{\tilde{\beta}_k \neq 0\} &= P\{\tilde{\beta}_k \neq 0, |\tilde{\beta}_k| \geq cn^{-r/(2r+1)}\} + P\{\tilde{\beta}_k \neq 0, |\tilde{\beta}_k| < cn^{-r/(2r+1)}\} \\ &< \varepsilon/2 + P\{\tilde{\beta}_k \neq 0, |\hat{\beta}_k| < cn^{-r/(2r+1)}\}. \end{aligned} \tag{5.31}$$

Using the same arguments as in the proof of (5.23), we can obtain that

$$\begin{aligned} n^{\frac{r}{2r+1}} p'_{\lambda_{1k}}(|\tilde{\beta}_k|)\mathrm{sgn}(\tilde{\beta}_k) &= n^{\frac{-r-1}{2r+1}} \sum_{i=1}^n \delta_i f(0|X_i,U_i) X_{ik} X_i^T(\beta_0-\tilde{\beta}) \\ &\quad + n^{\frac{-r-1}{2r+1}} \sum_{i=1}^n \delta_i f(0|X_i,U_i) X_{ik} W_i^T(\gamma_0-\tilde{\gamma}) \\ &\quad + n^{\frac{-r-1}{2r+1}} \sum_{i=1}^n \delta_i f(0|X_i,U_i) X_{ik} R(U_i) + o_p(1). \end{aligned} \tag{5.32}$$

From Lemma 5.2, we have $\|\tilde{\beta}-\beta_0\| = O_p(n^{\frac{-r}{2r+1}})$, $\|\tilde{\gamma}-\gamma_0\| = O_p(n^{\frac{-r}{2r+1}})$ and $\|R(U_i)\| = O_p(n^{\frac{-r}{2r+1}})$. Hence, the three terms on the right-hand side are of order $O_p(n \cdot n^{\frac{-r-1}{2r+1}} \cdot n^{\frac{-r}{2r+1}}) = O_p(1)$. Hence, for large n, there exists some c such that

$$P(n^{\frac{r}{2r+1}} p'_{\lambda_{1k}}(|\tilde{\beta}_k|) > c) < \varepsilon/2. \tag{5.33}$$

By condition C6, we have that

$$\inf_{|\beta_k| \leq cn^{-r/(2r+1)}} n^{\frac{r}{2r+1}} p'_{\lambda_{1k}}(|\beta_k|) = n^{\frac{r}{2r+1}} \lambda_{1k} \inf_{|\beta_k| \leq cn^{-r/(2r+1)}} \lambda_{1k}^{-1} p'_{\lambda_{1k}}(|\beta_k|) \to \infty. \tag{5.34}$$

That is, $\tilde{\beta}_k \neq 0$ and $|\tilde{\beta}_k| < cn^{-r/(2r+1)}$ imply that $n^{r/(2r+1)} p'_{\lambda_{1k}}(|\tilde{\beta}_k|) > c$ for large n. Then, invoking (5.33), we have that

$$P\{\hat{\beta}_k \neq 0\} < \varepsilon/2 + P(n^{r/(2r+1)} p'_{\lambda_{1k}}(|\hat{\beta}_k|) > C) < \varepsilon.$$

This completes the proof of this theorem. □

Proof of Theorem 3.2

Proof. From the proof of Lemma 5.1, we know that

$$\|\tilde{\theta}(u) - \theta(u)\|^2 \leq 2(\tilde{\gamma}-\gamma)^T H(\tilde{\gamma}-\gamma) + 2\int_0^1 R(u)^2 du, \tag{5.35}$$

where $H = \int_0^1 B(u)B^T(u)du$. From Lemma 5.2, we get that $\|\tilde{\gamma} - \gamma\| = O_p(n^{-r/(2r+1)})$. Hence invoking $\|R(U_i)\| = O(K^{-r}) = O_p(n^{-r/(2r+1)})$, a simple calculation yields

$$\|\tilde{\theta}(u) - \theta(u)\|^2 = O_p\left\{n^{\frac{-2r}{2r+1}}\right\}. \tag{5.36}$$

Then, we complete the proof of this theorem. □

Acknowledgments

The authors sincerely thank the referees and the editor for their constructive suggestions and comments, which substantially improved an earlier version of this paper. This paper is supported by the National Natural Science Foundation of China (11301569), the Chongqing Research Program of Basic Theory and Advanced Technology (cstc2016jcyjA1365), the Research Foundation of Chongqing Municipal Education Commission(KJ1500614), and the Scientific Research Foundation of Chongqing Technology and Business University (2015-56-06).

References

[1] Koenker, R. and Bassett, G.S. *Regression quantiles*, Econometrica **46**, 33-50, 1978.

[2] Buchinsky, M. *Recent advances in quantile regression models: A practical guide for empirical research*, Journal of Human Resources **33**, 88-126, 1998.

[3] Koenker, R. and Machado, J. *Goodness of fit and related inference processes for quantile regression*, Journal of the American Statistical Association **94**, 1296-1310, 1999.

[4] Tang, C.Y. and Leng, C.L. *An empirical likelihood approach to quantile regression with auxiliary information*, Statistics & Probability Letters **82**, 29-36, 2012.

[5] Hendricks, W. and Koenker, R. *Hierarchical spline models for conditional quantiles and the demand for electricity*, Journal of the American Statistical Association **87**, 58-68, 1992.

[6] Yu, K. and Jones, M.C. *Local linear quantile regression*, Journal of The American Statistical Association **93**, 228-237, 1998.

[7] Dabrowska, D.M. *Nonparametric quantile regression with censored data*, The Indian Journal of Statistics, Series A **54**, 252-259, 1992.

[8] Lee, S. *Efficient semiparametric estimation of a partially linear quantile regression model*, Econometric Theory **19**, 1-31, 2003.

[9] Sun, Y. *Semiparametric efficient estimation of partially linear quantile regression models*, The Annals of Economics and Finance **6**, 105-127, 2005.

[10] He, X. and Liang, H. *Quantile regression estimates for a class of linear and partially linear errors-in-variables models*, Statistica Sinica **10**, 129-140, 2000.

[11] Chen, S. and Khan, S. *Semiparametric estimation of a partially linear censored regression model*, Economic Theory **17**, 567-590, 2001.

[12] Lv, X. and Li, R. *Smoothed empirical likelihood analysis of partially linear quantile regression models with missing response variables*, Advances in Statistical Analysis **97**, 317-347, 2013.

[13] Tang, C.Y. and Leng, C. *Penalized high-dimensional empirical likelihood*, Biometrika **97**, 905-920, 2010.

[14] Ren, Y.W. and Zhang, X.S. *Variable selection using penalized empirical likelihood*, Science China Mathematics **54**, 1829-1845, 2011.

[15] Variyath, A.M., Chen, J. and Abraham, B. *Empirical likelihood based variable selection*, Journal of Statistical Planning and Inference **140**, 971-981, 2010.

[16] Wu, T.T., Li, G. and Tang, C. *Empirical likelihood for censored linear regression and variable selection*, Scandinavian Journal of Statistics **42**, 798-812, 2015.

[17] Zhao, P. X. *Empirical likelihood based variable selection for varying coefficient partially linear models with censored data*, Journal of Mathematical Research with Applications **33**, 493-504, 2013.

[18] Xi, R., Li, Y. and Hu, Y. *Bayesian Quantile Regression Based on the Empirical Likelihood with spike and slab priors*, Bayesian Analysis, Online first, DOI: 10.1214/15-BA975, 2016.

[19] Hou, W., Song, L.X., Hou, X.Y. and Wang, X.G. *Penalized empirical likelihood via bridge estimator in Cox's proportional hazard model*, Communications in Statistics-Theory and Methods **43**, 426-440, 2014.

[20] Frank, I. and Friedman, J. *A statistical view of some chemometrics regression tools*, Technometrics **35**, 109-135, 1993.

[21] Tibshirani, R. *Regression shrinkage and selection via the Lasso*, J R Stat Soc Ser B **58**, 267-288, 1996.

[22] Fan, J.Q. and Li, R.Z. *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association **96**, 1348-1360, 2001.

[23] Zhang, C. *Nearly unbiased variable selection underminimax concave penalty*, Ann Stat **38**, 894-942, 2010.

[24] Schumaker, L.L. Spline Functions, Wiley, New York, 1981.

[25] Kaslow, R.A., Ostrow, D.G., Detels, R., Phair, J.P., Polk, B.F. and Rinaldo Jr, C.R. *The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants*, American Journal of Epidemiology **126**, 310-318, 1987.

[26] Huang, J., Wu, C. and Zhou, L. *Varying-coefficient models and basis function approximations for the analysis of repeated measurements*, Biometrika **89**, 111-128, 2002.

[27] Tang, Y., Wang, H. and Zhu, Z. *Variable selection in quantile varying coefficient models with longitudinal data*, Comput Stat Data Anal **57**, 435-449, 2013.

[28] Wang, H., Zhu, Z. and Zhou, J. *Quantile regression in partially linear varying coefficient models*, Ann Stat. **37**, 3841-3866, 2009.

[29] Zhao, P.X. and Xue, L.G. *Variable selection for semiparametric varying coefficient partially linear errors-in-variables models*, Journal of Multivariate Analysis **101**, 1872-1883, 2010.

[30] Fan, J.Q. and Li, R. *New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis*, Journal of the American Statistical Association **99**, 710-723, 2004.

[31] Xue, L.G. and Zhu, L.X. *Empirical likelihood semiparametric regression analysis for longitudinal data*, Biometrika **94**, 921-937, 2007.

[32] Xue, L.G. *Empirical likelihood for linear models with missing responses*, Journal of Multivariate Analysis **100**, 1353-1366, 2009.

[33] Zhao, P.X. and Xue, L.G. *Empirical likelihood inferences for semiparametric varying coefficient partially linear models with longitudinal data*, Communications in Statistics-Theory and Methods **39**, 1898-1914, 2010.