Partially linear transformation cure models for interval-censored data

Hu, Tao; Xiang, Liming

2014

Hu, T., & Xiang, L. (2016). Partially linear transformation cure models for interval-censored data. *Computational Statistics and Data Analysis*, 93, 257-269.

<https://hdl.handle.net/10356/85081>

<https://doi.org/10.1016/j.csda.2014.08.014>

© 2014 Elsevier. This is the author created version of a work that has been peer reviewed and accepted for publication by *Computational Statistics and Data Analysis*, Elsevier. It incorporates referee's comments but changes resulting from the publishing process, such as copyediting, structural formatting, may not be reflected in this document. The published version is available at: [<http://dx.doi.org/10.1016/j.csda.2014.08.014>].

Downloaded on 26 Apr 2023 20:02:13 SGT

Partially Linear Transformation Cure Models for Interval-Censored Data

Tao Hu

School of Mathematical Sciences & BCMIS, Capital Normal University
Beijing 100048, China

Liming Xiang *

School of Physical and Mathematical Sciences, Nanyang Technological University
21 Nanyang Link, Singapore 637371

Abstract

There has been considerable progress in the development of semiparametric transformation models for regression analysis of time-to-event data. However, most of the current work focuses on right-censored data. Significantly less work has been done for interval-censored data, especially when the population contains a nonignorable cured subgroup. A broad and flexible class of semiparametric transformation cure models is proposed for analyzing interval-censored data in the presence of a cure fraction. The proposed modeling approach combines a logistic regression formulation for the probability of cure with a partially linear transformation model for event times of susceptible subjects. The estimation is achieved by using a spline-based sieve maximum likelihood method, which is computationally efficient and leads to estimators with appealing properties such as consistency, asymptotic normality and semiparametric efficiency. Furthermore, a goodness-of-fit test can be constructed for the proposed models based on the sieve likelihood ratio. Simulations and a real data analysis are provided for illustration of the methodology.

KEY WORDS: Cure rate; Interval censoring; Semiparametric efficiency; Sieve likelihood ratio test; Sieve maximum likelihood estimator; Transformation.

1 Introduction

Interval-censored data commonly arise in clinical and public health sciences, where the event time is never observed precisely but known at intermittent follow-up visits. Right and left censoring are special cases of interval censoring. A number of statistical methods for handling such data can be found, e.g., in Sun (2006). Due to advances in modern

*Corresponding Author. *Email Address:* lmxiang@ntu.edu.sg; Tel.: +65 6513 7451; Fax: +65 6515 8213.

medical techniques and health care, survival rates have substantially improved. This is often indicated in some studies by the presence of a significant "cured" proportion of subjects who are not at risk to experiencing the event of interest, such as failure or relapse. As a consequence, there may be a considerable number of subjects with large right-censored times in a sample.

Basically, there are two classes of cure rate models developed in the literature to allow for such a subgroup of subjects to be cured. One class is promotion time (non-mixture) cure models proposed by Yakovlev and Tsodikov (1996), in which a single model is used to describe the survival function of the whole population with right-censored data. For interval-censored data, Liu and Shen (2009) studied the promotion time cure model using a semiparametric maximum likelihood approach, while Hu and Xiang (2013) developed a spline-based sieve likelihood method for the same model, leading to more efficient estimators.

Models we consider in this paper belongs to the other class of cure rate models, named as mixture cure models, which assume that the study population is a mixture of susceptible and non-susceptible subjects, and then model the effects of covariates on the cure rate of the population and the survival function of non-susceptible subjects through the logistic regression and survival model components respectively. This modeling approach is straightforward and has attracted much attention in the literature (e.g. Farewell, 1982; Kuk and Chen, 1992; Li et al., 2001; Tsodikov et al., 2003). Extensions of the mixture cure models to interval-censored data have been discussed in recent years. Banerjee and Carlin (2004) proposed parametric cure rate models for interval-censored smoking relapse times from a Bayesian perspective . With current status data, known as a special case of interval-censored data, Lam and Xue (2005) discussed a semiparametric accelerated failure time (AFT) model based cure mixture using the sieve maximum likelihood estimation. Ma (2010) studied a semiparametric Cox cure model for mixed case interval-censored data using the maximum likelihood approach and weighted bootstrap for estimation and inference. More recently, Li and Ma (2010) applied a fully parametric AFT model and Xiang et al. (2011) and Lam and Wong (2014) considered the semiparametric Cox model for the susceptible group in the analysis of clustered interval-censored data, where the correlated nature of the data was accommodated through random effects in both the logistic regression and survival model components.

However, either the parametric survival model or the proportional hazards assumption is too restrictive for susceptible subjects in many biomedical applications and may result in unreliable estimates when the assumption is violated. This leads to the development of more accurate modeling approaches. The proportional odds model, as a useful alternative, provides better summary of data when the hazard ratio between two sets of covariate values is not constant but converges over time (Murphy *et al.*, 1997). The proportional odds cure rate models have been studied consequently by Mao and Wang (2010) and Gu *et al.* (2011). To encompass both the proportional hazards and proportional odds based mixture cure models as special cases, Fine (1999) and Lu and Ying (2004) exploited the mixture cure model with linear transformation models for the susceptible population using estimating equation methods. Their work has been further extended by Othus *et al.* (2009) to more general situations incorporating dependent censoring as well as time-dependent covariates. In addition, transformation models within the non-mixture cure model framework have been discussed by Zeng *et al.* (2006) given biological considerations and Yin (2008) for multivariate survival data. However, all these aforementioned studies have focused on right-censored data. Ma and Kosorov (2005) studied partly linear transformation models for current status data without considering the possibility of cure.

In this paper, we develop a semiparametric transformation modeling approach, in which survival time of a susceptible subject is specified by transforming it into a variable linked to covariates through a partially linear regression function while the cure fraction is modeled by a logistic regression. A key feature of semiparametric models with interval-censored data is that the baseline hazard function can not be eliminated from the likelihood or partial likelihood as that under right censoring. We approximate the unknown functions via a monotone B-spline (Schumaker, 1981) and construct a spline-based sieve semiparametric likelihood (Shen and Wong, 1994; Shen, 1997) to estimate parameters and the nonparametric components simultaneously. Because the sieve likelihood function usually involves fewer parameters to be estimated than the nonparametric likelihood function, the proposed estimation may reduce the computational burden in comparison with the non-parametric maximum likelihood (NPML) estimation techniques used in Ma and Kosorok (2005) and Liu and Shen (2009). After estimation, it is natural to construct hypothesis test for checking the goodness-of-fit of the model. We then develop a sieve likelihood ratio test for selecting sig-

nificant variables in the parametric component and checking the nonparametric component, in a similar spirit to the testing approach given by Shen and Shi (2005).

The rest of the paper is organized as follows. Section 2 presents the model, sieve maximum likelihood estimation and its computational issues. Section 3 studies asymptotic properties of the resulting estimator and then develops a goodness-of-fit test based on the sieve likelihood ratio. Simulation studies and a real data example are provided in Sections 4 and 5, respectively. Finally, Section 6 concludes the paper with a discussion. Sketches of proofs of theorems are given in the Appendix.

2 Model and Inference

2.1 Model and Likelihood function

Under the mixture cure modeling approach (Farewell, 1982), a decomposition of the event time is given by

$$T = YT^* + (1 - Y)\infty, \quad (2.1)$$

where $T^* < \infty$ denotes the failure time of a susceptible subject and Y indicates, by the value 1 or 0, whether the study subject is susceptible or not. Conditional on covariates $X \in \mathbb{R}^p$, $W \in \mathbb{R}$ and $Z \in \mathbb{R}^q$, we define the partially linear transformation cure models as follows:

$$\begin{aligned} H(T^*) &= -\beta^T X - \phi(W) + \epsilon, \\ p(Z) &= P(Y = 1|Z) = \frac{\exp(\alpha^T Z)}{1 + \exp(\alpha^T Z)}, \end{aligned} \quad (2.2)$$

where H is an unknown non-decreasing transformation, ϕ is an unknown smooth function, β and α are unknown regression parameter vectors of p - and q -dimension, respectively. Covariates Z and X may share some common components and Z includes 1 so that α contains an intercept term. ϵ is an error term with a known distribution F . Common choices of F include the extreme value distribution with $F_\epsilon(x) = 1 - \exp(-\exp(x))$ and the logistic distribution with $F_\epsilon(x) = \exp(x)/(1 + \exp(x))$ corresponding to the Cox proportional hazards model and the proportional odds model, respectively, for T^* when $\phi(W) = 0$.

We assume that the event time T may not be observed exactly, but known to be within a time interval $[U, V]$. That is, U is the last examination time before and V is the first examination time after the event, where $U < V$, U could be 0 and V could be ∞ . We

define $\delta_1 = I(0 = U \leq T \leq V < \infty)$ and $\delta_2 = I(U \leq T \leq V < \infty)$ according to whether U equals to zero or not, and denote the observation from a single subject by $\mathcal{W} = (\delta_1, \delta_2, U, V, X, Z, W)$. We also assume that conditional on X, Z and W , T is independent of (U, V) and the distribution of (U, V) is non-informative to T . Then the density of $\mathcal{W}$ is given by

$$\left\{ p(Z)F(H(V) + \beta^\tau X + \phi(W)) \right\}^{\delta_1} \left\{ p(Z)F(H(V) + \beta^\tau X + \phi(W)) - p(Z)F(H(U) + \beta^\tau X + \phi(W)) \right\}^{\delta_2} \left\{ 1 - p(Z)F(H(U) + \beta^\tau X + \phi(W)) \right\}^{1-\delta_1-\delta_2}.$$

Let $\theta = (\alpha^\tau, \beta^\tau, H, \phi)^\tau$ and $\theta_0 = (\alpha_0^\tau, \beta_0^\tau, H_0, \phi_0)^\tau$. We have the log-likelihood function of θ based on a single observation $(\delta_1, \delta_2, U, V, X, Z, W)$

$$\begin{aligned} l(\theta; \mathcal{W}) &= \delta_1 \log \{p(Z)\} + \delta_1 \log \{F(H(V) + \beta^\tau X + \phi(W))\} + \delta_2 \log \{p(Z)\} \\ &\quad + \delta_2 \log \{F(H(V) + \beta^\tau X + \phi(W)) - F(H(U) + \beta^\tau X + \phi(W))\} \\ &\quad + (1 - \delta_1 - \delta_2) \log \{1 - p(Z)F(H(U) + \beta^\tau X + \phi(W))\}. \end{aligned} \quad (2.3)$$

Under the partially linear transformation cure models, the log-likelihood of θ based on an i.i.d. sample $\mathcal{W}_i = (\delta_{1i}, \delta_{2i}, U_i, V_i, X_i, Z_i, W_i)$ for $i = 1, 2, \dots, n$ is given by

$$l_n(\theta) = \sum_{i=1}^n l(\theta; \mathcal{W}_i).$$

It is noted that with interval censoring the log-likelihood $l_n(\theta)$ depends on H only through its values at the distinct observed time points. Hence, we define $0 = t_0 < t_1 < t_2 < \dots < t_k < \infty$ to be the ordered distinct time points in the collection of observed interval end points $\{U_i, V_i : i = 1, \dots, n\}$. Semiparametric maximum likelihood estimates are obtained by maximizing $l_n(\theta)$ with respect to α, β and unknown functions $H(\cdot)$ and $\phi(\cdot)$. However, direct maximization of $l_n(\theta)$ is difficult and results in a challenging high-dimensional optimization problem when the sample size n is large. For example, the value of k could be up to $2n$ for cases without ties among finite values of $\{U_i, V_i\}$, $i = 1, \dots, n$. To overcome the curse of dimensionality, we propose a spline-based sieve maximum likelihood (ML) estimation method.

2.2 Sieve ML estimation for (α, β, H, ϕ)

We give a brief review of the generic sieve ML estimation first and then illustrate how it can be used to estimate parameters and nonparametric functions under the current model

setting.

The sieve method provides a convenient and general framework to obtain estimates over a subset (also known as a sieve space) of the infinite-dimensional parameter space and allows the subset to grow with the sample size (Grenander 1981). Geman and Hwang (1982) proposed a sieve ML estimation procedure to simultaneously estimate parameters as well as unknown functions involved in the likelihood. In the sieve ML estimation, a sieve log-likelihood is particularly constructed by approximating the unknown function through some nonparametric technique and estimates are obtained by maximizing the log-likelihood with respect to unknown parameters over a sieve space. To this end, we first approximate the transformation function $H(\cdot)$ and nonparametric regression term $\phi(\cdot)$ using splines, and then construct the sieve space.

In particular, $H(\cdot)$ is approximated by B-spline functions over interval $I = [a, b]$, where a and b are the lower and upper bounds of the finite observation times $\{(U_i, V_i) : i = 1, \dots, n\}$, $a \geq 0$ and $b > a$. Let $a = d_0 < d_1 < \dots < d_q = b$ be partition of I where $q = O(n^\nu)$ ($0 < \nu < 0.5$) is a positive integer such that $\max_{1 \leq j \leq q} |d_j - d_{j-1}| = O(n^{-\nu})$. Based on the method provided by Schumaker (1981) for choosing basis functions, we have $N = q + l$ normalized B-spline basis functions of order $l + 1$, which form a basis for the linear spline space. We denote these basis functions by $\boldsymbol{\pi}_1(t) = (B_1(t), \dots, B_N(t))^\tau$ with which $H(t)$ can be approximated by $H_n(t) = \boldsymbol{\pi}_1(t)^\tau \boldsymbol{\varphi}$, where $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_N)^\tau \in \mathbb{R}^N$ is a spline coefficient vector with $\boldsymbol{\varphi}_0$ corresponding to $H_0(t)$, basis functions $B_i \geq 0$ for $i = 1, \dots, N$ and $\sum_i B_i = 1$ at each point within $[a, b]$. Similarly, we can approximate $\phi(\cdot)$ by $\phi_n(w) = \boldsymbol{\pi}_2(w)^\tau \boldsymbol{\psi}$ on the support of $[c, d]$, where $c \geq 0$, $d > c$, $\boldsymbol{\psi} = (\psi_1, \dots, \psi_N)^\tau \in \mathbb{R}^N$ is a spline coefficient vector with $\boldsymbol{\psi}_0$ corresponding to $\phi_0(w)$, $\boldsymbol{\pi}_2(w) = (\mathbf{b}_1(w), \dots, \mathbf{b}_N(w))^\tau$, $\mathbf{b}_i \geq 0$, $i = 1, \dots, N$ and $\sum_i \mathbf{b}_i = 1$ at each point within $[c, d]$.

Denote the parameter space of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\tau, \boldsymbol{\beta}^\tau, H, \phi)^\tau$ by

$$\Theta = \{\boldsymbol{\theta} : \boldsymbol{\alpha} \in \mathcal{A}, \boldsymbol{\beta} \in \mathcal{B}, H \in \mathcal{H}, \phi \in \mathcal{M}\} = \mathcal{A} \times \mathcal{B} \times \mathcal{H} \times \mathcal{M},$$

where $\mathcal{A}$ is a compact subset of $\mathbb{R}^q$, $\mathcal{B}$ is a compact subset of $\mathbb{R}^p$, $\mathcal{H} = \{H : H \in C^m[a, b], |H^{(m)}(t_1) - H^{(m)}(t_2)| \leq L_1 |t_1 - t_2|^\gamma, \forall t_1, t_2 \in [a, b]\}$, and $\mathcal{M} = \{\phi : \phi \in C^m[c, d], |\phi^{(m)}(w_1) - \phi^{(m)}(w_2)| \leq L_2 |w_1 - w_2|^\gamma, \forall w_1, w_2 \in [c, d]\}$ for a nonnegative integer m and unknown positive constants L_1 and L_2 . $\gamma \in (0, 1]$ with $m + \gamma > 0.5$. The notation $C^m[a, b]$ denotes a class of continuous functions having m th derivative in $[a, b]$. Such smoothness assumptions

for both $\mathcal{H}$ and $\mathcal{M}$ are often used in nonparametric curve estimation.

We then describe the construct of the sieve space for parameters $\boldsymbol{\theta}$. First we introduce some more notation. For any $\boldsymbol{\theta}_i \in \Theta, i = 1, 2$, we define a distance

$$d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|^2 + \|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|^2 + \|H_1 - H_2\|_2^2 + \|\phi_1 - \phi_2\|_2^2)^{1/2},$$

where $\|a\|$ is the Euclidean norm of a vector a , $\|g(X)\|_2 = (\int g^2 dP)^{1/2}$ is the $L^2(P)$ norm of a function g for X being distributed according to the probability measure P . $\|H_1 - H_2\|_2^2 = \|H_1(U) - H_2(U)\|_2^2 + \|H_1(V) - H_2(V)\|_2^2$, where $\|H_1(U) - H_2(U)\|_2^2 = E\{H_1(U) - H_2(U)\}^2$ and $\|H_1(V) - H_2(V)\|_2^2 = E\{H_1(V) - H_2(V)\}^2$. We further denote sets

$$\begin{aligned}\mathcal{H}_n &= \left\{ H_n : H_n(t) = \sum_{i=1}^N B_i(t) \varphi_i, \varphi = (\varphi_1, \dots, \varphi_N)^\tau \in \mathcal{C}_{1n} \right\}, \\ \mathcal{M}_n &= \left\{ \phi_n : \phi_n(w) = \sum_{i=1}^N \mathbf{b}_i(w) \psi_i, \psi = (\psi_1, \dots, \psi_N)^\tau \in \mathcal{C}_{2n} \right\},\end{aligned}$$

$$\mathcal{C}_{1n} = \{ \varphi : \varphi_1 \leq \varphi_2 \leq \dots \leq \varphi_N, \max_{1 \leq i \leq N} |\varphi_i| \leq \iota_n \},$$

and

$$\mathcal{C}_{2n} = \{ \psi : \max_{1 \leq i \leq N} |\psi_i| \leq \iota_n \},$$

where $\iota_n \leq n^{(2l-1)/(2l'(2l+1))}$ with a constant l' arbitrarily close to l (Shen, 1997). The choice of sets $\mathcal{H}_n$ and $\mathcal{M}_n$ is similar to those given in Zhang et al.(2010) and Hu and Xiang (2013). The monotonicity constraints on φ_i ensure that each element of $\mathcal{H}_n$, as an estimate of the non-decreasing function H , is also non-decreasing. By Corollary 6.21 in Schumaker (1981), we then can use $\Theta_n = \mathcal{A} \times \mathcal{B} \times \mathcal{H}_n \times \mathcal{M}_n$ as a sieve space of Θ .

Note that each element of $\mathcal{H}_n$ is a non-decreasing function because of the monotonicity constraints on transformation H . This fact is a consequence of the variation diminishing properties of B-splines (see example 4.75 and theorem 4.76 of Schumaker, 1981). The sieve ML estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{H}, \hat{\phi})^\tau$ is obtained by maximizing $l_n(\boldsymbol{\theta})$ over the sieve space Θ_n . It is equivalent to maximizing $l_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}^\tau \boldsymbol{\varphi}, \boldsymbol{\pi}^\tau \boldsymbol{\psi})$ with respect to $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ over $\mathcal{A} \times \mathcal{B} \times \mathcal{C}_{1n} \times \mathcal{C}_{2n}$, at a substantially reduced computational cost since the number of basis functions N required grows much slower than the sample size n .

2.3 Computational issues

In the sieve ML estimation procedure proposed, we typically choose cubic splines of order $l = 3$ to approximate unknown functions smoothly. Other splines such as linear and quadratic splines could also be used if $H(t)$ and $\phi(w)$ are less smooth. Selecting the number of knots for the B-spline is often important for controlling the roughness of the functions approximated in the spline-based estimation. We use knots that are quantiles of the finite observation times $\{(U_i, V_i) : i = 1, 2, \dots, n\}$ with uniform percentile ranks, while the number of knots s is determined by a model selection criterion. Given the sieve ML estimator $\hat{\boldsymbol{\theta}}$, it is natural to determine s or equivalently $N = (s + l)$ by minimizing the following BIC,

$$\text{BIC} = -2l_n(\hat{\boldsymbol{\theta}}) + (2N + p + q) \log n,$$

where $l_n(\boldsymbol{\theta})$ is the sieve log-likelihood given in Section 2.2.

In addition, the numerical optimization of $l_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}^\tau \boldsymbol{\varphi}, \boldsymbol{\pi}^\tau \boldsymbol{\psi})$ in the estimation procedure can be readily implemented using an existing constrained nonlinear programming routine such as R packages *Rsolnp* and *Rdonlp2*, the SAS constrained optimization procedure *PROC NLP* or Matlab function *fmincon*. The programming effort is reasonable since users only need to code the objective function and monotonicity constraints.

3 Theoretical Properties

In this section, we study the asymptotic behavior of the estimator proposed in Section 2.2. We first present the regularity conditions required.

(A1). (a) There exists a minimal examination spacing $\zeta > 0$ such that $P(V - U \geq \zeta) = 1$; and (b) the union of the supports of finite U and V is contained in an interval $[a, b]$, where $0 < a < b < \infty$.

(A2). (a) $\boldsymbol{\alpha}_0 \in \mathcal{A}$ and $\mathcal{A}$ is compact subset of $\mathbb{R}^q$; (b) $\boldsymbol{\beta}_0 \in \mathcal{B}$ and $\mathcal{B}$ is compact subset of $\mathbb{R}^p$; (c) The function $H_0 \in \mathcal{H}$ is strictly increasing function on $[0, \tau]$; (d) The true nonparametric covariate effect $\phi \in \mathcal{M}$; (e) The error term distribution F is known and F has first and second derivatives f and $f^{(1)}$, where the support of f is $\mathbb{R}$ and $|f^{(1)}|$ is bounded.

(A3). (a) Z belongs to a bounded subset of $\mathbb{R}^q$; (b) X belongs to a bounded subset of $\mathbb{R}^p$; (c) The support of W is $[c, d]$, where $-\infty < c < d < \infty$.

(A4). (a) For covariates Z (or X), if there exists a constant c_0 and a constant vector $\tilde{\gamma}$ such that $\tilde{\gamma}^\tau Z = c_0$ (or $\tilde{\gamma}^\tau X = c_0$) almost surely, then $c_0 = 0$ and $\tilde{\gamma} = 0$; (b) $H(0) = 0$;

(c) There exists a constant τ , such that $P(V = \tau|X, Z) = P(V \geq \tau|X, Z) > 0$, and $F(H(\tau) + \beta^\tau X + \varphi(W)) = 1$ for $\beta \in \mathbb{R}^p$, $H \in \mathcal{H}$ and $\phi \in \mathcal{M}$.

Assumptions (A1) and (A2) are asymptotic frameworks for semiparametric regression with interval-censored data commonly used in the literature, for example Huang and Rossini (1997), Zhang et al. (2010) and Ma (2010). Assumption (A2) controls the rate of convergence of the sieve ML estimator, see Shen and Wong (1994) for a more detailed discussion. The boundedness assumption (A3) is required in the consistency and weak convergence proofs. Assumption (A4) ensures the identifiability and consistency. Similar assumptions were used in Zeng et al. (2006), Liu and Shen (2009), Ma (2010) and Hu and Xiang (2013). Particularly, (A4)(a) guarantees that there is no multicollinearity among covariates in the parametric parts. The fulfillment of (A4)(b) can be obtained by re-centering $H(v)$ to $H(v) - H(0)$. (A4)(c) describes the requirement for sufficiently long follow-up in which at least some subjects are cured so that τ can be treated as ∞ . We noted that the identifiability result of the proposed model holds for continuous or binary covariates, given that there are at least two covariates.

Under the regularity conditions given above, the theoretical results of estimator $\hat{\theta}$ are summarized in the following theorems.

Theorem 3.1 (*Strong Consistency*) Under Assumptions (A1)-(A4), $d(\hat{\theta}, \theta_0) \rightarrow 0$ almost surely.

Theorem 3.2 (*Rate of convergence*) Under Assumptions (A1)-(A4), $d(\hat{\theta}, \theta_0) = O_p(n^{-r\nu} + n^{-(1-\nu)/2})$, where r is the order of smoothness of the true regression function ϕ_0 and $0 < \nu < 0.5$. Moreover, $d(\hat{\theta}, \theta_0) = O_p(n^{-r/(1+2r)})$ if $\nu = 1/(1 + 2r)$.

By Theorem 3.2, the convergence rate of $\hat{\phi}$ is $n^{-r/(2r+1)}$ when $\nu = 1/(1 + 2r)$, which is the optimal global convergence rate of estimators for nonparametric regression according to Stone (1980). In practice, the value of r is unknown and has to be determined by specific considerations or by examination of the data. $r = 2$ (piecewise linear), 3 (quadratic) and 4 (cubic) are common choices to preserve smoothness and avoid the oscillation problem associated with higher order polynomials. Theorem 3.2 also implies that the smooth $\hat{H}$ can achieve the convergence rate $O_P(n^{-r/(2r+1)})$, faster than $n^{-1/3}$ -rate derived in the penalized estimation context by Ma and Kosorok (2005). Moreover, we can show that $\hat{H}$ is uniformly

consistent, i.e., $\|H - H_0\|_\infty = o_p(1)$, by applying Lemma 2 in Chen and Shen (1998).

Regarding the efficiency of the proposed estimator, an additional assumption for functional of $\boldsymbol{\theta}$ is required. We denote by $\mathcal{V}$ the linear span of $\Theta_0 - \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ and Θ_0 is the true parameter space. Let $l(\boldsymbol{\theta}; \mathcal{W})$ be the log-likelihood for a sample of size one and $\delta_n = n^{-r\nu} + n^{-(1/2-\nu)}$. For any $\boldsymbol{\theta} \in \{\boldsymbol{\theta} \in \Theta_0 : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| = O(\delta_n)\}$, we define the first order directional derivative of $l(\boldsymbol{\theta}; \mathcal{W})$ at the direction $v \in \mathcal{V}$ as

$$\dot{l}(\boldsymbol{\theta}; \mathcal{W})[v] = \left. \frac{dl(\boldsymbol{\theta} + tv; \mathcal{W})}{dt} \right|_{t=0},$$

the second order directional derivative as

$$\ddot{l}(\boldsymbol{\theta}; \mathcal{W})[v, \tilde{v}] = \left. \frac{d^2 l(\boldsymbol{\theta} + tv + \tilde{t}\tilde{v}; \mathcal{W})}{d\tilde{t}dt} \right|_{t=0, \tilde{t}=0} = \left. \frac{d\dot{l}(\boldsymbol{\theta} + \tilde{t}\tilde{v}; \mathcal{W})}{d\tilde{t}} \right|_{\tilde{t}=0},$$

the Fisher inner product on the space $\mathcal{V}$ as

$$\langle v, \tilde{v} \rangle = P\{\dot{l}(\boldsymbol{\theta}; \mathcal{W})[v]\dot{l}(\boldsymbol{\theta}; \mathcal{W})[\tilde{v}]\}$$

and the Fisher norm for $v \in \mathcal{V}$ as $\|v\|^{1/2} = \langle v, v \rangle$. Let $\bar{\mathcal{V}}$ be the closed linear span of $\mathcal{V}$ under the Fisher norm. Then $(\bar{\mathcal{V}}, \|\cdot\|)$ is a Hilbert space.

Furthermore, we define a smooth functional of $\boldsymbol{\theta}$

$$\varrho(\boldsymbol{\theta}) = \lambda_1^\tau \alpha + \lambda_2^\tau \beta + \int_a^b \lambda_3(w) \phi(w) dw + \int_0^\tau \lambda_4(t) H(t) dt,$$

where $\|\lambda_i\| \leq 1, i = 1, 2$, $\lambda_3 \in \mathcal{H}$ and $\lambda_4 \in \mathcal{M}$. For any $v \in \mathcal{V}$, we write

$$\dot{\varrho}(\boldsymbol{\theta}_0)[v] = \left. \frac{d\varrho(\boldsymbol{\theta}_0 + tv)}{dt} \right|_{t=0}$$

whenever the right hand-side limit is well defined and assume that:

(A5) for any $v \in \bar{\mathcal{V}}$, $\varrho(\boldsymbol{\theta}_0 + tv)$ is continuously differentiable in $t \in [0, 1]$ near $t = 0$, and

$$\|\dot{\varrho}(\boldsymbol{\theta}_0)\| = \sup_{v \in \bar{\mathcal{V}}: \|v\| > 0} \frac{|\dot{\varrho}(\boldsymbol{\theta}_0)[v]|}{\|v\|} < \infty.$$

Note that $\varrho(\boldsymbol{\theta}) - \varrho(\boldsymbol{\theta}_0) = \dot{\varrho}(\boldsymbol{\theta}_0)[\boldsymbol{\theta} - \boldsymbol{\theta}_0]$. Based on Assumption (A5) and the Riesz representation theorem, there exists $v^* \in \bar{\mathcal{V}}$ such that $\dot{\varrho}(\boldsymbol{\theta}_0)[v] = \langle v^*, v \rangle$ for all $v \in \bar{\mathcal{V}}$ and $\|v^*\|^2 = \|\dot{\varrho}(\boldsymbol{\theta}_0)\|$.

Theorem 3.3 *Suppose assumptions (A1)-(A5) hold, $0.25/r < \nu < 0.5$, $r \geq 2$, then $n^{1/2}(\varrho(\hat{\boldsymbol{\theta}}) - \varrho(\boldsymbol{\theta}_0)) \rightarrow N(0, \|\dot{\varrho}(\boldsymbol{\theta}_0)\|^2)$ in distribution and $\varrho(\hat{\boldsymbol{\theta}})$ is semiparametrically efficient.*

Theorem 3.3 offers ease of inference procedure, especially for parameters α and β . With $\lambda_3 = \lambda_4 = 0$, it yields that $n^{1/2}\{(\lambda_1^T(\hat{\alpha} - \alpha_0) + \lambda_2^T(\hat{\beta} - \beta_0))\} \rightarrow N(0, \|\dot{\varrho}(\theta_0)\|^2)$, and thus by the Gramer-Wold device and applying the general theory of Shen (1997), one can establish the asymptotic normality and semiparametric efficiency for $\hat{\alpha}$ and $\hat{\beta}$.

In general, there is no closed form for asymptotic variances of $\hat{\alpha}$ and $\hat{\beta}$. Instead, these asymptotic variances can be estimated by the weighted bootstrap method (Ma and Kosorok, 2005; Cheng and Huang, 2010) through the following steps. First, we generate n i.i.d. positive random weights $\omega_1, \dots, \omega_n$ from a known distribution with $E(\omega) = Var(\omega) = 1$ (e.g., weight $\omega \sim \text{EXP}(1)$). Next, we obtain a weighted sieve maximum likelihood estimator via

$$\tilde{\theta} = \arg \min_{\theta \in \Theta_n} \sum_{i=1}^n \omega_i l(\theta; \mathcal{W}_i),$$

where $\tilde{\theta} = (\tilde{\alpha}, \tilde{\beta}, \tilde{H}, \tilde{\phi})$ and $l(\theta; \mathcal{W}_i)$ is defined in (2.3). Given $\mathcal{W}_i$, it follows from Theorem 1 in Cheng and Huang (2010) that $\sqrt{n}(\tilde{\alpha} - \hat{\alpha})$ and $\sqrt{n}(\hat{\alpha} - \alpha_0)$, $\sqrt{n}(\tilde{\beta} - \hat{\beta})$ and $\sqrt{n}(\hat{\beta} - \beta_0)$ have the same limiting distributions which can be used to justify the weighted bootstrap for inference on $\hat{\alpha}$ and $\hat{\beta}$, respectively. We then approximate the distributions of $\hat{\alpha}$ and $\hat{\beta}$ by generating a large number of samples (we use 500 samples in the numerical results of this paper) from the distributions of $\tilde{\alpha}$ and $\tilde{\beta}$, respectively. The variance-covariance matrices of $\hat{\alpha}$ and $\hat{\beta}$ can consequently be computed based on their empirical distributions.

Theorem 3.3 can also be applied for assessing the goodness-of-fit of the nonparametric component in the proposed model (2.2). Let $\lambda_1 = 0$, $\lambda_2 = 0$, $\lambda_4 = 0$ and denote $\varrho_\phi(\theta) = \int_c^d \lambda_3(w)\phi(w)dw$. It follows from Theorem 3.3 that

$$n^{1/2} \int_c^d \lambda_3(w)(\phi(w) - \phi_0(w))dw \rightarrow N(0, \|\dot{\varrho}_\phi(\theta_0)\|^2), \quad (3.1)$$

where $\|\dot{\varrho}_\phi(\theta_0)\|^2$ can be consistently estimated by the weighted bootstrap. Using result (3.1), it is possible to consider the testing problem with null hypothesis on the nonparametric component $H_0 : \phi(w) = \phi_0(w)$, where $\phi_0(w)$ is a pre-specified parametric function. In particular, if we are interested in checking the linear or quadratic effect of covariate w in the model, we then test for the null hypothesis H_0 with

$$\phi_0(w) = \gamma_0 + \gamma_1 w, \quad \text{or} \quad \phi_0(w) = \gamma_0 + \gamma_1 w + \gamma_2 w^2.$$

Let $R_n = n^{1/2} \int_a^b \lambda_3(w)(\hat{\phi}(w) - \phi_0(w))dw$. To this end, a testing statistic can be defined as

$$\mathcal{R}_n = \frac{R_n^2}{\|\dot{\varrho}_\phi(\boldsymbol{\theta}_0)\|^2}. \quad (3.2)$$

It follows from (3.1) that $\mathcal{R}_n \sim \chi_1^2$ under H_0 .

More generally, we now consider the hypothesis test for a known smooth functional $\boldsymbol{\rho}(\boldsymbol{\theta}) = (\rho_1(\boldsymbol{\theta}), \dots, \rho_k(\boldsymbol{\theta}))^\tau : \Theta \rightarrow \mathbb{R}^k$,

$$\mathcal{H}_0 : \boldsymbol{\rho}(\boldsymbol{\theta}) = 0,$$

where $\boldsymbol{\rho}(\boldsymbol{\theta})$ is a vector of known functionals. Special cases of $\boldsymbol{\rho}(\boldsymbol{\theta})$ include $\boldsymbol{\rho}(\boldsymbol{\theta}) = \boldsymbol{\alpha} - \boldsymbol{\alpha}_0 \in \mathbb{R}^q$, $\boldsymbol{\rho}(\boldsymbol{\theta}) = \boldsymbol{\beta} - \boldsymbol{\beta}_0 \in \mathbb{R}^p$ and $\rho(w) = \phi(w) - \phi_0(w)$ with fixed w . Without loss of generality, we assume $\partial\rho_1(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\tau, \dots, \partial\rho_k(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\tau$ are linearly independent, otherwise a linear transformation can be performed for the hypothesis. Following the work of Shen and Shi (2005), we can develop a sieve likelihood ratio test for $\mathcal{H}_0$ under some conditions on $\rho_i(\boldsymbol{\theta})$. For completeness we list these conditions below.

(LR1). (a) For any $v \in \bar{\mathcal{V}}$, $\boldsymbol{\rho}(\boldsymbol{\theta}_0 + sv)$ is continuously differentiable in $s \in [0, 1]$ near $s = 0$, and

$$\|\dot{\boldsymbol{\rho}}(\boldsymbol{\theta}_0)\| = \sup_{v \in \bar{\mathcal{V}}: \|v\| > 0} \frac{|\dot{\boldsymbol{\rho}}(\boldsymbol{\theta}_0)[v]|}{\|v\|} < \infty.$$

(b) There exist constants $c > 0$, $\omega > 0$, and a small $\varepsilon > 0$ such that for any $v \in \bar{\mathcal{V}}$ with $\|v\| < \varepsilon$, we have

$$\left\| \boldsymbol{\rho}(\boldsymbol{\theta}_0 + v) - \boldsymbol{\rho}(\boldsymbol{\theta}_0) - \dot{\boldsymbol{\rho}}(\boldsymbol{\theta}_0)[v] \right\| \leq c\|v\|^\omega.$$

By the Riesz representation theorem, there exists $v_i^* \in \bar{\mathcal{V}}$ such that $\dot{\rho}_i(\boldsymbol{\theta}_0)[v] = \langle v, v_i^* \rangle$ for all $v \in \bar{\mathcal{V}}$ and $\|v_i^*\|^2 = \|\dot{\rho}_i(\boldsymbol{\theta}_0)\|^2$, $i = 1, 2, \dots, k$.

(LR2). For some positive sequence $\{\zeta_n, n \geq 1\}$, $\zeta_n \rightarrow 0$, $\liminf_{n \rightarrow \infty} n^{1/2}\zeta_n > 0$,

$$\limsup_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \sup_{\boldsymbol{\theta} \in \Theta_n: d(\boldsymbol{\theta}_0, \Pi_n \boldsymbol{\theta}_0) \geq M\zeta_n} [l_n(\boldsymbol{\theta}) > l_n(\Pi_n \boldsymbol{\theta}_0)] \geq 0 \right\} = 0.$$

In addition, $d(\boldsymbol{\theta}_0, \Pi_n \boldsymbol{\theta}_0) = O(\zeta_n)$, where $\Pi_n \boldsymbol{\theta}_0$ is a projection from Θ to the sieve space Θ_n .

Theorem 3.4 *Suppose assumptions (A1)-(A5), (LR1)-(LR2) hold, then*

$$\mathcal{T}_n = 2n \left\{ \max_{\boldsymbol{\theta} \in \Theta_n} l_n(\boldsymbol{\theta}) - \max_{\boldsymbol{\theta} \in \Theta_n: \rho(\boldsymbol{\theta})=0} l_n(\boldsymbol{\theta}) \right\} \rightarrow \chi_k^2$$

in distribution.

4 Simulation

We present simulation studies to examine the finite sample performance of the sieve ML estimator proposed in Section 2 and the power properties of test statistics derived in Section 3, respectively.

4.1 Performance of the estimation

We generate the cure indicator Y and event time T (if not cured) from the following partially linear cure model:

$$\begin{aligned} H(T) &= -\beta X - \phi(W) + \epsilon, \\ p(Z) &= P(Y = 1|Z) = \frac{\exp(\alpha_1 + \alpha_2 Z)}{1 + \exp(\alpha_1 + \alpha_2 Z)}, \end{aligned} \quad (4.1)$$

where the covariate Z corresponding to the cure rate is generated from the uniform $(0, 2)$ distribution and $X = Z$. The hazard function of error ϵ is

$$\lambda_\epsilon(t) = \exp(t) / \{1 + \nu \exp(t)\},$$

which reduces to **Model I**: the proportional hazards model if constant $\nu = 0$ and **Model II**: the proportional odds model if $\nu = 1$. We choose the transformation function $H(t) = \log(t/2)$, and set parameters $(\alpha_1, \alpha_2) = (2, -1)$, $\beta = 1$ and function $\phi(w) = 2 \sin(\pi w) - \exp(-0.5w)$. We consider the following two designs for the covariate W .

Design I: W is independent of Z and $W \sim Unif[0, 1]$;

Design II: W is dependent of Z , $W \sim Unif[0, 0.5]$ if $Z \leq 1$ and $W \sim Unif[0.5, 1]$ otherwise.

For each sample of size $n = 200$ and 400 , we generate interval-censored time observations by $U = \max(T - U^{(1)}, T + U^{(2)} - 1)$ and $V = \min(T + U^{(2)}, T - U^{(1)} - 1)$, where $U^{(1)}$ and $U^{(2)}$ are independent and uniformly distributed in the interval $(0, 1)$. We compute the sieve ML estimates following the procedure developed in Section 2 and estimate the standard errors of the estimated regression parameters using the weighted bootstrap method. For the cubic B-spline, the number of knots q or equivalently $N = (q + l)$ is chosen using BIC defined in Section 2.2.

500 replications are generated for each simulation. Tables 1-2 summarize the simulation results under models I and II, respectively, including the average bias (BIAS), sample

standard deviation (STD) and mean of the estimated standard error (ESE) of parameter estimates and the coverage proportion (CP) of the 95 percent confidence intervals. The results indicate that the proposed estimation performs satisfactorily for both the proportional hazards and proportional odds models. As expected, increasing sample size clearly leads to less biases and smaller standard errors of parameter estimates under the two covariate designs. Our parameter estimates basically unbiased, even when sample size is relative small. The standard deviations of the estimates shrink at approximately the $\sqrt{n}$ rate. The estimated standard errors are very close to the corresponding sample standard deviations, which supports validity of the weighted bootstrap methods. The 95% confidence intervals provide coverage generally close to the nominal level.

Table 1: Simulation results for Model I (the proportional hazards model) with 500 replications

n	α_1				α_2				β			
	BIAS	STD	ESE	CP	BIAS	STD	ESE	CP	BIAS	STD	ESE	CP
Design I												
200	0.048	0.085	0.079	0.921	0.047	0.073	0.085	0.921	-0.017	0.096	0.104	0.964
400	0.019	0.060	0.059	0.936	0.008	0.060	0.061	0.925	-0.023	0.069	0.063	0.957
Design II												
200	0.061	0.068	0.066	0.953	0.056	0.062	0.070	0.964	-0.014	0.096	0.097	0.936
400	0.011	0.040	0.044	0.953	0.031	0.059	0.057	0.932	-0.019	0.066	0.065	0.973

Table 2: Simulation results for Model II (the proportional odds model) with 500 replications

n	α_1				α_2				β			
	BIAS	STD	ESE	CP	BIAS	STD	ESE	CP	BIAS	STD	ESE	CP
Design I												
200	0.059	0.102	0.101	0.935	0.064	0.109	0.107	0.957	-0.056	0.121	0.142	0.923
400	0.025	0.067	0.081	0.965	0.017	0.082	0.081	0.982	-0.021	0.082	0.096	0.965
Design II												
200	0.045	0.082	0.075	0.954	0.063	0.114	0.097	0.934	0.031	0.098	0.097	0.959
400	0.010	0.044	0.047	0.945	0.003	0.048	0.042	0.934	-0.008	0.048	0.050	0.959

4.2 Power of the proposed test

To study the finite sample performance of the test statistics $\mathcal{T}_n$ and $\mathcal{R}_n$ in Section 3, we take the proportional hazards model with the same simulation settings as Model I in Subsection 4.1, though the results hold for the proportional odds model too. We first consider the test of significance for parameter β in the model, i.e., $H_0 : \beta = 0$ versus $H_1 : \beta = c_1 \neq 0$, where constant c_1 ranges from 0.1 to 0.5 with increment 0.1. The empirical size and power functions of $\mathcal{T}_n$ are calculated based on 500 i.i.d. samples of size $n = 200$ generated from Model I. Figure 1 presents the power functions versus c_1 evaluated at significance levels 0.05 and 0.1. As expected, the empirical size ($c_1 = 0$) is very close to the nominal level 0.05 and 0.1. The power of the test increases quickly as the signal c_1 gets stronger. The power becomes greater than 0.90 when $c_1 > 0.3$.

Figure 1. The power functions of the test $\mathcal{T}_n$ evaluated at significance levels 0.05(dashed) and 0.1 (solid).

Next we examine the performance of the test $\mathcal{R}_n$ for hypothesis: $H_0 : \phi(w) = 0$ versus $H_1 : \phi(w) \neq 0$. We compute the power of the test for a sequence of alternatives $H_1 : \phi(w) = c_2 \sin(2\pi x)$ indexed by c_2 , which controls the degree of departure from the null hypothesis. The value c_2 is taken to be a grid of equally spaced points on $[0, 0.3]$. Based on 1000 replications of sample sizes $n = 200$. Figure 2 depicts the power functions evaluated at significance level 0.05 and 0.1. It is evident that the power increases to 1 rapidly as c_2 increases. The empirical sizes of the test are 0.055 and 0.117, which are all close to the corresponding

significance levels 0.05 and 0.1, respectively.

Figure 2. The power functions of the test $\mathcal{R}_n$ at significance levels 0.05(dashed) and 0.1 (solid).

5 Real Data Application

For illustration purpose, we apply the proposed method to a study on effectiveness of a smoking intervention on risk of smoking resumption (Murray *et al.*, 1998). The data have been analyzed recently by researchers such as Banerjee and Carlin (2004) and Yu and Peng (2008) using parametric cure rate models, and Hu and Xiang (2013) through a non-mixture semiparametric cure model.

A total of 223 smokers who were known to have quit smoking at least once over the 5-year study period were included in this study. Each individual was observed once a year. If the individual resumed smoking after an initial attempt to quit, the primary outcome, relapse time, was observed as an approximate one-year time interval from the previous observation to the current observation. During the study, 158 individuals were found not having resumed smoking, providing a cessation rate of around 71 per cent. The four covariates X_1 , X_2 , X_3 and W were recorded for each individual, where X_1 was the sex indicator (0 = male, 1 = female), X_2 was the intervention type (1=special intervention, 0= usual care), X_3 was the number of cigarettes smoked per day, and W was the duration as smokers in years.

Unlike the existing work aforementioned, [we treat the effect of time-dependent covariate](#)

W nonparametrically in this analysis and consider a more flexible model framework, the partially linear transformation cure model, as follows.

$$\begin{aligned} H(T) &= -X_1\beta_1 - X_2\beta_2 - X_3\beta_3 - \phi(W) + \epsilon, \\ p(X_2) &= P(Y = 1|X_2) = \frac{\exp(\alpha_1 + \alpha_2 X_2)}{1 + \exp(\alpha_1 + \alpha_2 X_2)} \end{aligned} \quad (5.1)$$

where β_i , $i = 1, 2, 3$, denote the effects of the covariate X_i , $\phi_w(W)$ is an unknown smooth function that characterizes the main effect of W and Y denotes whether the study subject is susceptible or not. We choose only covariate X_2 in the cure part because it is particularly important in this example and is the only significant variable in the cure part revealed by Yu and Peng (2008) based on a Weibull cure model. We use cubic splines with $l = 3$ for approximating both unknown functions H and ϕ , and specify a general form for the baseline hazard function of ϵ as $e^t/(1 + \nu e^t)$. Note that ν and the knot number q of the B-spline are determined by minimizing the Bayesian information criterion:

$$\text{BIC}(\nu, q) = -2l_n(\hat{\boldsymbol{\theta}}) + (2(q + l) + 5) \log(n),$$

By running the estimation procedure described at Section 2, we choose ν to be 0.3 and q to be 2. The results of analysis are reported in Table 3. It is found that, similar to previous findings (Banerjee and Carlin, 2004; Yu and Peng, 2008), smokers with special intervention ($X_2 = 1$) have significantly higher cure rate than those with usual care and all three covariates are insignificant in the proportional hazards model part for the uncured group. When applying the proportional odds model for the uncured group, however, the effect of special intervention becomes significant in reducing on the relative risk of smoking resumption for smokers.

It is interesting to examine whether the nonlinear function $\phi(\cdot)$ is significantly different from zero, or whether the variable W has a significant effect on the smoking relapse time T of any kind (linear or nonlinear). For each $\nu = 0, 0.3$ and 1 , we apply $\mathcal{T}_n$ to test the hypothesis $H_0 : \phi(W) = 0$. All nonlinear functions are significantly different from 0 at level 0.05 with the p-values given as 0.018, 0.006, 0.031, respectively. We therefore conclude that (5.1) is more appropriate to fit this data set than the linear transformation cure mode, and the effect of variable W is statistically significant. To further check if the effect of W is linear on the smoking relapse time, it is equivalent to the test for the linearity of the unknown

Table 3: Parameter estimates in partially linear transformation cure models for the smoking cessation data

Parameters	Logistic regression for cure rate		Partially linear transformation model for uncured group		
	α_1	α_2	$-\beta_1$	$-\beta_2$	$-\beta_3$
$v=0.3$ (selected using BIC)					
Estimate	-0.572	-0.452	-0.190	0.472	-0.037
SE	0.135	0.139	0.146	0.140	0.051
$v = 0$ (Proportional hazards model)					
Estimate	-0.401	-0.311	-0.241	0.262	-0.009
SE	0.078	0.135	0.174	0.172	0.064
$v = 1$ (Proportional odds model)					
Estimate	-0.407	-0.408	-0.279	0.292	-0.005
SE	0.144	0.146	0.151	0.141	0.103

function $\phi(W)$, i.e., $H_0 : \phi(W) = \gamma_0 + \gamma_1 W$. Calculating the test statistic $\mathcal{T}_n$ for $\nu = 0, 0.3$ and 1, we obtain the corresponding p-values of 0.001, 0.030 and 0.027, respectively. This verifies the adequacy of the model (5.1) we used.

6 Discussion

This paper provides a model framework for a class of partially linear transformation cure models with interval-censored data. The spline-based sieve maximum likelihood method is utilized to estimate parameters and nonparametric components efficiently. Under some regularity conditions, the estimator is shown to be consistent and asymptotic normal. A sieve likelihood ratio test is developed for the significance of regression coefficients as well as the nonparametric component. Simulation studies have been carried out to examine the finite sample performance of the proposed method.

Extending model (2.2) to incorporate an unknown smooth function $\psi(W)$ in the logistic regression model so that

$$P(Y = 1|Z, W) = \frac{\exp(\boldsymbol{\alpha}^\tau Z + \psi(W))}{1 + \exp(\boldsymbol{\alpha}^\tau Z + \psi(W))}$$

would be a topic for future research when the sample size is reasonably large. This extension relaxes the parametric logistic regression assumption on the cure rate which is often hard to be verified in practice. Another interesting extension would be to accommodate two or more potentially non-linear explanatory variables to describe the event time for susceptible

subjects. For example, $H(T)$ in (2.2) can be modeled as

$$H(T) = -\beta^T X - \varphi_1(W_1) - \dots - \varphi_q(W_q) + \epsilon,$$

which yields a partially linear additive transformation cure model. Alternatively, a partially linear time-varying coefficient transformation cure model or partially linear single-index transformation cure model specifies $H(T)$ as

$$H(T) = -\beta^T X - \alpha(W)^T Z + \epsilon \quad \text{or} \quad H(T) = -\beta^T X - g(\gamma^T Z) + \epsilon.$$

An intuitively appealing approach to fit these models would be to approximate the nonparametric functions by splines, and then construct the sieve maximum likelihood estimation as studied in this paper. Further work on rigorous theoretical justification of this approach is warranted.

To decide which covariates enter the linear component and which covariates enter the nonparametric component is always an interesting and challenging topic in partially linear models. A possible way to approach this problem is to determine the covariates in a variable selection framework. Zhang et al. (2011) provided an automatic criterion to distinguish linear and nonlinear components using a penalization procedure in partially linear models with smoothing splines. However, the extension of the method to the context of partially linear transformation cure models for interval-censored data becomes seriously challenging. This would be a separate project, and we leave it for future study. For the moment, we rely on visual inspection to separate the covariates into the linear and nonparametric components in the analysis of smoking cessation data.

Acknowledgements

The authors thank a co-editor, an associate editor and the two anonymous referees for their constructive and valuable comments which are very helpful for improving the manuscript. The research of Xiang is partially supported by the Singapore MOE AcRF (ARC29/14, RG30/12). Hu's research is partially supported by the Natural Science Foundation of China (11201317, 11371062).

A Appendix

This section contains the sketch of proofs for Theorems 3.1-3.4. Throughout the following proofs, we denote $Pf = \int f(x)dP(x)$ and $P_nf = n^{-1} \sum_{i=1}^n f(X_i)$, the empirical process indexed by function $f(X)$. we let C represent a generic constant that may vary from place to place.

Proof of Theorem 3.1

By the calculation of Shen and Wong (1994, p. 597), for any $\varepsilon > 0$, the covering number (definition 2.1.5 in van der Vaart and Wellner, 1996) of the class $\mathcal{F} = \{l(\theta; \mathcal{W}) : \theta \in \Theta_n\}$ satisfies $N(\varepsilon, \mathcal{F}, L_1(P_n)) \leq C l_n^{2N} \varepsilon^{-(2N+p+q)}$. Adopting similar proofs of Theorem 37 in Pollard (1984), we have

$$\sup_{\theta \in \Theta_n} |P_n l(\theta; \mathcal{W}) - Pl(\theta; \mathcal{W})| \rightarrow 0, a.s. \quad (\text{A.1})$$

The proof of this result is given separately in Remark A1. Let $M(\theta; \mathcal{W}) = -l(\theta; \mathcal{W})$, and

$$\zeta_{1n} = \sup_{\theta \in \Theta_n} |P_n M(\theta; \mathcal{W}) - PM(\theta; \mathcal{W})|, \quad (\text{A.2})$$

$$\zeta_{2n} = P_n M(\theta_0; \mathcal{W}) - PM(\theta_0; \mathcal{W}). \quad (\text{A.3})$$

Denote $K_\epsilon = \{\theta : d(\theta, \theta_0) \geq \epsilon, \theta \in \Theta_n\}$.

$$\begin{aligned} \inf_{K_\epsilon} Pl(\theta; \mathcal{W}) &= \inf_{K_\epsilon} \{PM(\theta; \mathcal{W}) - P_n M(\theta; \mathcal{W}) + P_n M(\theta; \mathcal{W})\} \\ &\leq \zeta_{1n} + \inf_{K_\epsilon} P_n M(\theta; \mathcal{W}). \end{aligned} \quad (\text{A.4})$$

If $\hat{\theta}_n \in K_\epsilon$, we have

$$\begin{aligned} \inf_{K_\epsilon} P_n M(\theta; \mathcal{W}) &= P_n M(\hat{\theta}; \mathcal{W}) \\ &\leq P_n l(\theta_0; \mathcal{W}) = P_n M(\theta_0; \mathcal{W}) - PM(\theta_0; \mathcal{W}) + PM(\theta_0; \mathcal{W}) \\ &= \zeta_{2n} + PM(\theta_0; \mathcal{W}). \end{aligned} \quad (\text{A.5})$$

By identification condition (A3), we obtain that $\inf_{K_\epsilon} PM(\theta; \mathcal{W}) - PM(\theta_0; \mathcal{W}) = \delta_\epsilon > 0$.

Based on (A.4) and (A.5), we have

$$\inf_{K_\epsilon} PM(\theta; \mathcal{W}) \leq \zeta_{1n} + \zeta_{2n} + PM(\theta_0; \mathcal{W}) = \zeta_n + PM(\theta_0; \mathcal{W})$$

with $\zeta_n = \zeta_{1n} + \zeta_{2n}$. Hence, we can get that $\zeta_n \geq \delta_\epsilon$. Furthermore, we have $\{\hat{\theta}_n \in K_\epsilon\} \subseteq \{\zeta_n \geq \delta_\epsilon\}$. By (A.1) and the strong law of large number, we have $\zeta_{1n} = o(1)$, $\zeta_{2n} = o(1)$, a.s. Therefore, by $\cup_{k=1}^\infty \cap_{n=k}^\infty \{\hat{\theta}_n \in K_\epsilon\} \subseteq \cup_{k=1}^\infty \cap_{n=k}^\infty \{\zeta_n \geq \delta_\epsilon\}$, we can complete the proof.

Remark A1: Proofs of (A.1)

Choose $\delta_n = 1$, $\nu/2 < \phi_1 < 1/2$, and $\alpha_n = n^{-1/2+\phi_1}(\log n)^{1/2}$ where α_n is a non-increasing series, and for a fixed ϵ , choose $\epsilon_n = \epsilon\alpha_n$, then for any $l(\theta; \mathcal{W}) \in \mathcal{F}$ and n being large enough, we have

$$\text{var}(\mathbb{P}_n l(\theta; \mathcal{W})) / (4\epsilon_n)^2 \leq \frac{(1/n)Pl^2(\theta; \mathcal{W})}{16\epsilon^2\alpha_n^2} \leq \frac{C}{16\epsilon^2n\alpha_n^2} \ll \frac{1}{16\epsilon^2 \log n} < \frac{1}{2},$$

where $\ll$ is defined by Pollard (1984, p. 34) because if $a_n/b_n \rightarrow \infty$ for sequences $\{a_n\}$ and $\{b_n\}$, then $a_n \gg b_n$, or if $a_n/b_n \rightarrow 0$, then $a_n \ll b_n$.

By applying the results of Theorem II.31 of Pollard (1984), it follows that

$$\begin{aligned} & P\left(\sup_{\mathcal{F}} |P_n l(\theta; \mathcal{W}) - Pl(\theta; \mathcal{W})| > 8\epsilon_n\right) \\ & \leq 8N(\epsilon_n, \mathcal{F}, L_1(P_n)) \exp(-n\epsilon_n^2/128) P(\sup_{\mathcal{F}} |P_n l^2(\theta; \mathcal{W})| \leq 64) + P(\sup_{\mathcal{F}} |P_n l^2(\theta; \mathcal{W})| > 64) \\ & \leq 8N(\epsilon_n, \mathcal{F}, L_1(P_n)) \exp(-n\epsilon_n^2/128) \\ & \leq 8C l_n^{2N} \epsilon_n^{-(2N+p+q)} \exp(-n\epsilon_n^2/128) \\ & = 8C \exp\left\{-(2(n^\nu + l))\gamma \log n - (2(n^\nu + l) + p + q) \log(\epsilon n^{-1/2+\phi_1}(\log n)^{1/2}) - n\epsilon^2 n^{-1+2\phi_1} \log n / 128\right\} \\ & \leq 8C \exp\left\{(2(n^\nu + l) + p + q)[(\gamma + 1/2 - \phi_1) \log n - \log \log n / 2 - \log \epsilon] - \epsilon^2 n^{2\phi_1} \log n / 128\right\} \\ & \leq 8C \exp(-C n^{2\phi_1} \log n). \end{aligned}$$

Hence

$$\sum_{n=1}^{\infty} P\left(\sup_{\mathcal{F}} |P_n l(\theta; \mathcal{W}) - Pl(\theta; \mathcal{W})| > 8\epsilon_n\right) < \infty.$$

By the Borel-Cantelli lemma, we have

$$\sup_{\mathcal{F}} |P_n l(\theta, W) - Pl(\theta, W)| \rightarrow 0, a.s. P_{\theta_0}.$$

Hence the proof of (A.1) is completed.

Remark A2: Proofs of Model Identification

Let $f(\mathcal{W}; \theta)$ be the probability density function of $\mathcal{W} = (\delta_1, \delta_2, U, V, Z, X)$ measured at parameter value $\theta = (\alpha, \beta, H, \psi)$. Suppose $f(\mathcal{W}; \theta) = f(\mathcal{W}; \theta_0)$ with probability 1. We

first claim that, to establish the identifiability, it suffices to show that $\theta = \theta_0$. To see this, consider $\delta_1 = 1$, then

$$\frac{\exp(\alpha^\tau Z)}{1 + \exp(\alpha^\tau Z)} F(H(V) + X^\tau \beta + \psi(W)) = \frac{\exp(\alpha_0^\tau Z)}{1 + \exp(\alpha_0^\tau Z)} F(H_0(V) + X^\tau \beta_0 + \psi_0(W)).$$

Let $V = \tau$, by Assumption (A4)(c), then the above equation leads to

$$\frac{\exp(\alpha^\tau Z)}{1 + \exp(\alpha^\tau Z)} = \frac{\exp(\alpha_0^\tau Z)}{1 + \exp(\alpha_0^\tau Z)}.$$

It follows from the monotonicity of $\exp(x)/(1 + \exp(x))$ and Assumption (A4)(a) that $\alpha = \alpha_0$. Considering the case in which $\delta_1 = \delta_2 = 0$, by Assumption (A4)(b) and the monotonicity of $F(\cdot)$, we can get $H(U) + X^\tau \beta + \psi(W) = H_0(U) + X^\tau \beta_0 + \psi_0(W)$. Therefore $X^\tau(\beta - \beta_0) = H_0(U) - H(U) + \psi_0(W) - \psi(W)$. by (A4)(a), we have $\beta = \beta_0$, $H_0(U) - H(U) + \psi_0(W) - \psi(W) = 0$. So $H_0(U) + \psi_0(W) = H(U) + \psi(W)$. Then taking the partial derivative with respect to U results in $H'(U) = H'_0(U)$. Because $H(0) = 0$, $H(U) = \int_0^U H'(u) du = \int_0^U H'_0(u) du = H_0(U) - H_0(0) = H_0(U)$. Thus $\psi_0(W) = \psi(W)$. This completes the proof.

Proof of Theorem 3.2

For any $\eta > 0$, define the class $\mathcal{F}_\eta = \{l(\theta_{n0}; \mathcal{W}) - l(\theta; \mathcal{W}) : \theta \in \Theta_n, d(\theta, \theta_{n0}) \leq \eta\}$ with $\theta_{n0} = (\alpha_0, \beta_0, H_{n0}, \psi_{n0})$. Following the calculation of Shen and Wong (1994), we can establish that $\log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \leq CN \log(\eta/\varepsilon)$. Moreover, some algebraic calculations lead to $\|l(\theta_{n0}; \mathcal{W}) - l(\theta; \mathcal{W})\|_2^2 \leq C\eta^2$ for any $l(\theta_{n0}; \mathcal{W}) - l(\theta; \mathcal{W}) \in \mathcal{F}_\eta$.

Therefore, by Lemma 3.4.2 of van der Vaart and Wellner (1996), we obtain

$$E_P \|n^{1/2}(P_n - P)\|_{\mathcal{F}_\eta} \leq C J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) \left\{ 1 + \frac{J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)}{\eta^2 n^{1/2}} \right\}, \quad (\text{A.6})$$

where $J_\eta(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2) = \int_0^\eta \{1 + \log N_{[]}(\varepsilon, \mathcal{F}_\eta, \|\cdot\|_2)\}^{1/2} d\varepsilon \leq CN^{1/2}\eta$. The right-hand side of (A.6) yields $\phi_n(\eta) = C(N^{1/2}\eta + N/n^{1/2})$. It is easy to see that $\phi_n(\eta)/\eta$ decreasing in η , and $r_n^2 \phi_n(1/r_n) = r_n N^{1/2} + r_n^2 N/n^{1/2} < 2n^{1/2}$, where $r_n = N^{-1/2} n^{1/2} = n^{(-\nu+1)/2}$, $0 < \nu < 1/2$. Hence $n^{(1-\nu)/2} d(\hat{\theta}, \theta_{n0}) = O_P(1)$ by Theorem 3.2.5 of van der Vaart and Wellner (1996). This, together with $d(\theta_{n0}, \theta_0) = O_p(n^{-r\nu})$ (see Theorem 12.7 in Schumaker, 1981), yields that $d(\hat{\theta}, \theta_0) = O_p(n^{-(1/2-\nu)} + n^{-r\nu})$. This completes the proof.

Proof of Theorem 3.3

Let ε_n be any positive sequence satisfying $\varepsilon_n = o(n^{-1/2})$. For any $v^* \in \Theta_0$, by Theorem 12.7 of Schumaker (1981), there exists $\Pi_n v^* \in \Theta_n$ such that $\|\Pi_n v^* - v^*\| = o(1)$ and

$\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$. Also define $\rho[\theta - \theta_0; \mathcal{W}] = l(\theta; \mathcal{W}) - l(\theta_0; \mathcal{W}) - \dot{l}(\theta; \mathcal{W})[\theta - \theta_0]$.

Then by the definition of $\hat{\theta}$, we have

$$\begin{aligned}
0 &\leq P_n[l(\hat{\theta}; \mathcal{W}) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*; \mathcal{W})] \\
&= (P_n - P)[l(\hat{\theta}; \mathcal{W}) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*; \mathcal{W})] + P[l(\hat{\theta}; \mathcal{W}) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*; \mathcal{W})] \\
&= \pm \varepsilon_n P_n \dot{l}(\theta; \mathcal{W})[\Pi_n v^*] + (P_n - P) \left\{ \rho[\theta - \theta_0; \mathcal{W}] - \rho[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0; \mathcal{W}] \right\} \\
&\quad + P \left\{ \rho[\theta - \theta_0; \mathcal{W}] - \rho[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0; \mathcal{W}] \right\} \\
&= \mp \varepsilon_n P_n \dot{l}(\theta; \mathcal{W})[v^*] \pm \varepsilon_n P_n \dot{l}(\theta; \mathcal{W})[\Pi_n v^* - v^*] + (P_n - P) \left\{ \rho[\theta - \theta_0; \mathcal{W}] - \rho[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0; \mathcal{W}] \right\} \\
&\quad + P \left\{ \rho[\theta - \theta_0; \mathcal{W}] - \rho[\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0; \mathcal{W}] \right\} \\
&:= \mp \varepsilon_n P_n \dot{l}(\theta; \mathcal{W})[v^*] + I_1 + I_2 + I_3.
\end{aligned}$$

By (A1) and Chebyshev's inequality, and $\|\Pi_n v^* - v^*\| = o(1)$, we have $I_1 = \varepsilon_n \times o_p(n^{-1/2})$.

For I_2 , we have

$$\begin{aligned}
I_2 &= (P_n - P) \left\{ l(\hat{\theta}; \mathcal{W}) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*; \mathcal{W}) \pm \varepsilon_n \dot{l}(\theta_0; \mathcal{W})[\Pi_n v^*] \right\} \\
&= \mp \varepsilon_n (P_n - P) \left\{ \dot{l}(\tilde{\theta}; \mathcal{W}) - \dot{l}(\theta_0; \mathcal{W})[\Pi_n v^*] \right\},
\end{aligned}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta} \pm \varepsilon_n \Pi_n v^*$. By Theorem 2.8.3 in of van der Vaart and Wellner (1996), we know that $\{\dot{l}(\theta; \mathcal{W})[\Pi_n v^*] : \|\theta - \theta_0\| = O(\delta_n)\}$ is Donsker class. Therefore, by Theorem 2.11.23 of van der Vaart and Wellner (1996), we have $I_2 = \varepsilon_n \times o_p(n^{-1/2})$.

Note that

$$\begin{aligned}
P(\rho[\theta - \theta_0; \mathcal{W}]) &= P\{l(\theta; \mathcal{W}) - l(\theta_0; \mathcal{W}) - \dot{l}(\theta_0; \mathcal{W})[\theta - \theta_0]\} \\
&= 2^{-1} P\{\ddot{l}(\tilde{\theta}; \mathcal{W})[\theta - \theta_0, \theta - \theta_0] - \ddot{l}(\theta_0; \mathcal{W})[\theta - \theta_0, \theta - \theta_0]\} \\
&\quad + 2^{-1} P\{\ddot{l}(\theta_0; \mathcal{W})[\theta - \theta_0, \theta - \theta_0]\} \\
&= 2^{-1} P\{\ddot{l}(\theta_0; \mathcal{W})[\theta - \theta_0, \theta - \theta_0]\} + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where $\tilde{\theta}$ lies between θ_0 and θ and the last equation is due to Taylor expansion, (A1)-(A3) and $r \geq 2$. Therefore,

$$\begin{aligned}
I_3 &= -2^{-1} \{ \|\hat{\theta} - \theta_0\|^2 - \|\hat{\theta} \pm \varepsilon_n \Pi_n v^* - \theta_0\|^2 \} + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, \Pi_n v^* \rangle + 2^{-1} \|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + 2^{-1} \|\varepsilon_n \Pi_n v^*\|^2 + \varepsilon_n \times o_p(n^{-1/2}) \\
&= \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2})
\end{aligned}$$

where the last equality holds since $\delta_n \|\Pi_n v^* - v^*\| = o(n^{-1/2})$, Cauchy-Schwartz inequality, and $\|\Pi_n v^*\|^2 \rightarrow \|v^*\|^2$. Combing the above facts, together with $Pl(\theta_0; \mathcal{W})[v^*] = 0$, we can establish that

$$\begin{aligned} 0 &\leq P_n\{l(\hat{\theta}; \mathcal{W}) - l(\hat{\theta} \pm \varepsilon_n \Pi_n v^*; \mathcal{W})\} \\ &= \mp \varepsilon_n P_n \dot{l}(\theta_0; \mathcal{W})[v^*] \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}) \\ &= \mp \varepsilon_n (P_n - P)\{\dot{l}(\theta_0; \mathcal{W})[v^*]\} \pm \varepsilon_n \langle \hat{\theta} - \theta_0, v^* \rangle + \varepsilon_n \times o_p(n^{-1/2}). \end{aligned}$$

Therefore, we obtain $\sqrt{n} \langle \hat{\theta} - \theta_0, v^* \rangle = \sqrt{n}(P_n - P)\{\dot{l}(\theta_0; \mathcal{W})[v^*]\} + o_p(1) \rightarrow N(0, \|v^*\|^2)$, where the asymptotic normality is guaranteed by Central Limits Theorem and the asymptotic variance being equal to $\|v^*\|^2 = \|\dot{l}(\theta_0; \mathcal{W})[v^*]\|^2$. This, together with (A5) imply $n^{1/2}(\rho(\hat{\theta}) - \rho(\theta_0)) = n^{1/2} \langle \hat{\theta} - \theta_0, v^* \rangle + o_p(1) \rightarrow N(0, \|v^*\|^2)$ in distribution. The semiparametric efficiency can be established by applying the result of Bickel and Kwon (2001).

Proof of Theorem 3.4

It is quite similar to the proof of Theorem 2.1 in Shen and Shi (2005), we omit it here.

References

- Bickel, P. and J. Kwon (2001). Inference for semiparametric models: some questions and an answer. *Statistica Sinica* **11**, 863–960.
- Banerjee, S., Carlin, B. P. (2004). Parametric spatial cure rate models for interval-censored time-to-relapse data. *Biometrics* **60**, 268–275.
- Cheng, G. and Huang, J. (2010). Bootstrap Consistency for General Semiparametric M-estimation. *Annals of Statistics*, **38**, 2884–2915.
- Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with longterm survivors. *Biometrics* **38**, 1041–1046.
- Fine, J.P. (1999). Analysing competing risks data with transformation models. *Journal of the Royal Statistical Society, Ser. B* **61**, 817–830.
- Geman, A. and Hwang, C. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Annals of Statistics*, **10**, 401–414.

- Grenander, U. (1981). *Abstract Inference*. New York: Wiley.
- Gu, Y, Sinha, D and Banerjee, S. (2011). Analysis of cure rate survival data under proportional odds model. *Lifetime Data Analysis* **17**, 123–134.
- Hu, T and Xiang L. (2013). Efficient estimation for semiparametric cure models with interval-censored data. *Journal of Multivariate Analysis*, **121**, 139–151.
- Huang, J. and Rossini, A. J. (1997). Sieve estimation for the proportional odds failure-time regression model with interval censoring. *Journal of the American Statistical Association* **92**, 960–967.
- Kuk, A. Y. C. and Chen, C. H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79**, 531–541.
- Lam, K. F. and Wong, K.-Y. (2014). Semiparametric analysis of clustered interval-censored survival data with a cure fraction. *Computational Statistics & Data Analysis* **79**, 165–174.
- Lam, K. F. and Xue, H. (2005). A semiparametric regression cure model with current status data. *Biometrika* **92**, 573–586.
- Li, J. and Ma, S. (2010). Interval Censored Data with Repeated Measurements and a Cured Subgroup. *Journal of the Royal Statistical Society Ser. C (Applied Statistics)* **59**, 693–705.
- Li, C. S., Taylor, J. M. G. and Sy, J. P. (2001). Identifiability of cure models. *Statist. Probab. Lett.* **54**, 389–395.
- Liu, H. and Shen, Y. (2009). A semiparametric regression cure model for interval censored data. *Journal of the American Statistical Association* **104**, 1168–1178.
- Lu, W. and Ying, Z. (2004). On semiparametric transformation cure models. *Biometrika* **91**, 331–343.
- Ma, S. (2010). Mixed case interval censored data with a cured subgroup. *Statistica Sinica* **20**, 1165–1181.

- Ma, S. and Kosorok, M. R. (2005). Penalized Log-Likelihood Estimation for Partly Linear Transformation Models With Current Status Data. *Annals of Statistics* **33**, 2256–2290.
- Mao, M. and Wang, J.-L. (2010). Semiparametric Efficient Estimation for a Class of Generalized Proportional Odds Cure Models. *Journal of the American Statistical Association* **105**, 302–311.
- Murphy, S.A., Rossini, A.J. and van der Vaart, A.W. (1997). Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association* **92**, 968–976.
- Murray, R., Anthonian, N.R., Connett, J.E., Wise, R.A., Lindgren, P.G., Green, P.G., Nides, M.A. (for the Lung Health Study Research Group) (1998). Effects of multiple attempts to quit smoking and relapses to smoking on pulmonary function. *Journal of Clinical Epidemiology* **51**, 1317–1326.
- Othus, M., Li, Y. and Tiwari, R.C. (2009). A class of semiparametric mixture cure survival models with dependent censoring. *J. American Statistical Association* **104**, 1241–1250.
- Pollard, D. (1984), *Convergence of Stochastic Processes*. Springer, New York.
- Schumaker, L. (1981). *Spline Functions: Basic Theory*. Wiley, New York.
- Shen, X. (1997). On methods of sieves and penalization. *Annals of Statistics* **25**, 2555–2591.
- Shen, X., and Shi, J. (2005). Sieve Likelihood ratio inference on general parameter space. *Science in China Series A: Mathematics*, **48**(1), 67–78.
- Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. *Annals of Statistics* **22**, 580–615.
- Stone, C. (1980). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* **10** 1040–1053.
- Sun, J. (2006). *The Statistical Analysis of Interval-censored Failure Time Data*. Springer.

- Tsodikov, A., Ibrahim, J. G. and Takovlev. A. (2003). Estimating cure rates from survival data: An alternative to two-component mixture models. *Journal of the American Statistical Association* **98**, 1063–1078.
- Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- Xiang L., Ma, X. and Yau K.K.W. (2011). Mixture cure model with random effects for clustered and interval-censored survival data. *Statistics in Medicine*, **30**, 995–1006.
- Yakovlev, A.Y. and Tsodikov, A.D. (1996). *Stochastic models of tumor latency and their biostatistical applications*. World Scientific, Singapore.
- Yin, G. (2008). Bayesian transformation cure frailty models with multivariate failure time data. *Statistics in Medicine* **27**, 5929–5940.
- Yu, B. and Peng, Y. (2008). Mixture cure models for multivariate survival data. *Computational Statistics & Data Analysis* **52**, 1524–1532.
- Zeng, D., Yin, G. and Ibrahim, J.G. (2006). Semiparametric Transformation Models for Survival Data with a Cure Fraction. *Journal of the American Statistical Association* **101**, 670–684.
- Zhang, H., Cheng, G. and Liu, Y. (2011). Linear or nonlinear? automatic structure discovery for partially linear models. *Journal of the American Statistical Association* **106**, 1099–1112.
- Zhang, Y., Hua, L., and Huang, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scandinavian Journal of Statistics* **37**, 338–354.