Structure Identification in Varying-Coefficient Partially Linear Accelerated Failure Time Models via the Group MCP

Yue Fu [1, a], Kaifeng Hu [2, b, *]

[1] Department of Statistics, School of Economics, Jinan University, Guangzhou, 510632, China

[2] China People's Police University, Guangzhou, 510663, China

[a] fuyue@stu2020.jnu.edu.cn, [*, b] Corresponding Author Email: hukaifeng@cppu.edu.cn

Abstract. The accelerated failure time (AFT) model is commonly used for the analysis of survival data in the presence of right censored due to the interpretability. In some practical cases, especially when some covariates are time-related, it is not realistic to assume the linear predictors in the AFT model. We propose a varying-coefficient partially linear AFT model for right censored data, allowing the nonlinear effects of covariates. To tackle challenges in estimation, we propose a penalized profile likelihood approach which utilizes a group minimax concave penalty to determine the nonlinear effects of covariates. Under the suitable conditions, we show that the proposed method can correctly select linear components and nonlinear components with high probability. Simulation results demonstrate the satisfactory performance of the proposed method in finite sample cases.

Keywords: AFT model; B-spline; group MCP.

1. Introduction

Let T_i be logarithm of the failure time for individual i, u_i is a one-dimensional covariate and $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$is a p-dimensional covariate vector. Consider the varying-coefficient partially linear AFT model:

$$T_i = \mu + \left\{\sum_{j\in S_1} \beta_{0j} x_{ij} + \sum_{j\in S_2} \beta_{1j} u_i x_{ij} + \sum_{j\in S_3} g_j(u_i) x_{ij}\right\} + \varepsilon_i, 1 \le i \le n \qquad (1)$$

where μ is the intercept, S_1,S_2 and $S_3$are mutually exclusive and complementary subsets of $\{1,\dots,\mathrm{p}\}$, $\{\beta_{0j}: j \in S_1\}$are constant regression coefficients of the covariates with indices in S_1, $\{\beta_{1j}: j \in S_2\}$ are linear regression coefficients of the covariates with indices in S_2, and $\{g_j(u): j \in S_3\}$ are unknown functions. In this model, the mean response is linearly related to the covariates in $S_1$and S_2, while its relation with the remaining covariates is not specified up to any finite number of parameters. The model combines the flexibility of varying-coefficient regression and the parsimony of linear regression. When the relationship between T_i and $\{x_{ij}: j \in S_1\}$ or $\{x_{ij} u_i: j \in S_2\}$ is of main interest and can be approximated by a linear function, it offers more interpretability than a purely varying-coefficient model. When T_i is subject to right censoring, the observed data is (Y_i, δ_i, X_i), i = 1, 2, . . . n, where $Y_i = min\{T_i, C_i\}$, $\mathrm{C_i}$ is logarithm of the censoring time, and $\delta_i = 1_{\{T_i \le C_i\}}$ is the censoring indicator.

The AFT model was first proposed by Pieruschka[10] in 1961, which specifies the effect of a fixed covariate is to act multiplicatively on the failure time. The AFT model has received considerable attention in the statistical literature over the past decade or so. Buckley and James(1979)[1] proposed the Buckley-James estimator which adjusts censored observations using the Kaplan-Meier estimator. Ying(1993)[17] proposed the rank based estimator motivated by the score function of the partial likelihood function. Zhou(1992)[19] and Stute(1993)[13] calculated censored data with Kaplan-Meier weights and obtained the weighted least squares loss function.

Many punishment methods have been used to calculate the least squares loss function of the AFT model, within which the most commonly used punishment method is LASSO. Gui and Li(2005)[3], Ma and Huang(2007)[8] and Wang(2008)[14] introduced LASSO into the survival analysis to analyse gene expression under survival data. Huang and Ma(2010)[4] used the bridge penalization for regularized estimation and gene selection. Park and Ha(2019)[9] used h-likelihood penalty for variable selection in the AFT models. Most of the existing studies considered the AFT model with fixed coefficients, which unable to fully explain the relationship between covariates and response variables when the dynamic feature presents. To incorporate the dynamical pattern and the nonlinear effects of the covariates, the varying-coefficient partially linear AFT model is utilized.

The varying-coefficient partially linear model not only has the advantage of the interpretability of the parametric model, but also has the flexibility of the nonparametric model. It has been widely used in many fields. Lam and Fan(2008)[6] proposed the accelerated profilekernel algorithm for computing profile likelihood estimates under the generalized varying coefficient partially linear model. For the di-verging number of parameters in a varying coefficient partially linear model, Li et al. (2012) [7] adopted a bias-corrected empirical likelihood (BCEL) to avoid such strong structure assumptions on the model and inconvenience of estimation implementation. Wei et al.(2015)[15] constructed an efficient estimator for the parametric component of the varying coefficient partially linear model by applying the weighted profile least-squares approach, without the exploration of variable selection. Recently, many scholars have proposed variable selection methods that use penalty functions in non-parametric or semiparametric settings. Huang et al. (2012) [5] proposed a semiparametric regression tracking method to identify covariates with linear effects by using group MCP penalties in semiparametric partial linear models. Yang et al.(2017)[16] proposed a profle approach to identify the covariates with linear effect or nonlinear effect by using a group MCP in a varying-coefcient partially linear models model.

In this article, we consider the varying-coefficient partially linear AFT model and screen out linear components and nonlinear components. We approximate the nonparametric components by spline functions and determine the covariates that has linear/nonlinear effects via the penalty method. We derive a group coordinate descent algorithm for calculating the group MCP estimators. It is proved that it can correctly determine linear components and nonlinear components with high probability.

This article is organized as follows. In Section 2, we describe the process of selecting and estimating based on the basis expansion for the varying-coefficient partially linear AFT model using a group MCP. Section 3 presents the results on the selection consistency of the group MCP. Section 4 provides a simulation study to illustrate the proposed methods. Concluding remarks are contained in section 5.

2. Method

The varying-coefficient partially linear AFT model (1.1) can be embedded into the varying-coefficient model,

$$T_i = \mu + \alpha_1(u_i)x_{i1} + \ldots + \alpha_p(u_i)x_{ip} + \varepsilon_i \tag{2}$$

where $\alpha_j(u_i)$, are functions of u_i. Suppose that $u_1, \ldots, u_n$are from a distribution with support [a, b], where a<b are finite constants. If some of the $\alpha_j(u)$'s are linear, then (2.1) becomes the varying-coefficient partially linear AFT model (1.1). The problem is that of determining which $\alpha_j(u)$'s have a linear forms and which do not.For this purpose, we write $\alpha_j(u)$ as

$$\alpha_j(u) = \beta_{0j} + \beta_{1j}u + g_j(u) \tag{3}$$

Consider a truncated series expansion for approximating $g_j(u)$,

$$g_{nj}(u) = \sum_{k=1}^{m_n} \theta_{jk}\psi_k(u), 1 \le j \le p \tag{4}$$

where $\psi_1(u), \ldots, \psi_{m_n}(u)$ are basis functions, m_n is allowed to increase with n and here we can take the B-spline basis functions. According to the classical book of Schumaker (1981)[11], the basis satisfies (i) $\psi_k(u) \ge 0, k = 1, \ldots, m_n$, (ii) $\sum_{k=1}^{m_n} \psi_k(u) \equiv 1$. The varying-coefficient partially linear AFT model (1.1) can be written as

$$T_i = \mu + \left\{\sum_{j\in S_1} \beta_{0j}x_{ij} + \sum_{j\in S_2} \beta_{1j}u_i x_{ij} + \sum_{j\in S_3} \sum_{k=1}^{m_n} \theta_{jk}\psi_k(u_i)x_{ij}\right\} + \varepsilon_i \tag{5}$$

furthermore, there doesn't exist identification problem when the nonparametric function is expanded by using B-splines. If $\theta_{jk}=0, 1\le k\le m_n$, then $\alpha_j(u)$'s is linear structure. Therefore, with this formulation, the problem is to determine which groups $\{\theta_{jk};\ 1\le k\le m_n\}$, $j\in S_3$ are zero.

We calculated the censored data with Kaplan-Meier weights w_is, which are computed as $w_1=\frac{\delta_{(1)}}{n}, w_i=\frac{\delta_{(i)}}{n-i+1}\prod_{j=1}^{i-1}(\frac{n-j}{n-j+1})^{\delta_{(i)}}, 1\le i\le n$. Here $Y_{(1)}\le,\dots,\le Y_{(n)}$ are the order statistics of Y_i's, $\delta_{(1)},\dots,\delta_{(n)}$ are the associated censoring indicators, and $X_{(1)},\dots,X_{(n)}$ are the associated covariates. The loss function is

$$\frac{1}{2}\sum_{i=1}^{n} w_i\left\{Y_{(i)}-\mu-\sum_{j\in S_1}\beta_{0j}x_{(i)j}-\sum_{j\in S_2}\beta_{1j}u_{(i)}x_{(i)j}-\sum_{j\in S_3}\sum_{k=1}^{m_n}\theta_{jk}\psi_k\left(u_{(i)}\right)x_{(i)j}\right\}^2 \tag{6}$$

We center the responses and covariates. Let $\bar{Y}=\frac{\sum_{i=1}^{n}w_iY_{(i)}}{\sum_{i=1}^{n}w_i}, \bar{x}_j=\frac{\sum_{i=1}^{n}w_ix_{(i)j}}{\sum_{i=1}^{n}w_i}, \overline{ux_j}=\frac{\sum_{i=1}^{n}w_iu_{(i)}x_{(i)j}}{\sum_{i=1}^{n}w_i}, \overline{\psi_{jk}}=\frac{\sum_{i=1}^{n}w_i\psi_k(u_{(i)})x_{(i)j}}{\sum_{i=1}^{n}w_i}$, and $Y_{(i)}^*=Y_{(i)}-\bar{Y}, x_{(i)j}^*=x_{(i)j}-\bar{x}_j, u_{(i)}x_{(i)j}^*=u_ix_{(i)j}-\overline{ux_j}, \psi_{jk}(u_{(i)})x_{(i)j}{}^*=\psi_k(u_{(i)})x_{(i)j}-\overline{\psi_{jk}}, 1\le i\le n, 1\le j\le p, 1\le k\le m_n$.

For simplicity, let

$$x_j^*=\begin{pmatrix}\sqrt{w_1}x_{(1)j}^*\\ \sqrt{w_2}x_{(2)j}^*\\ \vdots\\ \sqrt{w_n}x_{(n)j}^*\end{pmatrix}, ux_j^*=\begin{pmatrix}\sqrt{w_1}u_{(1)}x_{(1)j}^*\\ \sqrt{w_2}u_{(2)}x_{(2)j}^*\\ \vdots\\ \sqrt{w_n}u_{(n)}x_{(n)j}^*\end{pmatrix}, X^*=\left(x_1^*,\dots,x_p^*,ux_1^*,\dots,ux_p^*\right)$$

$$Y^*=\begin{pmatrix}\sqrt{w_1}Y_{(1)}^*\\ \sqrt{w_2}Y_{(2)}^*\\ \vdots\\ \sqrt{w_n}Y_{(n)}^*\end{pmatrix}, z_{ij}^*=\sqrt{w_i}\begin{pmatrix}\psi_{j1}(u_{(i)})x_{(i)j}{}^*\\ \psi_{j2}(u_{(i)})x_{(i)j}{}^*\\ \vdots\\ \psi_{jm_n}(u_{(i)})x_{(i)j}{}^*\end{pmatrix}, z_j^*=\begin{pmatrix}z_{1j}^*\\ z_{2j}^*\\ \vdots\\ z_{nj}^*\end{pmatrix}, Z^*=\left(z_1^*,\dots,z_p^*\right)$$

$$\beta_0=\begin{pmatrix}\beta_{01}\\ \beta_{02}\\ \vdots\\ \beta_{0p}\end{pmatrix}, \beta_1=\begin{pmatrix}\beta_{11}\\ \beta_{12}\\ \vdots\\ \beta_{1p}\end{pmatrix}, \beta=(\beta_0^T,\beta_0^T)^T, \theta_j=\begin{pmatrix}\theta_{j1}\\ \theta_{j2}\\ \vdots\\ \theta_{jm_n}\end{pmatrix}, \theta=\left(\theta_1^T,\dots,\theta_p^T\right)^T$$

we can write the loss function as

$$L(\beta,\theta;\lambda,\gamma)=\frac{1}{2}\|Y^*-X^*\beta-Z^*\theta\|^2 \tag{7}$$

Define the penalized least squares loss function

$$L(\beta,\theta;\lambda,\gamma)=\frac{1}{2}\|Y^*-X^*\beta-Z^*\theta\|^2+\sum_{j=1}^{p}\rho\left(\|\theta_j\|_{A_j};\sqrt{m_n}\lambda,\gamma\right) \tag{8}$$

Where $\rho(\cdot;\cdot,\cdot)$is a penalty function depending on the penalty parameter $\lambda\ge 0$and a regularization parameter γ. The norm $\|\theta_j\|_{A_j}=\left(\theta_j^TA_j\theta_j\right)^{1/2}$ for a given positive definite matrix A_j. In theory, any positive definite matrix can be used as A_j, since $\|\theta_j\|_{A_j}=0$ if and only if $\theta_j=0$. However, it is important to choose a suitable $A_j$to make the amount of penalization comparable across the groups and to facilitate the computation. We use the minimax concave penalty (MCP) and detail discussions on the MCP can be found in Zhang (2010)[18].

The penalty in (2.7) is a composite of the penalty function $\rho(\cdot;\lambda,\gamma)$ and a weighted l_2-norm of θ_j. The $\rho(\cdot;\lambda,\gamma)$ is a penalty for individual variable selection. When applied to a norm of θ_j, it selects the coefficients in θ_j as a group. This is desirable, since the nonlinear components are represented by the coefficients in the θ_j's as groups. Based on (2.7), it is natural to call it the group minimax concave penalty (gMCP).

For a given(λ , γ), the penalized least squares solution is

$$(\hat{\beta}_n, \hat{\theta}_n) = arg \min_{\beta,\theta} L(\beta, \theta; \lambda, \gamma)$$

To compute $(\hat{\beta}_n, \hat{\theta}_n)$, we can use a penalized profile least squares approach. For any given θ_n, the $\hat{\beta}_n$that minimizes L necessarily satisfies

$$X^{*T}(Y^* - X^*\beta - Z^*\theta_n) = 0$$

Thus $\hat{\beta}_n = (X^{*T}X^*)X^{*T}(Y^* - Z^*\theta_n$. Let $Q = I - P_{X^*}$,where $P_{X^*} = X^*(X^{*T}X^*)^{-1}X^{*T}$ is the projection matrix onto the column space of X^*. The penalized objective function of θ_n is

$$L(\theta_n; \lambda, \gamma) = \frac{1}{2}\|Q(Y^* - Z^*\theta_n)\|^2 + \sum_{j=1}^{p} \rho\left(\|\theta_{nj}\|_{A_j}; \sqrt{m_n}\lambda, \gamma\right) \quad (9)$$

As noted above, any positive definite matrix can be used for A_j. Like Yang et al.(2017)[16], we use $A_j = n^{-1}U_j^T Q U_j$, where Q is an orthonormal matrix. Detailed discussion can be found in that article.

Here, we will follow Huang (2012)[5]. The set of indices of the covariates that are estimated to have the linear structure in the regression model (1.1) is $\widehat{S_1 \cup S_2} \equiv \left\{j: \|\hat{\theta}_{nj}\|_2 = 0\right\}$. Thus, $\hat{g}_{nj}(u) = 0, j \in \widehat{S_1 \cup S_2}, \hat{g}_{nj}(u) = \sum_{k=1}^{m_n} \hat{\theta}_{jk}\psi_{jk}(u), j \notin \widehat{S_1 \cup S_2}$ .Let $\hat{X}^*_{(1)} = \left((x_j^*, ux_j^*), j \in \widehat{S_1 \cup S_2}\right), \hat{Z}^*_{(2)} = \left(Z_j^*, j \notin \widehat{S_1 \cup S_2}\right), \hat{\theta}_{n(2)} = \left(\widehat{\theta}^{T}_{nj}, j \notin \widehat{S_1 \cup S_2}\right)^T$,we have $\hat{\beta}_n = (X^{*T}X^*)X^{*T}(Y^* - \hat{Z}^*_{(2)}\hat{\theta}_{n(2)})$ .The estimator of the coefficients of the linear components is $\hat{\beta}_{n1} = \left(\hat{\beta}_j: j \in \widehat{S_1 \cup S_2}\right)^T$. Let

$$\hat{\alpha}_{nj}(u) = (1, u)\hat{\beta}_j + \hat{g}_j(u), j \notin \widehat{S_1 \cup S_2}$$

Where $\hat{\beta}_j$ corresponds to the coefficients of the linear components. Let us write $\hat{\alpha}_n(u) = \left(\hat{\alpha}_{n1}(u), \ldots, \hat{\alpha}_{np}(u)\right)^T$. Then the estimator of the coefficient vector of the linear components can also be written as

$$\hat{\beta}_{n1} = \left(\hat{X}^{*T}_{(1)}\hat{X}^*_{(1)}\right)^{-1}\hat{X}^{*T}_{(1)}\left(Y^* - \sum_{j \notin \widehat{S_1 \cup S_2}} \hat{\alpha}_{nj}(u)x_j^*\right)$$

3. Theoretical properties

In this section, we established the model consistency and large sample properties of the proposed profile estimator of varying-coefficient partially linear AFT models. In particular, our model consistency result shows that the proposed method can correctly determine the linear and nonlinear components in the varying-coefficient partially linear AFT models with high probability.

Denote the underlying regression components by $\alpha_{0j}(u)$ and write

$$\alpha_{0j}(u) = \beta_{0j} + \beta_{1j}u + g_{0j}(u)$$

Suppose the series expansion for approximating $g_{0j}(u)$ is

$$g_{0j}(u) = \sum_{k=1}^{m_n} \theta_{0jk}\psi_k(u), 1 \le j \le p$$

Let $\theta_{0jn} = \left(\theta_{0j1}, .., \theta_{0jm_n}\right)^T$, and let $\|g\|_2 = [Eg^2(u)]^{1/2}$ for any square integrable function g on [a, b]. We have $S_1 \cup S_2 = \left\{j: \left\|g_{0j}(u)\right\|_2 = 0\right\}$ and $\| \theta_{0nj}\|_2 = 0$ for $j \in S_1 \cup S_2$. Let $\theta_{0n} = \left(\theta_{0n1}^T, .., \theta_{0np}^T\right)^T$.

Let $q = |S_1 \cup S_2|$ be the cardinality of $S_1 \cup S_2$, the number of linear components in the regression model. Take

$$\widetilde{\theta}_n = \arg\min_{\theta} \left\{\frac{1}{2}\|Q(Y^* - Z^*\theta_n)\|^2 : \theta_{nj} = 0, j \in S_1 \cup S_2\right\} \tag{10}$$

This is the oracle estimator of θ 0n that takes the identity of the linear components as known.

Analogous to the estimates defined at the end of Section 2, write the oracle estimators as

$$\tilde{g}_{nj}(u) = 0, j \in S_1 \cup S_2, \tilde{g}_{nj}(u) = \sum_{k=1}^{m_n} \tilde{\theta}_{jk}\psi_{jk}(u), j \notin S_1 \cup S_2$$

Let $X_{(1)}^* = \left(\left(x_j^*, ux_j^*\right), j \in S_1 \cup S_2\right), Z_{(2)}^* = \left(Z_j^*, j \in S_3\right), \tilde{\theta}_{n(2)} = \left(\tilde{\theta}_{nj}^T, j \in S_3\right)^T$, and

$\tilde{\alpha}_{nj}(u) = \tilde{\beta}_{0j} + \tilde{\beta}_{0j}u + \tilde{g}_{nj}(u), j \in S_3$. The oracle estimator of the coefficients of the linear components is

$$\tilde{\beta}_{n1} = \left(X_{(1)}^{*T}X_{(1)}^*\right)^{-1}X_{(1)}^{*T}\left(Y^* - \sum_{j \notin S_1 \cup S_2} \tilde{\alpha}_{nj}(u))x_j^*\right)$$

Without loss of generality, suppose that $S_1 \cup S_2 = \{1, \ldots, q\}$. Write$\tilde{\theta}_n = \left(0_{qm_n}^T, \tilde{\theta}_{n(2)}^T\right)^T$, where $0_{qm_n}$is a $(qm_n$)-dimensional vector of zeros and

$$\tilde{\theta}_{n(2)} = \left(Z_{(2)}^{*T}QZ_{(2)}^*\right)^{-1}Z_{(2)}^{*T}QY^* \tag{11}$$

where$Z_{(2)}^*$ is the design matrix related to $\theta_{n(2)}$. Let $\theta_* = \min_{j \in S_1 \cup S_2} \| \theta_{0nj}\|$, the smallest norm of the coefficients in the spline expansions of the nonlinear components.

Let k be a non-negative integer, and let $\alpha \in (0,1]$ be such that $d = k + \alpha > 0.5$. Let G be the class of functions g on [0, 1] whose kth derivative g(k) exists and satisfies a Lipschitz condition of order α:

$$\left|g^{(k)}(s) - g^{(k)}(t)\right| \leq C|s - t|^\alpha \quad s, t \in [a, b]$$

Then there exists a positive constant C_1 such that

$$\left\|g_{0j}(u) - g_{0n}(u)\right\| = O(m_n^{-2r-2\alpha})$$

We make the following assumptions.

(A1) The errors ($\varepsilon_1, \ldots, \varepsilon_n$) are independent of the Kaplan-Meier weights $(w_1, \ldots, w_n)$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent and identically distributed with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. Furthermore, $p(|\varepsilon_i| > x) \leq Kexp(-Cx^2)$, i = 1, . . . , n, for all $x \geq 0$ for some constants C and K.

(A2) p and q are fixed and the covariate vector u has a continuous density and there exist constants C_1 and C_2 such that the density function$\alpha_j(\cdot)$ of u satisfies $0 < C_1 \leq \alpha_j(u) \leq C_2 < \infty$, on [a, b] for every $1 \leq j \leq p$.

(A3) The observations $(Y_i, X_i, \delta_i), 1 \leq i \leq n$ are independent and identically distributed and there exists a positive constant M such that $\left|x_{ij}\right| \leq M, 1 \leq i \leq n, 1 \leq j \leq p$.

Theorem 3.1. Suppose that$m_n = O(n^{1/(2d+1)})$, $1/\sqrt{m_n}\gamma$ is less than the smallest eigenvalue of $Z^{*\prime}QZ^*/n$, and

$$\frac{1}{m_n^{(2d-1)/2}(\theta_* - \lambda\gamma)} + \frac{1}{\lambda\sqrt{n}} \to 0 \quad (12)$$

Then unnder (A1)—(A3), $P(\hat{\theta}_n \neq \tilde{\theta}_n) \to 0$. Consequently,$P(\widehat{S_1 \cup S_2} = S_1 \cup S_2) \to 1$, and

$P(\hat{\beta}_{n1} = \tilde{\beta}_{n1}) \to 1$, and $P(\|\hat{\alpha}_{nj}(u) - \tilde{\alpha}_{nj}(u)\|_2 = 0, j \in S_3) \to 1$.

Therefore, under the conditions of Theorem 3.1, the proposed estimator can correctly distinguish linear and nonlinear components with high probability. Furthermore, the proposed estimator has the oracle property in the sense that it is the same as the oracle estimator assuming the identity of the linear and nonlinear components were known, except on an event with probability tending to zero.

The main extra condition here is (3.3), which requires both $\theta_* > \lambda\gamma + a_n m_n^{-(2d-1)/2}$ and $\lambda = o(n^{-1/2})$ for some $a_n \to \infty$. The first part of this requirement ensures that the bias resulting from the penalty is so small that it does not interfere with selection, and the second part requires that the smallest norm $\theta_*$of the coefficients in the spline expansions of the (nonzero) nonlinear components be larger than the penalty level plus a term due to the spline approximation error.

Theorem 3.2. Under (A1)—(A3) hold, we have

$$\sum_{j=1}^{p} \|\hat{\alpha}_{nj}(u) - \alpha_{0j}(u)\|_2^2 \leq O_P(m_n/n) + O(1/m_n^{2d}) + O(m_n\lambda^2)$$

This theorem gives rate of convergence of the proposed estimator of nonparametric regression coefficient under the the varying-coefficient partially linear AFT model (1.1). In particular, if we assume that each nonparametric function component in (1.1) is second-order differentiable (d = 2) and take $m_n = O(n^{1/5})$ and $\lambda = n^{-1/2+\delta}$ for a small$\delta > 0$, then $\sum_{j=1}^{p}\|\hat{\alpha}_{nj}(u) - \alpha_{0j}(u)\|_2^2 \leq O_P(n^{-4/5})$, which is the optimal rate of convergence in nonparametric function.

We now consider the asymptotic distribution of $\hat{\beta}_{n1}$. According to section 3.3 of Fan and Huang (2005)[2], we can obtain Theorem 3 as follows.

Theorem 3.3. If the conditions (A1)—(A3) are satisfied, and Theorems 1 and 2 also hold, then

$$\sqrt{n}(\hat{\beta}_{n1} - \beta_{(1)}) \xrightarrow{d} N(0, \Sigma)$$

where $\Sigma = \left(E(X_{(1)}^{*}X_{(1)}^{*T}) - E[E(X_{(1)}^{*}Z_{(2)}^{*T})\{E(Z_{(2)}^{*}Z_{(2)}^{*T})\}^{-1}E(Z_{(2)}^{*}X_{(1)}^{*T})]\right)^{-1}\sigma^2$ and $\beta_{(1)} = (\beta_j : j \in S_1 \cup S_2)^T$. The limit distribution of Theorem 3.3 is the same as that of the oracle estimator $\tilde{\beta}_{n1}$.

4. Numerical studies

We conduct extensive simulation studies to evaluate the finite-sample performance of the proposed methods. Suppose the failure time T follows the varying-coefficient partial linear model:

$$T_i = \mu + \left\{\sum_{j \in S_1} \beta_{0j}x_{ij} + \sum_{j \in S_2} \beta_{1j}u_i x_{ij} + \sum_{j \in S_3} g_j(u_i)x_{ij}\right\} + \varepsilon_i, 1 \leq i \leq n$$

where $u_i, x_{ij}, \varepsilon_i, 1 \leq i \leq n, 1 \leq j \leq p$, independent and subject to the standard normal distribution. The coefficient are

$$\mu = 0.5, \alpha_1(u) = 2, \alpha_2(u) = 1.5u,$$

$$\alpha_3(u) = 2(2u-1)^2, \alpha_4(u) = 3\,sin\,2\pi u,$$

$$\alpha_5(u) = 4\,cos(2\pi u)/(2-u)^2, \alpha_6(u) = 4\,sin(2\pi u)/(2 - sin(2\pi u))$$

Two scenarios were considered in the simulation.

Scenario 1. Let p = 4 and consider the model

$$T = 0.5 + 2x_1 + 1.5ux_2 + \alpha_3(u)x_3 + \alpha_4(u)x_4 + \varepsilon_i$$

Here the first variable is constant effect, the second variable is linear effect and the last two have nonlinear effect.

Scenario 2. Let p = 6 and consider the model

$$T = 0.5 + 2x_1 + 1.5ux_2 + \alpha_3(u)x_3 + \alpha_4(u)x_4 + \alpha_5(u)x_5 + \alpha_6(u)x_6 + \varepsilon_i$$

Here the first variable is the constant effect, the second variable is the linear effect and the last four have the nonlinear effect.

In each scenario, two sample sizes (n=500,1000) and two censoring rates (censoring=20%,40%) were conducted. We use the cubic B-spline basis functions to approximate each α_j and utilzied the group coordinate descent algorithm for computing the group MCP solutions. The group coordinate descent algorithm optimizes a target function with respect to a single group at a time, iteratively cycling through all groups until convergence is reached which is described in details in section 3 of Huang et al.(2012)[5].

In our application, we apply the bayesian information criterion (BIC) (Schwarz (1978))[12] to select (λ , γ) for the group MCP. The BIC is given by

$$BIC(\lambda,\gamma) = log(RSS_{\lambda,\gamma}) + log_n \frac{m_n df_{\lambda,\gamma}}{n}$$

where RSS is the residual sum of squares, $df_{\lambda,\gamma}$ is the number of the nonzero selected groups for a given (λ , γ). Recall that m_n is the number of spline basis functions given in (2.3). We choose λ from a sequence of 100 values, starting from$\lambda_{max}$to $0.01\lambda_{max}$. Here $\lambda_{max} = \max_{1\le j\le p}\left\|n^{-1}\tilde{Z}_j^T\tilde{Y}\right\|$, which is the smallest value of λ that forces all the solutions to be zero. For the γ, we consider a grid of equally spaced points in the interval $[\gamma_{max}, \gamma_{min}] = [8.0,1.1]$with grid size 0.1.

The simulation results based on 100 replications are presented in Tables 1-4. The columns in Tables 1 and 2 are the average number of nonlinear components being selected (NL), and the percentage of occasions in which the correct nonlinear components were included in the selected model (IN%). Enclosed in parentheses are the corresponding standard errors. In Tables 3 and 4, we examine the performance of the proposed method for estimating the nonlinear components in the simulated models.

Table 1. Simulation results for Examples 1. NL, the average number of the nonlinear components being selected; IN%, the percentage of occasions in which the correct nonlinear components are included in the selected model, averaged over 100 replications. Enclosed in parentheses are the corresponding standard errors.

Example 1	censoring=20%		censoring=40%	
	NL	IN%	NL	IN%
n=500, group Lasso	2.01	0.69	1.93	0.72
	(0.969)	(0.464)	(0.956)	(0.451)
group MCP	3.03	0.89	2.94	0.93
	(0.881)	(0.314)	(0.886)	(0.256)
n=1000, group Lasso	1.05	0.82	0.98	0.42
	(0.821)	(0.501)	(0.821)	(0.496)
group MCP	3.13	0.94	3.08	0.95
	(0.824)	(0.238)	(0.872)	(0.219)

Table 1 and Table 2 compare the results of the two sample sizes when the censoring rate is 20% and 40%. It shows that the proposed method with the group MCP performed better than the proposed method with the group Lasso in terms of the NL and the IN% in the final model. For instance, in

Example 2 with n= 1000, censoring rate=20%, the percentage of correct selection NL was 4.41 with the group MCP and 1.96 with the group Lasso. From table 1 and table 2, we show that the group MCP was more accurate in distinguishing the nonlinear functions than the group Lasso.

Table 2. Simulation results for Examples 2. The terms of NL and IN% are interpreted as shown in Table 1. Enclosed in parentheses are the corresponding standard errors.

Example 2	censoring=20%		censoring=40%	
	NL	IN%	NL	IN%
n=500, group Lasso	2.58	0.90	2.30	0.89
	(1.190)	(0.301)	(1.193)	(0.314)
group MCP	4.25	0.98	4.09	0.97
	(1.306)	(0.141)	(1.198)	(0.171)
n=1000, group Lasso	1.96	0.83	1.72	0.74
	(1.100)	(0.377)	(1.054)	(0.440)
group MCP	4.41	1	4.40	0.99
	(1.198)	(0.000)	(1.247)	(0.100)

In Tables 3 and 4, we examine the performance of the proposed method for estimating the nonlinear components in the simulated models when the censoring rate is 20%. We show that the group MCP picked out the nonlinear part more times than the group Lasso in 100 replications. For instance, in Example 2 with n= 1000, $\alpha_3$was picked out 74 times with the group MCP and 39 with the group Lasso. Overall, the proposed method with the group MCP was effective in distinguishing the nonlinear ones in the simulation models.

Table 3. Number of times each component was selected as a nonlinear component by the group Lasso and group MCP methods in the 100 replications when the censoring rate is 20%, in Example 1.

	$\alpha_3(\cdot)$	$\alpha_4(\cdot)$	$\alpha_3(\cdot)$	$\alpha_4(\cdot)$
	n=500		n=1000	
Example 1, group Lasso	45	45	20	32
	(0.500)	(0.500)	(0.402)	(0.468)
Example 1,group MCP	73	71	79	877
	(0.446)	(0.456)	(0.409)	(0.423)

Table 4. Number of times each component was selected as a nonlinear component by the group Lasso and group MCP methods in the 100 replications when the censoring rate is 20%, in Example 2.

n=500	$\alpha_3(\cdot)$	$\alpha_4(\cdot)$	$\alpha_5(\cdot)$	$\alpha_6(\cdot)$
Example 2,group Lasso	33	50	44	45
	(0.472)	(0.503)	(0.499)	(0.500)
Example 2,group MCP	71	67	71	67
	(0.456)	(0.472)	(0.456)	(0.472)
n=1000	$\alpha_3(\cdot)$	$\alpha_4(\cdot)$	$\alpha_5(\cdot)$	$\alpha_6(\cdot)$
Example 2,group Lasso	39	31	35	26
	(0.490)	(0.464)	(0.479)	(0.441)
Example 2,group MCP	74	75	73	75
	(0.440)	(0.435)	(0.446)	(0.435)

5. Conclusion

In this article, we combine the variable-factor partial linear models with the AFT models. Censored data is processed with K-M weights to ensure full utilization of the data. We propose a method for distinguishing linear components and nonlinear components in varying-coefficient partial linear AFT models based on group MCP. Compared to the standard varying-coefficient inference approach, the proposed method does not require a priori assumption that the covariates have linear and nonlinear effects. We proved the asymptotic oracle properties of the proposed estimator under the group MCP method, meaning that it is the same as the standard varying-coefcient partially linear regression estimator assuming the model structure is known. The asymptotic rates of the penalty parameters required for our theoretical results are derived. However, in this paper, we studied automatic structure of the nonlinear part, and did not consider to estimate the nonlinear part. This is an important and challenging problem that requires further investigation, but is beyond the scope of the current paper. Our simulation study indicates that the proposed method works well in finite sample situations.

References

[1] Buckley J, James I. Linear regression with censored data[J]. Biometrika, 1979, 66:429-436.

[2] Fan J, Huang T. Profle likelihood inferences on semiparametric varying-coefcient partially linear models[J]. Bernoulli, 2005, 11(6):1031–1057.

[3] Gui J, Li H Z. Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data[J]. Bioinformatics, 2005, 21:3001- 3008.

[4] Huang J, Ma S G. Variable selection in the accelerated failure time model via the bridge method[J]. Lifetime data analysis, 2010,16:176-195.

[5] Huang J, Wei F, Ma S. Semiparametric regression pursuit[J]. Statistica Sinica, 2012, 22(4):1403-1426.

[6] Lam C, Fan J Q. Profle-kernel likelihood inference with diverging number of parameters[J]. The Annals of Statistics, 2008, 36: 2232–2260.

[7] Li G R, Lu L, Zhu L X. Empirical likelihood for a varying coefficient partially linear model with diverging number of parameters[J]. Journal of Multivariate Analysis, 2012, 105:85-111.

[8] Ma S, Huang J. Additive risk survival model with microarray data[J]. BMC Bioinformatics, 2007, 8:192.

[9] Park E, Ha I D. Penalized variable selection for accelerated failure time models with random effects[J]. Statistics in Medicine, 2019, 38:878-892.

[10] Pieruschka E. Relation between lifetime distribution and the stress level causing the failures[J]. Sunnyvale:lockheed missile and space division, 1961, 1-30.

[11] Schumake L L. Spline Functions: Basic Theory[M]. New York :John Wiley & Sons, Inc,1981.

[12] Schwarz G. Estimating the dimension of a model[J]. The Annals of Statistics, 1978, 6(2):461-464.

[13] Stute W. Consistent estimation under random censorship when covariables are available[J]. Journalof Multivariate Analysis, 1993, 45:89-103.

[14] Wang H S, Xia Y C. Shrinkage estimation of the varying coefficient model[J]. Journal of the American Statistical Association, 2008, 104(486):0162-1459.

[15] Wei C H, Wan L J, Liu C L. Efficient Estimation in Heteroscedastic Partially Linear Varying Coefficient Models[J]. Communications in Statistics - Simulation and Computation, 2015, 44:892–901.

[16] Yang G R, Sun Y Q, Cui X. Automatic structure discovery for varying-coefficient partially linear models[J]. Communications in Statistics - Theory and Methods, 2017, 46(15):7703-7716.

[17] Ying A L. A large sample study of rank estimation for censored regression data[J]. Institute of Mathematical Statistics, 1993, 21(1):76-99.

[18] Zhang C H. Nearly unbiased variable selection under minimax concave penalty[J]. The Annals of Statistics, 2010, 38(2):894-942.

[19] Zhou M. M-estimation in censored linear models[J]. Biometrika, 1992, 79(4): 837-41.