Statistica Sinica **25** (2015), 1437-1458
doi:http://dx.doi.org/10.5705/ss.2013.317

SEMIPARAMETRIC PARTIAL LINEAR QUANTILE REGRESSION OF LONGITUDINAL DATA WITH TIME VARYING COEFFICIENTS AND INFORMATIVE OBSERVATION TIMES

Xuerong Chen, Jianguo Sun and Lei Liu

Southwestern University of Finance and Economics, University of Missouri and Northwestern University

Abstract: Regression analysis of longitudinal data has been a popular topic in many fields for long time. However, only limited research exists for the case where observation times may be informative and for quantile regression of longitudinal data. In particular, to our knowledge, there does not exist any established method for quantile regression of longitudinal data with informative observation times, the focus of this paper. More specifically, we discuss this problem and present a semiparametric partial linear model with time-varying coefficients. For estimation, B-splines are used to approximate the time-varying coefficients and in addition to the estimation approach, model checking and selection procedures are also provided. The latter can be used to determine the covariates that indeed have time-varying effects on the longitudinal process of interest. The proposed method can identify the underlying true model structure and estimate the parameters simultaneously. Also we establish the convergence rate of the proposed estimators and the asymptotic normality of the estimated time-independent regression parameters. For the assessment of the finite sample performance of the proposed methods, an extensive simulation study is conducted and suggests that they work well for practical situations. They are applied to a set of longitudinal medical cost data on chronic heart failure patients that motivated this study.

Key words and phrases: B-splines, group penalized model selection, informative observation times, semiparametric time-varying coefficient model.

1. Introduction

Longitudinal data occur in many fields including epidemiological studies, medical follow-up studies, and observational studies and extensive literature has been established for their analysis (Diggle et al. (2013)). A typical feature of longitudinal data is that study subjects are usually observed repeatedly at different and irregular time points and for their analysis, a question that one needs to pay attention is how these observation times are generated. A simple situation, discussed the most in the literature, is that they can be treated as constants and thus one can perform some conditional analysis, see Fan and Li (2004), Fan, Huang, and Li (2007), Fan and Wu (2008) etc. A more general situation is that the mechanism behind the observation time may be different for different subjects or that they depend on covariates. Among others, Lin and Ying (2001) discussed this case and treated the observation times as realizations of some underlying point processes, often referred to as observation processes. They modeled the observation process by using the proportional rate model (Cook and Lawless (2007)). In this case the longitudinal response process of interest is observed only at the time points where the observation process jumps. We consider a more complicated situation where the observation process may be informative, as defined below (Sun et al. (2005); Sun, Sun, and Liu (2007)).

This paper was motivated by the analysis of medical cost data from a study of chronic heart failure patients at the University of Virginia Health System. The study consists of 1,475 patients, aged 60-89 years, who were first diagnosed with heart failure and treated in 2004. For each patient, the observed information includes the clinical visit or observation times in months, the medical cost for each clinical visit and three baseline covariates: age, gender, and race. All patients were followed until the end of the study, July, 2006, or their death and one main objective of the study was to investigate the relationship between the medical cost and the covariates. Preliminary studies indicated that the patient visiting the hospital more often tends to pay more for each visit. Thus the observation times contain some information about medical cost and are thus informative about the response process of interest (Liu, Huang, and O'Quigley (2008)). It has been shown that one can obtain biased or misleading results by ignoring such information (Sun et al. (2005)). In addition to the informative observation times, another common characteristic of medical cost data is that they are highly skewed to the right.

Several approaches have been developed for regression analysis of longitudinal data with informative observation times. For example, among others, Sun et al. (2005) considered the problem and proposed a marginal model approach for the analysis, while Sun, Sun, and Liu (2007) and Liang, Lu, and Ying (2009) developed some procedures that model the longitudinal process and the observation process jointly through latent variables. Note that these methods and other existing approaches are based on mean regression and assume that covariate effects are constant. Instead of mean regression, quantile regression is also often used for the analysis of longitudinal data, and it is well-known that the latter is usually more explicable and robust than the former when the data are skewed or contain outliers. However, There is no established approach for quantile regression analysis of longitudinal data with informative observation times.

To relax the restriction of constant covariate effects, time-varying coefficient models are often used (Sun, Sun, and Zhou (2013)). For example, Martinussen and Scheike (1999, 2000, 2001) and Sun and Wu (2005) considered some mean regression models with time-varying coefficients for the analysis of longitudinal data, while Kim (2007) and Wang, Zhu, and Zhou (2009) gave some time-varying coefficient models for quantile regression of longitudinal data. On the last two references, the former employed multipolynomials spline to approximate the varying coefficient function and the latter made use of B spline approximations. Under i.i.d sample, Kai, Li, and Zou (2011) considered the quantile regression of varying coefficient partially linear model by using local linear approximation technique. It should be noted that all of the methods above assume that observation times are noninformative. With the use of time-varying coefficient models, a natural question is how to determine if a covariate has time-varying or constant effect. To address this, in the case of mean regression, a common method is to apply generalized likelihood ratio tests (Fan, Zhang, and Zhang (2001)), while one can apply the rank score test for quantile regression (Kim (2007); Wei and He (2006)). Recently, the group penalized method (Huang, Wei, and Ma (2012); Zhang, Cheng, and Liu (2011)) has attracted a lot of attention because it can determine or select the model structure and estimate parameters simultaneously.

We consider semiparametric quantile regression of longitudinal data in the presence of informative observation times and time-varying coefficients or covariate effects. In particular, a semiparametric partial linear time-varying coefficient model is presented and the counting process is used to describe observation times or processes. The framework allows the observation processes to depend on covariates. This has not been considered in the case of quantile regression. In the proposed approach, we employ B-splines to approximate the time-varying coefficients and develop some sieve-based estimating equations for estimation of unknown parameters. Also a group penalized procedure is proposed to select the covariates or assess the model structure, and the MM algorithm is used to deal with the difficulties caused by the nonsmooth of the checking function in quantile regression, as well as the penalty term.

The remainder of the paper is organized as follows. Section 2 introduces some notation and the models to be used throughout the paper. Section 3 discusses a sieve estimating procedure, and the asymptotic properties of the proposed estimators are established in Section 4. A model checking procedure is presented in Section 4 for the assessment of the adequacy of the proposed models. In Section 5, we consider the assessment of the nature of covariate effects or the model structure and a group penalized-based model selection procedure is presented with the use of the MM algorithm. In addition, the selection of interior knots and penalty tuning parameter is discussed. Section 6 gives some results obtained from an extensive simulation study conducted to evaluate the finite sample performance of the proposed methodology. In Section 7, we apply the methodology to medical cost data, and some concluding remarks are given in Section 8.

2. Notation and Models

Consider a longitudinal study that consists of n independent subjects. For subject i, let $Y_i(t)$ denote the longitudinal process of interest and suppose that there exist a p-dimensional vector of possibly time-dependent covariates, denoted by $X_i(t)$, and a follow-up time C_i, $i = 1, \ldots, n$. Suppose that $Y_i(t)$ is observed only at time points $t_{i1} < t_{i2} < \cdots < t_{im_i}$, where m_i denotes the total number of observations on the ith subject. Let $N_i(t) = \sum_{j=1}^{m_i} I(t_{ij} \le t) = N_i^*(\min(t, C_i))$, where $N_i^*(t)$ is the underlying counting or observation process that characterizes the observation times. We assume that the covariate history $\{X_i(t) : 0 \le t \le C_i\}$ is observed for each subject.

To describe the effects of covariates on $Y_i(t)$ as well as $N_i^*(t)$, let $\mathcal{F}_{it} = \{N_i(s), 0 \le s < t\}$, the observation history up to time t for the ith subject, $i = 1, \ldots, n$. Suppose that we can write the covariates as $X_i(t) = (X_{1i}^T(t), X_{2i}^T(t))^T$, where $X_{1i}(t)$ represents the covariates that may have time-varying effects on $Y_i(t)$ and $X_{2i}(t)$ denotes the covariates that only have time-independent effects. For a given quantile level $\tau \in (0,1)$, we assume that $Y_i(t)$ satisfies the quantile partial linear time-varying coefficient model

$$Y_i(t) = \alpha^T(\tau, t) X_{1i}(t) + \beta^T(\tau) X_{2i}(t) + \varrho^T(\tau) H(\mathcal{F}_{it}) + \epsilon_i(\tau, t). \tag{2.1}$$

Here $\alpha(\tau, t)$ is a p_1-dimensional vector of time-varying coefficient, $\beta(\tau)$ is a p_2-dimensional vector of unknown regression parameters, $\varrho(\tau)$ is a p_3-dimensional vector of regression coefficients, H is a vector of known functions of the observation process up to time $t-$, and $\epsilon_i(\tau, t)$ is random error whose τth quantile is zero. The error term $\epsilon_i(\tau, t)$ may be time-dependent. For simplicity, we suppress the τ in $\alpha(\tau, t)$, $\beta(\tau)$, $\varrho(\tau)$ and $\epsilon_i(\tau, t)$ below.

The model at (2.1) has the longitudinal response process depending on the covariates, and the history of the observation process. Similar models have been used in Sun et al. (2005) among others. With respect to the function vector H in (2.1), there are several possible choices. A simple and natural one is $H(\mathcal{F}_{it}) = N_i(t-)$, which means that $\mathcal{F}_{it}$ affects the conditional quantile of the response variable through the total number of observations. Another choice is $H(\mathcal{F}_{it}) = (N_i(t-) - N_i(t-u))$, that the number of the observations in the past u time units contains relevant information about the response variable. In general, the selection of H should be based on the problem of interest and the prior knowledge about the possible relationship between the response process and the observation process.

For the observation process, we assume that $N_i^*(t)$ is a nonhomogeneous Poisson process satisfying the proportional rate model

$$E\{ dN_i^*(t)|X_i(t) \} = e^{\gamma^T X_i(t)}\, d\Lambda_0(t) \tag{2.2}$$

(Cook and Lawless (2007)), $i = 1, \ldots, n$. Here γ is a vector of unknown regression parameters and $\Lambda_0(t)$ is the unspecified baseline cumulative mean function of $N_i^*(t)$. We assume the follow-up time C_i depends on covariates $X_i(t)$ in an arbitrary fashion, but is independent of $N_i^*(t)$ and $Y_i(t)$, given $X_i(t)$ and $\mathcal{F}_{it}$, in the sense that

$$E\{ dN_i^*(t)|X_i(t), C_i \geq t \} = E\{ dN_i^*(t)|X_i(t) \},$$
$$E\{ Y_i(t)|X_i(t), \mathcal{F}_{it}, C_i \geq t \} = E\{ Y_i(t)|X_i(t), \mathcal{F}_{it} \}.$$

For simplicity, we restrict inference to a finite time interval $[0, t_0]$.

3. Estimation Procedure

In this section, we develop the estimation procedure for models (2.1) and (2.2). Let $Z_i(t) = (X_{2i}^T(t), H^T(\mathcal{F}_{it}))^T$, $\theta = (\beta^T, \varrho^T)^T$ and $\xi_i(t) = I(C_i \geq t)$ and first assume that γ and $\Lambda_0(t)$ are known. Here if $\gamma = 0$, a natural approach for estimating α and β is to minimize

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \rho_\tau(\epsilon_i(t))\, dN_i(t) \tag{3.1}$$

by following the quantile regression principle of Koenker and Bassett (1978), where $\rho_\tau(\epsilon) = \epsilon \times (\tau - I(\epsilon < 0))$ is the checking function. It is easy to see that this is equivalent to solving the estimating equation

$$\frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \begin{pmatrix} X_i(t) \\ H(\mathcal{F}_{it}) \end{pmatrix} \Big[I(\epsilon_i(t) \leq 0) - \tau \Big]\, dN_i(t) = o_p(n^{-1/2}). \tag{3.2}$$

Motivated this, for $\gamma \neq 0$, we can consider the function

$$\begin{aligned} g_i(\alpha, \theta, \gamma, \Lambda_0) = \int_0^{t_0} \begin{pmatrix} X_{1i}(t) \\ Z_i(t) \end{pmatrix} \Big[& I(Y_i(t) - \alpha^T(t) X_{1i}(t) - \theta^T Z_i(t) \leq 0) dN_i(t) \\ & - \tau \xi_i(t) e^{\gamma^T X_i(t)} d\Lambda_0(t) \Big] \end{aligned} \tag{3.3}$$

and thus the estimating equation

$$G_n(\alpha, \theta, \gamma, \Lambda_0) = \frac{1}{n} \sum_{i=1}^{n} g_i(\alpha, \theta, \gamma, \Lambda_0) = 0.$$

It is easy to show that $E\{g_i(\alpha_0, \theta_0, \gamma, \Lambda_0)\} = 0$ under models (2.1) and (2.2) and the assumptions, and thus this estimating equation is unbiased.

It is almost impossible to solve this estimating equation directly due to the dimension of $\alpha(t)$. To deal with this, we propose to first approximate $\alpha(t)$ by B-splines (Schumarker (1981)). For this, for given integers l, m, and $r = l + m$, let

$$\mathcal{H}_{rj} = \{\, g(\cdot) : |g^{(l)}(t_1) - g^{(l)}(t_2)| \le c_j |t_1 - t_2|^m, \ \text{for any}\, 0 \le t_1, t_2, \le t_0 \,\},$$

where c_j is a finite positive constant, $j = 1, \ldots, p_1$. Take $\mathcal{H}_r = \prod_{j=1}^{p_1} \mathcal{H}_{rj}$ and assume that the parameter space $\mathcal{B}$ for $\theta = (\beta^T, \varrho^T)^T$ is a compact subset of $\mathbb{R}^{p_2+p_3}$. We assume that $\alpha(t) = (\alpha_1(t), \ldots, \alpha_{p_1}(t))^T$ belongs to $\mathcal{H}_r$ and thus the whole parameter space is given by $\Theta = \mathcal{H}_r \times \mathcal{B}$.

Let $B_n(\cdot) = \{b_1(\cdot), \ldots, b_{k_n+l+1}(\cdot)\}^T$ denote a set of B-spline basis functions of order $l + 1$ with knots $0 = d_0 < \cdots < d_{k_n} < d_{k_n+1} = t_0$ and satisfying $\max_{1 \le k \le k_n} |d_k - d_{k+1}| = O(n^{-v})$. By assumption, the $\alpha_j(t)$'s are the lth differential functions. This suggests that one can approximate the jth component $\alpha_j(t)$ of $\alpha(t)$ by

$$\alpha_{nj}(t) = \sum_{k=1}^{k_n+l+1} b_k(t)\, \vartheta_{jk} = B_n^T(t)\, \vartheta_j.$$

Let $\mathcal{A}$ denote the parameter space of ϑ_j and assume that it is a bounded subset of $\mathbb{R}^{k_n+l+1}$. Take $\mathcal{H}_r^n = \prod_{j=1}^{p_1} \mathcal{H}_{rj}^n$ and $\Theta_n = \mathcal{H}_r^n \times \mathcal{B}$, where

$$\mathcal{H}_{rj}^n = \{\, \alpha_{nj} : \alpha_{nj}(t) = B_n^T(t)\vartheta_j, \ \ \vartheta_j \in \mathbb{R}^{k_n+l+1}, t \in [0, t_0] \,\}.$$

It is easy to see that $\{\Theta_n \in \Theta, n = 1, 2, \ldots\}$ is a sieve for the parameter space Θ. Let $\tilde{X}_{1i}(t) = (X_{1i1}(t)B_n^T(t), \ldots, X_{1ip_1}(t)B_n^T(t))^T \in \mathbb{R}^{p_{k_n}}$, $\tilde{Z}_i(t) = (\tilde{X}_{1i}^T(t), Z_i^T(t))^T$, and $\vartheta = (\vartheta_1^T, \ldots, \vartheta_{p_1}^T)^T \in \mathbb{R}^{p_{k_n}}$, where $p_{k_n} = p_1(k_n + l + 1)$. Motivated by (3.3), it is natural to consider the sieve estimating function

$$\begin{aligned} &g_i^s(\vartheta, \theta, \gamma, \Lambda_0) \\ &= \int_0^{t_0} \tilde{Z}_i(t) \Big[I(Y_i(t) - \vartheta^T \tilde{X}_{1i}(t) - \theta^T Z_i(t) \le 0) dN_i(t) - \tau \xi_i(t) e^{\gamma^T X_i(t)} d\Lambda_0(t) \Big] \end{aligned}$$

and therefore the sieve estimating equation

$$G_n^s(\vartheta, \theta, \gamma, \Lambda_0) = \frac{1}{n} \sum_{i=1}^{n} g_i^s(\vartheta, \theta, \gamma, \Lambda_0) = 0 \tag{3.4}$$

for estimation of ϑ and θ.

In this, it has been assumed that both γ and $\Lambda_0(t)$ are known, not really true. On the other hand, they can be easily estimated. Specifically, by following Lin et al. (2000), one can estimate γ by solving the estimating equation

$$\sum_{i=1}^{n} \int_0^{t_0} \{X_i(t) - \bar{X}(t;\gamma)\} dN_i(t) = 0.$$

Here $\bar{X}(t;\gamma) = S^{(1)}(t;\gamma)/S^{(0)}(t;\gamma)$ and

$$S^{(k)}(t;\gamma) = \frac{1}{n} \sum_{j=1}^{n} \xi_j(t) X_j(t)^{\otimes k} \exp(\gamma^T X_j(t)),$$

$k = 0, 1, 2$, where $a^{\otimes 2} = aa^T$ for any vector a. Let $\widehat{\gamma}$ denote this estimator of γ. Then we can estimate $\Lambda_0(t)$ by the Nelson-Aalen type estimator

$$\widehat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(u)}{S^{(0)}(t;\widehat{\gamma})}.$$

By plugging $\widehat{\gamma}$ and $\widehat{\Lambda}_0(t)$ into (3.4), we obtain

$$G_n^s(\vartheta, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) = \frac{1}{n} \sum_{i=1}^{n} g_i^s(\vartheta, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) = 0, \tag{3.5}$$

which is asymptotically unbiased.

For large n and k_n, the number of knots, solving of these estimating equations is not easy due to the high dimension. Also the derivation of the asymptotic properties of the resulting estimators would be difficult. To simplify, take $\epsilon_{ni}(t) = Y_i(t) - \alpha_n^T(t) X_{1i}(t) - \theta^T Z_i(t)$. Then, since $\vartheta^T \tilde{X}_{1i}(t) = \alpha_n^T(t) X_{1i}(t)$ we can rewrite $G_n^s(\vartheta, \theta, \widehat{\gamma}, \widehat{\Lambda}_0)$ as

$$G_n^s(\vartheta, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \tilde{Z}_i(t) \Big\{ \Big[I(\epsilon_{ni}(t) \le 0) - \tau \Big] dN_i(t) + \tau d\widehat{M}_i(t) \Big\},$$

where $\widehat{M}_i(t) = N_i(t) - \int_0^t \xi_i(s) \exp(\widehat{\gamma}^T X_i(s))\, d\widehat{\Lambda}_0(t)$. Motivated by (3.1) and (3.2), we propose to consider the sieve objective function

$$\Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0) = \frac{1}{n} \sum_{i=1}^{n} \int_0^{t_0} \Big\{ \rho_\tau(\epsilon_{ni}(t)) dN_i(t) - \tau \epsilon_{ni}(t) d\widehat{M}_i(t) \Big\}. \tag{3.6}$$

Then solving (3.5) is equivalent to minimizing the objective function (3.6).

Define the estimators of $\alpha(t)$ and θ as

$$(\widehat{\alpha}_n(t), \widehat{\theta}) = \arg \min_{(\alpha_n, \theta) \in \Theta_n} \Psi_n(\alpha_n, \theta, \widehat{\gamma}, \widehat{\Lambda}_0).$$

For the determination of $\widehat{\alpha}_n(t)$ and $\widehat{\theta}$, one can employ the existing result about M-estimators.

4. Asymptotic Properties and Model Checking

In this section, we establish the asymptotic properties of $\widehat{\alpha}_n(t)$ and $\widehat{\theta}$ and discuss their variance estimation. In addition, a procedure is presented for checking the appropriateness of the proposed models.

For any $\alpha(t), \tilde{\alpha}(t) \in \mathcal{H}_r$, let

$$\|\alpha - \tilde{\alpha}\|_{\mathcal{H}_r}^2 = \sum_{j=1}^{p_1} \int_0^{t_0} [\alpha(t) - \tilde{\alpha}(t)]^2 \, dt$$

and $\rho(\eta, \tilde{\eta}) = (\|\alpha - \tilde{\alpha}\|_{\mathcal{H}_r}^2 + \|\theta - \tilde{\theta}\|^2)^{1/2}$ for any $\eta = (\alpha, \theta), \tilde{\eta} = (\tilde{\alpha}, \tilde{\theta}) \in \Theta$. Let $\widehat{\eta}_n = (\widehat{\alpha}_n(t), \widehat{\theta})$ and $\eta_0 = (\alpha_0(t), \theta_0)$ denote the true values of η. Also take

$$\begin{aligned}
A_n^* &= \frac{1}{n} \sum_{i=1}^n E\bigg\{ \int_0^{t_0} Z_i^{**\otimes 2}(t) f_{\epsilon_i}(0|X_i(t), \mathcal{F}_{it}) dN_i(t) \bigg\}, \\
V_n^* &= \frac{1}{n} \sum_{i=1}^n E\bigg\{ h_i^{**}(\alpha_0, \theta_0) + \tau \int_0^{t_0} \bigg[Z_i^{**}(t) - \bar{z}^{**}(t; \gamma_0) \bigg] dM_i(t) \\
&\quad - \tau P_n^* \Omega_n^{-1} \int_0^{t_0} \bigg[X_i(t) - \bar{x}(t; \gamma_0) \bigg] dM_i(t) \bigg\}^{\otimes 2},
\end{aligned}$$

where such involved quantities as Z_i^{**} and $h_i^{**}(\alpha_0, \theta_0)$ are defined in the proof that is available online.

To establish asymptotic properties, we need some regularity conditions. Suppose that γ belongs to a compact parameter space $\tilde{\Theta}$.

(A_1) For some $r \geq 1, \alpha_0(t) \in \mathcal{H}_r$.

(A_2) Both $X(t)$ and $H(\cdot)$ have bounded total variation on $[0, t_0]$.

(A_3) The density functions $f_{\epsilon_i}(.\,|X_i(t), \mathcal{F}_{it}), i = 1, \ldots, n$ of the random errors is uniformly bounded away from zero and infinity.

(A_4) $E[\max_{i,t \in [0,t_0]} \|Z_i^*(t)\|^2] < \infty$.

(A_5) The eigenvalues of A and V(the limits of A_n^* and V_n^*) are bounded away from infinity and zero for sufficiently large n.

(A_6) $\{N_i(\cdot), X_i(\cdot), \xi_i(\cdot)\}, i = 1, 2, \ldots, n$ are i.i.d..

(A_7) $P(C_i \geq t_0) > 0, i = 1, 2, \ldots, n$.

(A_8) $N_i(t_0), i = 1, 2, \ldots, n$ are bounded by a constant.

(A_9) The following matrix is positive definite:

$$E\bigg\{\int_0^{t_0}\bigg[X_i(t)-\bar{x}(t;\gamma_0)\bigg]^{\otimes 2}\xi_i(t)\exp(\gamma_0^T X_i(t))d\Lambda_0(t)\bigg\}.$$

Condition (A_1) holds if $\alpha_0(\cdot)$ has a bounded rth order derivative on $[0,t_0]$ and is commonly used in the spline-based literature. Condition (A_2) is common in the literature on time-varying covariate effect models, while (A_3) is a standard assumption used in quantile regression. Condition (A_5) is needed to ensure that the asymptotic covariance of $\widehat{\theta}$ exists, and $(A_6)-(A_9)$ are required for the derivation of the asymptotic normality and weak convergence of $\widehat{\gamma}$ and $\widehat{\Lambda}_0(t)$. Similar conditions have been used in Lin et al. (2000) among others.

Theorem 1. *If $(A_1)-(A_3)$ and $(A_6)-(A_9)$ hold, and $k_n\to\infty$ and $k_n/n\to 0$ as $n\to\infty$, $r\geq 1$, then $\rho(\widehat{\eta}_n,\eta_0)=O_p((k_n/n)^{1/2}+k_n^{-r})$. For $k_n\asymp n^{1/(2r+1)}$, we have $\rho(\widehat{\eta}_n,\eta_0)=O_p(n^{-r/(2r+1)})$.*

Theorem 2. *If $(A_1)-(A_9)$ hold, and $k_n\to\infty, k_n^2/n\to 0, nk_n^{-4r}\to 0$ as $n\to\infty$, $r\geq 1$, then $\sqrt{n}(\widehat{\theta}-\theta_0)\xrightarrow{D} N(0,\Sigma)$, where $\Sigma=A^{-1}VA^{-1}$.*

The proofs of these theorems are sketched in the supplementary material. Theorem 1 tells us that $\widehat{\alpha}_n(t)$ is consistent, and can achieve the optimal convergence rate in the usual nonparametric regression setting (Stone (1980)). To use the results, we need to estimate the covariance matrix Σ and, for this, we propose a bootstrap procedure.

Let B denote an integer and select B random samples each of size n with replacement from $(Y_i(t),X_{1i}(t),Z_i(t),C_i), i=1,2,\ldots,n$. Let the j-th bootstrapped sample be $(Y_i^{(j)}(t),X_{1i}^{(j)}(t),Z_i^{(j)}(t),C_i^{(j)})$, $i=1,2,\ldots,n$, $j=1,2,\ldots,B$, and with corresponding observation time points $t_{i1}^{(j)}<t_{i2}^{(j)}<\cdots<t_{im_i^{(j)}}^{(j)}$. Then $N_i^{(j)}(t)=\sum_{k=1}^{m_i^{(j)}}I(t_{ik}^{(j)}\leq t)$. Let $(\widehat{\alpha}_n^{(j)}(t),\widehat{\theta}^{(j)})$ be the minimizer of the bootstrapped objective function

$$\frac{1}{n}\sum_{i=1}^n\int_0^{t_0}\bigg\{\rho_\tau(\epsilon_{ni}^{(j)}(t))dN_i^{(j)}(t)-\tau\epsilon_{ni}^{(j)}(t)d\widehat{M}_i^{(j)}(t)\bigg\}.$$

Here, $\epsilon_{ni}^{(j)}(t)$ and $\widehat{M}_i^{(j)}(t)$ are defined as $\epsilon_{ni}(t)$ and $\widehat{M}_i(t)$, but based on the j-th bootstrapped sample and the resulting estimators $\widehat{\gamma}^{(j)}$ and $\widehat{\Lambda}_0^{(j)}(t)$. Then one can use the empirical variance of $\{\widehat{\alpha}_n^{(1)}(t),\widehat{\theta}^{(1)},\ldots,\widehat{\alpha}_n^{(B)}(t),\widehat{\theta}^{(B)}\}$ to estimate the asymptotic variance of $\widehat{\alpha}_n(t)$ and $\widehat{\theta}$. By Cheng and Huang (2010), the resulting estimators are consistent and thus can be used to make inference on $\alpha(t)$ and θ.

For this method, the question of the adequacy of models (2.1) and (2.2) arises. For model (2.2), one can directly employ the approach given in Lin et al. (2000). To check model (2.1), similar to Chen, Wei, and Parzen (2004) and Sun et al. (2012), we propose to consider the cumulative sums of residuals

$$\mathcal{L}_\tau(u, x_1, z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^u I(X_{i1} \le x_1, Z_i \le z) \Big[I(\widehat{\epsilon}_{ni}(t) \le 0)\, dN_i(t) - \tau \xi_i(t) e^{\widehat{\gamma}^T X_i(t)} d\widehat{\Lambda}_0(t) \Big],$$

where $\widehat{\epsilon}_{ni}(t) = Y_i(t) - \widehat{\alpha}_n^T(t) X_{1i}(t) - \widehat{\theta}^T Z_i(t)$, and the event $I(X_{i1} \le x_1, Z_i \le z)$ means that each component of X_{1i} and Z_i is no larger than the corresponding component of x_1 and z. We show in Appendix II of the supplementary material that the null distribution of $\mathcal{L}_\tau(u, x_1, z)$ can be approximated by a zero-mean Gaussian process

$$\tilde{\mathcal{L}}_\tau(u, x_1, z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{1i}(u) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{2i}(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{3i}(u),$$

where $\widehat{\mathcal{L}}_{ki}, k = 1, 2, 3$, are obtained by replacing the unknown quantities in $\mathcal{L}_{ki}$, defined in technical proof supplementary material, with their estimators.

It is difficult to estimate the null distribution analytically. To handle this problem, using a resampling approach similar to that used in Cheng, Wei, and Ying (1997) and Sun et al. (2012) one can approximate the null distribution of $\mathcal{L}_\tau(u, x_1, z)$ by the conditional distribution of $\widehat{\mathcal{L}}_\tau(u, x_1, z)$, where

$$\widehat{\mathcal{L}}_\tau(u, x_1, z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{1i}(u) U_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{2i}(u) U_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widehat{\mathcal{L}}_{3i}(u) U_i$$

with the U_i's being a random sample from the standard normal distribution. In reality, one can obtain a large number of realizations from $\widehat{\mathcal{L}}_\tau(u, x_1, z)$ by repeatedly generating $(U_1, \ldots, U_n)$ while fixing the observation data. An unusual pattern of $\mathcal{L}_\tau(u, x_1, z)$ comparied to $\widehat{\mathcal{L}}_\tau(u, x_1, z)$ would suggest a lack-of-fit of model (2.1). Since $\mathcal{L}_\tau(t, x_1, z)$ is expected to fluctuate randomly around 0 under model (2.1), a formal lack-of-fit test could also be constructed based on the supremum statistic $\sup_{0 \le u \le t_0, x_1, z} |\mathcal{L}_\tau(u, x_1, z)|$. The p-value of this test can be obtained by comparing the observed value of $\sup_{0 \le u \le t_0, x_1, z} |\mathcal{L}_\tau(u, x_1, z)|$ to a large number of realizations of $\widehat{\mathcal{L}}_\tau(u, x_1, z)$.

5. Model Selection Procedure

In practice, one may not know which of the covariates of interest have time-varying or time-independent effects. To address this, we present a selection procedure developed based on the group smoothly clipped absolute deviation (SCAD) penalized method (Fan and Li (2001)). Let $p_\lambda(b)$ denote the SCAD penalty function for $b > 0$, with

$$p'_\lambda(b) = \lambda\left\{I(b \le \lambda) + \frac{(a\lambda - \|\bar{\vartheta}_{j*}\|)_+}{(a-1)\lambda} I(b > \lambda)\right\},$$

where $a > 2$ and $\lambda > 0$ are tuning parameters. This penalty function is symmetric around the origin.

We first assume that both $X_i(t)$ and $H(\mathcal{F}_{it})$ in model (2.1) have time-varying effects and in this case, model (2.1) has the form

$$Y_i(t) = \alpha^T(t)X_{1i}(t) + \beta^T(t)X_{2i}(t) + \varrho^T(t)H(\mathcal{F}_{it}) + \epsilon_i(t). \tag{5.1}$$

Let $\bar{B}_n(\cdot) = AB_n(\cdot) = \{1, \bar{B}_{2n}(\cdot)\} = \{1, \bar{b}_2(\cdot), \ldots, \bar{b}_{k_n+l+1}(\cdot)\}^T$. As before, we approximate $\alpha_j(t)$, $\beta_j(t)$, and $\varrho_j(t)$ by

$$\begin{aligned}
\alpha_j(t) &\approx \bar{\vartheta}_{j1} + \sum_{k=2}^{k_n+l+1} \bar{b}_k(t)\bar{\vartheta}_{jk} = \bar{\vartheta}_{j1} + \bar{B}_{2n}^T(t)\bar{\vartheta}_{j*} = \bar{B}_n^T(t)\bar{\vartheta}_j,\ j = 1, \ldots, p_1,\\
\beta_j(t) &\approx \bar{\vartheta}_{j1} + \bar{B}_{2n}^T(t)\bar{\vartheta}_{j*} = \bar{B}_n^T(t)\bar{\vartheta}_j,\ j = p_1 + 1, \ldots, p_2,\\
\varrho_j(t) &\approx \bar{\vartheta}_{j1} + \bar{B}_{2n}^T(t)\bar{\vartheta}_{j*} = \bar{B}_n^T(t)\bar{\vartheta}_j\ ,\ j = p_2 + 1, \ldots, p_3,
\end{aligned}$$

respectively, where $\bar{\vartheta}_{j*} \in \mathbb{R}^{k_n+l}$ and $\bar{\vartheta}_j = (\bar{\vartheta}_{j1}, \bar{\vartheta}_{j*}^T)^T$. Here $\{\, \bar{\vartheta}_{j1}, j = 1, \ldots, p = p_1 + p_2 + p_3 \,\}$ correspond to the constant part of the coefficients, while $\{\, \bar{\vartheta}_{j*}, j = 1, \ldots, p \,\}$ correspond to the time-varying part.

Let $\bar{\vartheta} = (\bar{\vartheta}_1^T, \ldots, \bar{\vartheta}_p^T)^T$ and $\mathcal{X}_i(t) = [X_{1i1}(t)\bar{B}_n^T(t), \ldots, X_{1ip_1}(t)\bar{B}_n^T(t), X_{2i1}(t)\bar{B}_n^T(t), \ldots, X_{2ip_2}(t)\bar{B}_n^T(t), H_1(\mathcal{F}_{it})\bar{B}_n^T(t), \ldots, H_{p_3}(\mathcal{F}_{it})\bar{B}_n^T(t)]^T$. For $j = 1, \ldots, p$, if $\|\bar{\vartheta}_{j*}\| = (\bar{\vartheta}_{j*}^T\bar{\vartheta}_{j*})^{1/2} = 0$, the jth covariate only has constant or time-independent effect, otherwise, it has time-varying effect. Let $\bar{\epsilon}_{ni}(t) = Y_i(t) - \bar{\vartheta}^T\mathcal{X}_i(t)$. We consider the penalized loss function

$$\Psi_n^P(\bar{\vartheta}, \widehat{\gamma}, \widehat{\Lambda}_0) = \sum_{i=1}^n \int_0^{t_0} \left\{\rho_\tau(\bar{\epsilon}_{ni}(t))dN_i(t) - \tau\bar{\epsilon}_{ni}(t)d\widehat{M}_i(t)\right\} + n\sum_{j=1}^p p_\lambda(\|\bar{\vartheta}_{j*}\|). \tag{5.2}$$

Without loss of generality, assume that $\alpha_0(t)$ is the vector of the time-varying coefficients and $\theta_0(t) = (\beta_0(t)^T, \varrho_0(t)^T)^T = (\beta_0^T, \varrho_0^T)^T = \theta_0$, the vector of constant coefficients. Let the $\widehat{\bar{\vartheta}}_j$'s denote the values of the $\bar{\vartheta}_j$'s that minimize the loss function. Then $\alpha_j(t)$ and θ can be estimated by $\widehat{\alpha}_{nj}^P(t) = \bar{B}_n^T(t)\widehat{\bar{\vartheta}}_j$ and $\widehat{\theta}^P = (\widehat{\bar{\vartheta}}_{(p_1+1)1}, \ldots, \widehat{\bar{\vartheta}}_{p_21}, \widehat{\bar{\vartheta}}_{(p_2+1)1}, \ldots, \widehat{\bar{\vartheta}}_{p1})^T$, respectively.

To establish the asymptotic properties of $\widehat{\alpha}_n^P(t)$ and $\widehat{\theta}^P$, let $\widehat{\eta}_n^P = (\widehat{\alpha}_n^P(t), \widehat{\theta}^P)$ and take

$$\overline{A}_n = \frac{1}{n}\sum_{i=1}^{n} E\Big\{\int_0^{t_0} \mathcal{X}_i(t)\mathcal{X}_i^T(t) f_\epsilon(0|X_i(t),\mathcal{F}_{it})dN_i(t)\Big\}.$$

We need a condition, to ensure that the asymptotic variance of $\widehat{\theta}^P$ exists.

($\tilde{A}_5$) The eigenvalues of $A, \bar{A}$ and V are bounded away from infinity and zero for sufficiently large n.

Theorem 3. *Assume the conditions* $(A_1)-(A_4), (\tilde{A}_5)$ *and* $(A_6)-(A_9)$ *hold. If* $k_n \to \infty$, $k_n^2/n \to 0$, $nk_n^{-4r} \to 0$, $\lambda = \lambda_n \to 0$, *and* $\lambda_n/((k_n/n)^{1/2} + k_n^{-r}) \to \infty$ *as* $n \to \infty$, $r \geq 1$, *then*

(a) $\widehat{\theta}^P$ *is a constant vector with probability approaching* 1*;*

(b) $\rho(\widehat{\eta}_n^P, \eta_0) = O_p((k_n/n)^{1/2} + k_n^{-r})$*;*

(c) $\sqrt{n}(\widehat{\theta}^P - \theta_0) \xrightarrow{D} N(0,\Sigma)$, *where* Σ *is as in Theorem* 2.

The proof is sketched in the supplementary material. The objective function at (10) has three terms inside the summation and its minimization not straightforward in general. We suggest an iterative method that replaces two terms by smoother surrogate functions. After its kth iteration, let $\bar{\vartheta}^k$ denote the minimizer obtained from the kth step objective function, and ε_1 and ε_2 be the perturbation constants defined in the MM algorithms of quantile regression (Hunter and Lange (2000)) and variable selection (Hunter and Li (2005)), respectively. Take

$$Q_\varepsilon(\bar{\vartheta}|\bar{\vartheta}^k) = \sum_{i=1}^{n}\int_0^{t_0} \zeta_\tau^\varepsilon(\bar{\epsilon}_{ni}(t)|\bar{\epsilon}_{ni}^k(t))dN_i(t),\ \zeta_\tau^\varepsilon(\bar{\epsilon}_n|\bar{\epsilon}_n^k) = \frac{1}{4}\Big[\frac{(\bar{\epsilon}_n)^2}{\varepsilon + |\bar{\epsilon}_n^k|} + (4\tau-2)\bar{\epsilon}_n + c\Big]$$

and for a scaler b,

$$\Phi_{\varepsilon_2}(b|b^k) = p_{\lambda,\varepsilon_2}(|b^k|) + \frac{(b^2 - b^{k2})p'_\lambda(|b|+)}{2(\varepsilon_2 + |b^k|)},$$
$$p_{\lambda,\varepsilon_2}(|b^k|) = p_\lambda(|b^k|) - \varepsilon_2 \int_0^{|b^k|} \frac{p'_\lambda(t)}{\varepsilon_2 + t}dt.$$

At the $(k+1)$th iteration, we consider the surrogate objective function

$$\begin{aligned}&\tilde{\Psi}_n^{(k+1),P}(\bar{\vartheta}, \widehat{\gamma}, \widehat{\Lambda}_0|\bar{\vartheta}^k)\\ &\quad = Q_{\varepsilon_1}(\bar{\vartheta}|\bar{\vartheta}^k) - \frac{1}{n}\sum_{i=1}^{n}\int_0^{t_0} \tau\bar{\epsilon}_{ni}(t)d\widehat{M}_i(t) + \sum_{j=1}^{p}\Phi_{\varepsilon_2}\big(\|\bar{\vartheta}_{j*}\| \,\big|\, \|\bar{\vartheta}_{j*}^k\|\big). \qquad (5.3)\end{aligned}$$

This function is smooth and can be minimized by using Newton's method. This, one can obtain the penalized estimator and select the model structure at the same time.

To implement this procedure, one needs to choose the number of interior knots k_n and the tuning parameter λ. We suggest two-step procedure based on the BIC. For fixed k_n, we take λ_{k_n} to be the minimizer of

$$\mathrm{BIC}_1(\lambda) = \log\left(\sum_{i=1}^{n}\int_0^{t_0}\left\{\rho_\tau(\widehat{\widehat{\epsilon}}_{ni}(t,k_n))dN_i(t) - \tau\widehat{\widehat{\epsilon}}_{ni}(t,k_n)d\widehat{M}_i(t)\right\}\right) + \frac{\log(n/k_n)}{(n/k_n)}V_\lambda + \frac{\log n}{n}C_\lambda.$$

Here, $\widehat{\widehat{\epsilon}}_{ni}(t,k_n) = Y_i(t) - \widehat{\widehat{\vartheta}}^T(k_n)\mathcal{X}_i(t)$, with $\widehat{\widehat{\vartheta}}(k_n)$ denoting the minimizer of (10) based on given k_n, while V_λ and C_λ are, respectively, the numbers of time-varying coefficients and constant coefficient selected by minimizing the penalized loss function with tuning parameter λ. Once λ_{k_n} is obtained, we choose k_n as the minimizer of

$$\mathrm{BIC}_2(k) = \log\left(\sum_{i=1}^{n}\int_0^{t_0}\left\{\rho_\tau(\widehat{\widehat{\epsilon}}_{ni}(t,\lambda_k))dN_i(t) - \tau\widehat{\widehat{\epsilon}}_{ni}(t,\lambda_k)d\widehat{M}_i(t)\right\}\right) + \frac{\log n}{n}\{V_{\lambda_k}(k+l+1) + C_{\lambda_k}\}.$$

Here $\widehat{\widehat{\epsilon}}_{ni}(t,\lambda_k) = Y_i(t) - \widehat{\widehat{\vartheta}}^T(\lambda_k)\mathcal{X}_i(t)$ with $\widehat{\widehat{\vartheta}}(\lambda_k)$ denoting the penalized quantile regression estimator obtained by minimizing penalized objective function (5.2) with the tuning parameter λ_{k_n}, and V_λ and C_λ are as above corresponding to λ_{k_n}. For the tuning parameter a, we use $a = 3.7$ following the suggestion of Fan and Li (2001).

6. A Simulation Study

In this section we present some results obtained from a simulation study conducted to evaluate the finite sample performance of the proposed procedures. We took the model

$$Y_i(t) = \alpha_1(t) + \alpha_2(t)X_{i1} + \beta_1 X_{i2} + \beta_2 X_{i3} + \varrho N_i(t-) + (1+\sigma X_{i1})\epsilon_i(t)$$

for the longitudinal response of interest, $i = 1, 2, \ldots, n$. We took $\alpha_1(t) = \sin(2t) + 1$, $\alpha_2(t) = t^2 - t + 1$, $\beta_1 = -1$, $\beta_2 = 1$, and $\varrho = 1.5$, and generated X_{i1}, X_{i2}, and X_{i3} from the standard normal distribution, the uniform distribution over interval (0,1), and the Bernoulli distribution with success probability 0.5, respectively. Then, the τth conditional quantile of $Y_i(t)$ is

$$Q_\tau(Y_i(t)|X_i, N_i(t-)) = \alpha_1(\tau,t) + \alpha_2(\tau,t)X_{i1} + \beta_1(\tau)X_{i2} + \beta_2(\tau)X_{i3} + \varrho(\tau)N_i(t-),$$

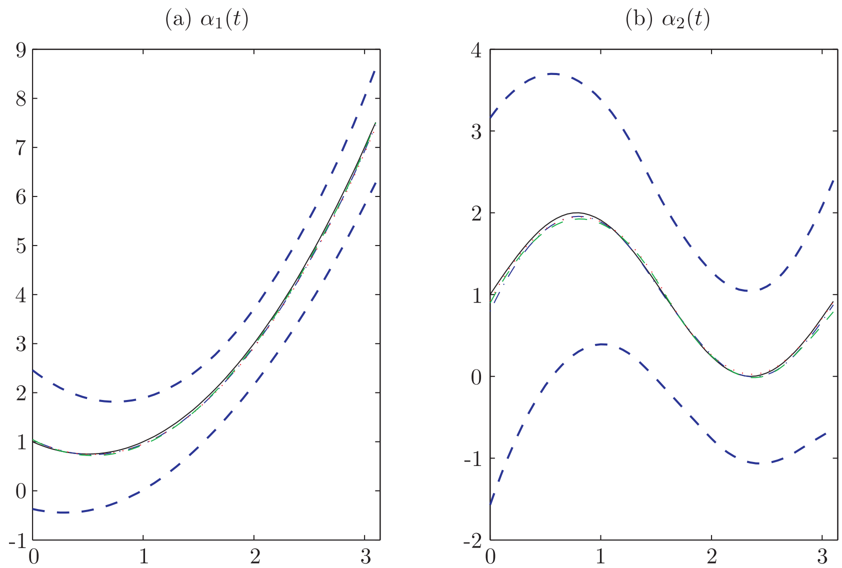

Figure 1. (**Case I**) The estimated coefficient functions for three quantiles: τ=0.25 (red dotted curve), τ=0.50 (blue dashed dotted curve) and τ=0.75 (green dashed curve). The black solid curve represents the true curve. The blue dash curves are 95% point-wise confidence bands.

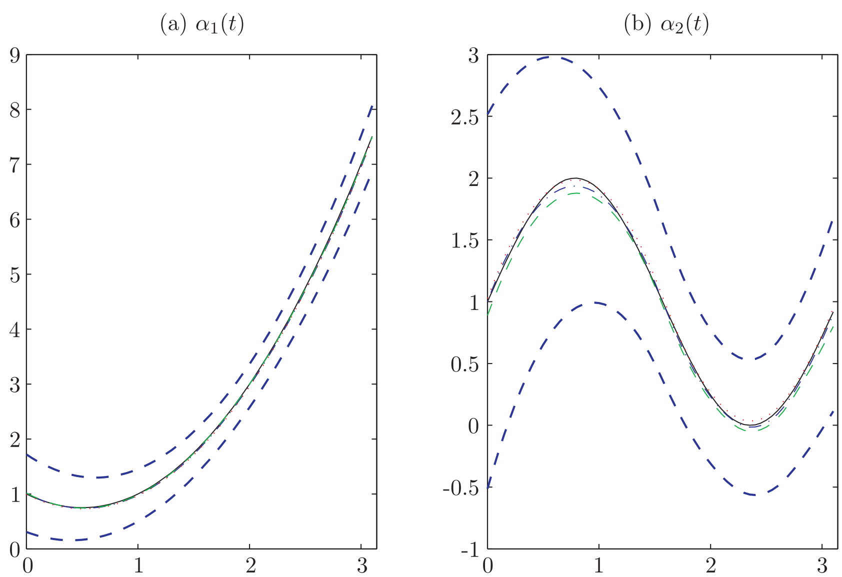

Figure 2. (**Case II**) The estimated coefficient functions for three quantiles: τ=0.25 (red dotted curve), τ=0.50 (blue dashed dotted curve) and τ=0.75 (green dashed curve). The black solid curve represents the true curve. The blue dash curves are 95% point-wise confidence bands.

Table 1. Estimation results on time-independent effects.

		$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$		
Case		$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$
I	Bias	-0.0028	0.0285	0.0103	0.0253	0.0220	0.0126	0.0248	0.0370	0.0169
	SD	0.5317	0.2902	0.0661	0.4760	0.2691	0.0606	0.5207	0.2916	0.0724
	SE	0.5229	0.3019	0.0623	0.4846	0.2802	0.0611	0.4992	0.2958	0.0670
	CP	0.9300	0.9580	0.9300	0.9420	0.9520	0.9380	0.9360	0.9340	0.9260
II	Bias	0.0007	0.0121	0.0050	0.0150	0.0120	0.0078	0.0104	0.0267	0.0100
	SD	0.2480	0.1315	0.0368	0.2202	0.1221	0.0340	0.2493	0.1395	0.0408
	SE	0.2689	0.1538	0.0356	0.2468	0.1407	0.0347	0.2608	0.1520	0.0388
	CP	0.9580	0.9700	0.9440	0.9760	0.9660	0.9380	0.9640	0.9620	0.9360

Table 2. The frequency table for the selection of time-varying coefficients.

			Case I					Case II				
n	k_n	τ	α_1	α_2	β_1	β_2	ϱ	α_1	α_2	β_1	β_2	ϱ
50	2	0.25	500	500	0	0	135	500	500	0	0	101
		0.5	500	500	0	0	120	500	500	0	0	122
		0.75	500	500	0	3	103	500	500	0	0	95
	4	0.25	500	500	0	3	139	500	500	0	0	113
		0.5	500	500	1	2	119	500	500	0	0	124
		0.75	500	500	0	11	109	500	500	0	2	91
100	2	0.25	500	500	0	0	13	500	500	0	0	4
		0.5	500	500	0	0	10	500	500	0	0	7
		0.75	500	500	0	3	20	500	500	0	0	9
	4	0.25	500	500	0	0	18	500	500	0	0	5
		0.5	500	500	0	0	21	500	500	0	0	11
		0.75	500	500	0	0	21	500	500	0	0	13

where $\alpha_1(\tau, t) = \alpha_1(t) + Q(\tau)$, $\alpha_2(\tau, t) = \alpha_2(t) + \sigma Q(\tau)$, $\beta_1(\tau) = \beta_1$, $\beta_2(\tau) = \beta_2$, and $\varrho(\tau) = \varrho$, and $Q(\tau)$ denotes the τth quantile of ϵ_i.

For the random error terms ϵ_i's, we considered Cases I and II. In Case I, they were the standard normal distribution with setting $\sigma = 0$, in Case II, we took $\sigma = 1$ and generated them from $N(0, 0.25)$. With respect to the observation process, model (2.2) was set to have the form

$$E\{\, dN_i^*(t)|X_i\,\} = \lambda_0(t)\, e^{\gamma_1 X_{i1} + \gamma_2 X_{i2} + \gamma_3 X_{i3}}\, dt$$

over the interval $[0, \pi]$ with $\gamma_1 = 0.5$, $\gamma_2 = -0.25$, $\gamma_3 = 1$, and $\lambda_0(t) = 2t$. For the follow-up time, we took $C_i = \pi \approx 3.14$ and used 500 resamples for the variance estimation. The results given are based on 500 replication.

Table 1 presents the results obtained based on the simulated data on estimation of three constant regression coefficients with $n = 100$, for both random error cases, and at the 0.25th, 0.5th, and 0.75th quantiles, respectively. Our focus

Table 3. IMSE and MSE of penalized estimators and oracle estimators for Case I.

		Penalized estimator					Oracle estimator				
n	τ	α_1	α_2	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$	α_1	α_2	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$
		$k_n = 2$									
	0.25	0.4078	0.4250	0.5358	0.1703	0.0453	0.3994	0.4229	0.5367	0.1704	0.0070
50	0.5	0.2800	0.4327	0.4321	0.1430	0.0562	0.2726	0.4239	0.4302	0.1416	0.0062
	0.75	0.3694	0.8284	0.4372	0.1632	0.2058	0.3672	0.8204	0.4378	0.1618	0.0081
	0.25	0.1404	0.2164	0.2021	0.0712	0.0064	0.1395	0.2160	0.2027	0.0714	0.0038
100	0.5	0.1319	0.2077	0.1889	0.0663	0.0053	0.1316	0.2078	0.1883	0.0663	0.0038
	0.75	0.1583	0.3305	0.2272	0.0762	0.0100	0.1581	0.3294	0.2272	0.0763	0.0053
		$k_n = 4$									
	0.25	0.4330	0.6023	0.5154	0.1651	0.1404	0.4184	0.5905	0.5173	0.1616	0.0069
50	0.5	0.3236	0.6896	0.4291	0.1409	0.1752	0.3107	0.6780	0.4268	0.1399	0.0063
	0.75	0.7519	2.0085	0.4268	0.1858	1.3341	0.5770	1.6219	0.4198	0.1592	0.0086
	0.25	0.1516	0.2456	0.2004	0.0696	0.0068	0.1504	0.2442	0.2012	0.0693	0.0037
100	0.5	0.1449	0.2498	0.1902	0.0656	0.0155	0.1440	0.2509	0.1898	0.0655	0.0037
	0.75	0.1817	0.4669	0.2261	0.0766	0.0223	0.1810	0.4650	0.2270	0.0765	0.0052

Table 4. IMSE and MSE of penalized estimators and oracle estimators for Case II.

		Penalized estimator					Oracle estimator				
n	τ	α_1	α_2	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$	α_1	α_2	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\varrho}$
		$k_n = 2$									
	0.25	0.0864	0.4506	0.1464	0.0421	0.0103	0.0846	0.4485	0.1481	0.0424	0.0023
50	0.5	0.0708	0.1443	0.1025	0.0342	0.0158	0.0692	0.1472	0.1022	0.0345	0.0021
	0.75	0.1168	0.6097	0.1185	0.0429	0.0245	0.1035	0.5995	0.1177	0.0425	0.0029
	0.25	0.0355	0.4191	0.0500	0.0160	0.0013	0.0355	0.4181	0.0502	0.0161	0.0011
100	0.5	0.0316	0.0666	0.0418	0.0142	0.0013	0.0315	0.0675	0.0419	0.0141	0.0011
	0.75	0.0421	0.4628	0.0550	0.0175	0.0027	0.0416	0.4622	0.0548	0.0176	0.0017
		$k_n = 4$									
	0.25	0.0951	0.5317	0.1436	0.0413	0.0308	0.0913	0.5285	0.1434	0.0411	0.0023
50	0.5	0.0835	0.2037	0.1017	0.0345	0.0554	0.0792	0.1994	0.1016	0.0343	0.0022
	0.75	0.2223	0.9308	0.1174	0.0428	1.1055	0.2071	0.9236	0.1177	0.0425	0.0031
	0.25	0.0380	0.4491	0.0486	0.0156	0.0016	0.0380	0.4498	0.0487	0.0156	0.0011
100	0.5	0.0346	0.0712	0.0425	0.0138	0.0021	0.0344	0.0717	0.0425	0.0138	0.0011
	0.75	0.0486	0.4767	0.0544	0.0172	0.0056	0.0482	0.4760	0.0546	0.0172	0.0017

here is to evaluate the performance of the proposed estimation procedure and for this, we assumed that the true model structure is known. Here we used the cubic B-spline and the number of knots was $k_n = 4$, the largest integer smaller than $n^{1/3}$ with $n = 100$. In addition, the 0.2th, 0.4th, 0.6th and 0.8th quantiles of the observation times were used as the knots. The results in the table include the averages of the estimated biases, the sample standard deviations of the estimates (SD), the averages of the estimated standard errors (SE), and the 95%

empirical coverage probabilities (CP). The corresponding estimated time-varying coefficients for Cases I and II are given in Figures 1 and 2, respectively, along with their 95% point-wise confidence bands.

In each part of Figures 1 and 2, there are three curves or estimates given by $\alpha_1(t) = \widehat{\alpha}_1(\tau, t) - Q(\tau)$ or $\alpha_2(t) = \widehat{\alpha}_2(\tau, t) - \sigma Q(\tau)$ for three different τ values. Thus, three estimates of the same function rather than estimates of three different functions $\widehat{\alpha}_1(\tau, t)$ or $\widehat{\alpha}_2(\tau, t)$. These results indicate that the proposed estimation procedure appear to work reasonably well. This is especially the case for the time-independent regression parameters as the estimators seem to be unbiased and the variance estimation is close to the sample variance. Table 1 also suggests that the covariate effects at the median can be more readily estimated than those at the 0.25th and 0.75th quantiles.

We considered the performance of the model selection procedure given in Section 5. For this, we assumed that the true model structure is unknown and considered two criteria: the number of times a coefficient is selected to be time-varying, and the integrated mean squared error (IMSE) for the penalized estimators of time-varying coefficients or the mean squared error (MSE) for the penalized estimators of constant coefficients. The IMSE of $\widehat{\alpha}_{nj}(t)$ is taken as

$$\mathrm{IMSE}\{\widehat{\alpha}_{nj}(t)\} = \frac{1}{100} \sum_{k=1}^{100} \{\, \widehat{\alpha}_{nj}(t_k) - \alpha_j(t_k)\, \}^2,$$

where $t_1 < \cdots < t_{100}$ are equally spaced time points over $[0, 3.14]$, $j = 1, 2$.

Table 2 gives the frequencies that each of the five coefficients was selected to be time-varying, the results on the IMSE or MSE are in Tables 3 and 4. Here we considered both random error cases with $n = 50$ and 100. To save the computer burden, we took $k_n = 2$ or 4 and given k_n, the parameter λ was selected by using the BIC criterion given in Section 5. We tried several other values for k_n and obtained similar results. In Tables 3 and 4, for comparison, we calculated the IMSE or MSE for the estimators given by the estimation procedure proposed in Section 3 assuming that the true model structure is known, referred to as the oracle estimator. One can see from Table 2 that the selection is almost always correct for the first four parameters even with $n = 50$ and the selection for the last parameter dramatically improved from $n = 50$ to $n = 100$.

With respect to the IMSE and MSE, they decrease as the sample size increases, as expected, and grow closer.

7. An Application

We applied the proposed methodology to the monthly medical cost data discussed earlier. The study involved 1,475 patients whose age were 60 years or

Table 5. Estimation of time-independent effects for medical cost data.

	$\tau = 0.25$			$\tau = 0.5$		
	age	gender	race	age	gender	race
Estimate	-0.3474	0.1424	-0.3228	-0.5272	0.2382	-0.4220
SE	0.0089	0.1106	0.1354	0.0098	0.1336	0.1560
	$\tau = 0.75$			observation process		
	age	gender	race	age	gender	race
Estimate	-1.5100	0.1696	-1.3145	-0.0030	0.0411	0.0072
SE	0.0228	0.3483	0.6047	0.0026	0.0408	0.0445

above and who had the first diagnosed heart failure in 2004. For each patient, the observed information includes clinical visit or observation time point (in months) and the corresponding monthly medical cost as well as the baseline covariates of gender, race, and age. The follow-up time was either July 31, 2006, the end of the study, or their death. The median of the medical cost was \$350, while the mean was \$2670. Cost was highly skewed to the right. As discussed earlier, it appears that the observation process contains relevant information about the cost (Liu, Huang, and O'Quigley (2008); Sun et al. (2012)). The main objective here is to estimate the trajectory of the medial cost and its relationship with the three baseline covariates.

Let $Y_i(t)$ denote the cubic root of the medical cost at month t for patient i, $i = 1, \ldots, 1{,}475$. We used the cubic root is used here to avoid large response values but still keep the skewness of the data. For patient i, let $X_{0i} = 1$ for the intercept term, X_{1i} be centered age, $X_{2i} = 1$ if the patient is male and 0 otherwise, and $X_{3i} = 1$ if the patient is white and 0 otherwise. We assume that $Y_i(t)$ can be described by model (2.1) with $H(\mathcal{F}_{it}) = N_i(t-)$. Thus, medical cost depends on the observation process through the total number of the medical visits. In the analysis, we assumed that all coefficients in model (2.1) are time-varying coefficients and applied the model selection procedure of Section 6 with k_n and the penalized tuning parameter λ selected by the BIC procedures. This means the terms corresponding to the intercept X_{0i} and the observation process $N_i(t-)$ with time-varying effects, while the three baseline covariates have constant effects.

Table 5 presents the estimates of three constant coefficients; the results indicate that two of the three baseline covariates, age and race, had significant effects on the monthly medical cost. In particular, the cost tends to get smaller as the patient gets older, which may be because their treatments are less aggressive. The difference between genders is not significant. The three covariates did not have significant effects on the observation process.

Figure 3 gives the estimates of the two time-varying coefficients at the quantiles $\tau = 0.25$, 0.5, and 0.75 and their 95% point-wise confidence bands, the three estimates have similar shapes for the effects corresponding to the intercept and

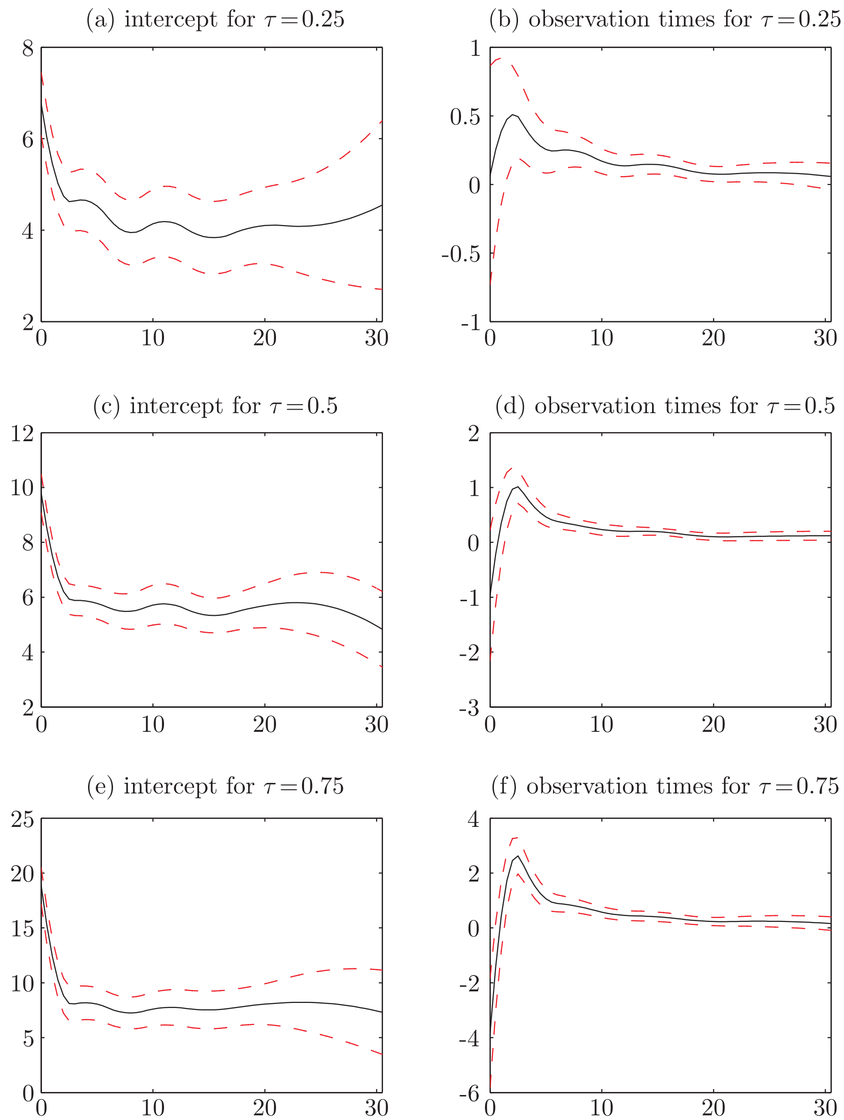

Figure 3. The estimated coefficient functions (solid curves) and their 95% point-wise confidence bands(dash curves) for the medical cost data and at three quantiles: τ=0.25 , τ=0.5 and τ=0.75.

the observation process. Parts (a), (c) and (e) of the figure indicate that given other factors, medical cost starts high, and then decreases, staying stable after about 15 months. This is to be expected. From the observation process point of view, parts (b), (d) and (f) of the figure tell us that the relationship between medical cost and the process is complicated at the beginning and, as expected, higher numbers of visits means higher costs. After about 5 to 15 months, the effect of the number of medical visits seems to decrease to nonsignificance.

To check the adequacy of model (2.1) for the problem here, we applied the model checking procedure described in Section 5 and obtained the supremum test statistic $\sup_{0\leq u\leq t_0,x_1,z}|\mathcal{L}_\tau(u,x_1,z)| = 11.4651, 20.2721, 28.7268$ at $\tau = 0.25, 0.5, 0.75$, respectively. These correspond to p-values of 0.4103, 0.4906 and 0.4472 and indicate that the model (2.1) fits the data well.

8. Concluding Remarks

The key difference between the estimation procedure given in Section 3 and the model selection procedure of Section 5 is the use of the SCAD penalty function in the latter. In practice, if the covariates with time-varying effects are known, one can directly apply the estimation procedure, otherwise, one may want to employ the model selection procedure first. One could apply such other penalty functions as the LASSO (Tibshirani (1996, 1997)) or SELO (Dicker, Huang, and Lin (2012)) and develop the corresponding model selection procedure.

Model (2.1) is a conditional model with respect to the observation process; alternatively one could model the longitudinal process $Y_i(t)$ and the observation process $N_i^*(t)$ jointly through the use of some latent variables or processes (Sun, Sun, and Zhou (2013)). For the observation process model (2.2), it was supposed that covariates have only constant effects. One might also allow some covariates to have time-varying effects and it is apparent that, in this case, a different method is needed for estimation of the model.

In model (2.1), we have assumed that the effect of observation processes is time-independent. For this situation, as pointed out by a reviewer, a question of interest is what effect one would expect to see on estimation of other parameters by treating $\rho(t)$ as time-independent. We conducted a simulation study and the numerical results suggest that the effect depends on the shape of $\rho(t)$. In general, it tends to reduce the variances of the estimators but increase the biases.

Another assumption is that the follow-up time C_i is independent of both $Y_i(t)$ and $N_i^*(t)$ given covariates. Thus we say C_i is informative about $Y_i(t)$ or $N_i^*(t)$ and one possibility is that C_i is generated by a dependent terminal event such as death. A large literature exists for the situation where C_i is informative about $Y_i(t)$ in the context of longitudinal data analysis or about $N_i^*(t)$ in the context of recurrent event data analysis. A common approach then is joint modeling. There seems to be no established method in the context of quantile regression.

Acknowledgement

We wish to thank the Editor Qiwei Yao, associate editor and two reviewers for their many helpful and insightful comments and suggestions that greatly improved the paper. Liu's research is partly supported by AHRQ grant R01

HS020263. Chen's research is supported by Fundamental Research Funds for the Central Universities JBK140507 and JBK141111.

References

Chen, L. Wei, L. and Parzen, M. (2004). Quantile regression for correlated observations. *Proceedings of the Second Seattle Symposium in Biostatistics,* 51-69 *Lecture Notes in Statistics*, 179.

Cheng, G. and Huang, J. Z. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38**, 2884-2915.

Cheng, S. C., Wei, L. J. and Ying, Z. (1997). Predicting survival probabilities with semiparametric transformation models. *J. Amer. Statist. Assoc.* **92**, 227-235.

Cook, R., and Lawless, J. (2007). *The Statistical Inference of Recurrent Events.* Springer Science Press, New York.

Dicker, L., Huang, B. and Lin, X. (2012). Variable selection and estimation with the seamless-L_0 penalty. *Statist. Sinica* **23**, 929-962.

Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2013). *Analysis of Longitudinal Data.* 2nd edition. Oxford University Press, New York.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *Ann. Statist.* **7**, 1-26.

Efron, B. (1981). Censored data and the bootstrap. *J. Amer. Statist. Assoc.* **72**, 316-319.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its Oracle properties. *J. Amer. Statist. Assoc.* **96**, 1348-1360.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *J. Amer. Statist. Assoc.* **99**, 710-723.

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *J. Amer. Statist. Assoc.* **102**, 632-641.

Fan, J., and Wu, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data. *J. Amer. Statist. Assoc.* **103**, 1520-1533.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153-193.

Huang, J., Wei, F. and Ma, S. (2012). Semiparametric regression pursuit. *Statist. Sinica* **22**, 1403-1426.

Hunter, D. R. and Li, R. (2005). Variable selection using MM algorithms. *Ann. Statist.* **33**, 1617-1642.

Hunter, D. R. and Lange, K. (2000). Quantile regression via an MM algorithm. *J. Comput. Graph. Statist.* **9**, 60-77.

Kai, B., Li, R. and Zou, H. (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Statist.* **39**, 305-332.

Kim, M. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35**, 92-108.

Koenker, R. and Bassett, G. S. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

Liang, Y., Lu, W. and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. *Biometrics* **65**, 377-384.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96**, 103-126.

Liu, L., Huang, X. and O'Quigley, J. (2008). Analysis of longitudinal data in the presence of informative observational times and a dependent terminal event, with application to medical cost data. *Biometrics* **64**, 950-958.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika* **86**, 691-702.

Martinussen, T. and Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitudinal data. *Ann. Statist.* **28**, 1000-1025.

Martinussen, T. and Scheike, T. H. (2001). Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scand. J. Statist.* **28**, 303-323.

Schumarker, L. L. (1981). *Spline Functions: Basic Theory.* Wiley, New York.

Stone, C. (1980). Optimal rate of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.

Sun, L., Song, X., Zhou, J. and Liu, L. (2012). Joint analysis of longitudinal data with informative observation times and dependent terminal event. *J. Amer. Statist. Assoc.* **107**, 688-700.

Sun, J., Park, D. Sun, L. and Zhao, X. (2005). Semiparametric regression analysis of longitudinal data with informative observation times. *J. Amer. Statist. Assoc.* **100**, 882-889.

Sun, J., Sun, L. and Liu, D. (2007). Regression analysis of longitudinal data in the presence of informative observation and censoring times. *J. Amer. Statist. Assoc.* **102**, 1397-1406.

Sun, Y., Sun, L. and Zhou, J. (2013). Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime Data Anal.* **19**, 317-349.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scand. J. Statist.* **32**, 21-47.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. Ser. B* **58**, 267-288.

Tibshirani, R. (1997). The Lasso method for variable selection in the Cox model. *Stat. Med.* **16**, 385-395.

Wang, H. X., Zhu, Z. Y. and Zhou, J. H. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37**, 3841-3866.

Wei, Y. and He, X. (2006). Conditional growth charts. *Ann. Statist.* **34**, 2069-2097.

Zhang, H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? Automatic structure discovery for partially linear models. *J. Amer. Statist. Assoc.* **106**, 1099-1112.

School of Statistics, Southwestern University of Finance and Economics, Chengdu, Sichuan, China.

E-mail: chenxr522@foxmail.com

Department of Statistics, University of Missouri, Columbia, USA.

E-mail: sunj@missouri.edu

Department of Preventive Medicine and Robert H. Lurie Cancer Center, Northwestern University, Chicago, USA.

E-mail: lei.liu@northwestern.edu

(Received October 2013; accepted September 2014)