Lifetime Data Analysis (2020) 26:518–544
https://doi.org/10.1007/s10985-019-09488-8

Varying coefficient transformation cure models for failure time data

Man-Hua Chen[1] · Xingwei Tong[2]

Received: 29 July 2015 / Accepted: 25 September 2019 / Published online: 9 October 2019

Abstract
This article discusses regression analysis of right-censored failure time data where there may exist a cured subgroup, and also covariate effects may be varying with time, a phenomena that often occurs in many medical studies. To address the problem, we discuss a class of varying coefficient transformation models along with a logistic model for the cured subgroup. For inference, a sieve maximum likelihood approach is developed with the use of spline functions, and the asymptotic properties of the proposed estimators are established. The proposed method can be easily implemented, and the conducted simulation study suggests that the proposed method works well in practical situations. An illustrative example is provided.

Keywords Cure model · Maximum likelihood estimation · Regression analysis · Spline smoothing

1 Introduction

Failure time data commonly occur in many areas such as economic and medical studies as well as social science, and a large literature has been established for their analysis (Kalbfleisch and Prentice 2002; Klein and Moeschberger 2003). In a typical failure time study, an underlying assumption is that every subject in the study is susceptible to the event of interest. In reality, however, this may not be true. For this situation, we usually say that there exists a cured subgroup, and in this paper, we will discuss regression analysis of such failure time data.

Some methods have been developed for the analysis of failure time data with a cured subgroup (Choi et al. 2014; Demarqui et al. 2014; Kuk and Chen 1992; Zeng et al. 2006). For example, one of the early works was proposed by Farewell (1982), who

✉ Man-Hua Chen
mchen@mail.tku.edu.tw

[1] Department of Statistics, Tamkang University, Tamsui 25137, Taiwan

[2] School of Statistics, Beijing Normal University, Beijing 100875, China

considered the logistic model to estimate the cured probability. Lu and Ying (2004) gave some estimating equation approaches for the failure times arising from linear transformation models, and Lu (2010) investigated the problem under the accelerated failure time model. In addition, Wang et al. (2012) considered the nonparametric spline estimates for both nonsusceptible and susceptible individuals, and Chen et al. (2013) discussed the association estimation for clustered failure time data. The point is that all methods mentioned above assume that covariate effects are constant or time-independent, and it is apparent that this may not be true.

Varying coefficient models have been discussed under many different contexts including nonparametric regression, generalized linear models, nonlinear time series models, and longitudinal and functional data analysis (e.g., Hastie and Tibshirani 1990; Fan and Zhang 1999; Cai et al. 2000; Fan et al. 2006; Cai et al. 2007). For these situations, a primary assumption is that treatment effects are usually nonlinear or time-dependent. Chen and Tong (2010), among others, developed a class of varying coefficient transformation models for regression analysis of failure time data, in the absence of a cured subgroup. In this article, we present a varying-coefficient mixture model that allows covariates to interact with each other nonlinearly, for both susceptible and nonsusceptible subjects. The features of the kidney transplant data (Klein and Moeschberger 2003) are the possible existence of a cured subgroup and non-linear interaction of gender or race in terms of age. In Sect. 6, we find that the data has a nonsusceptible subgroup and cross survival curves. We believe that varying coefficients achieve the purpose of capturing the non-linearly changing effect of gender to age or race to age, and a cure rate achieves the purpose of capturing the existing nonsusceptible subgroup.

The rest of the article is organized as follows. In Sect. 2, we first introduce some notation and assumptions and describe the assumed models. A sieve maximum likelihood estimation procedure is then developed in Sect. 3, and an adaptive EM algorithm is also provided for determining the proposed estimator. Section 4 establishes the asymptotic behavior of the proposed estimator, and Sect. 5 gives some results of a simulation study for evaluating the finite sample performance of the proposed method, indicating that the proposed method works well. Section 6 provides an illustrative example, and Sect. 7 gives some discussion and concluding remarks.

2 Notation, models and assumptions

Consider a failure time study that may include a cured subgroup and let T denote the underlying failure time of interest. In the following, we assume that T can be described as

$$T = \eta T^* + (1 - \eta) \infty . \tag{1}$$

In the above, T^* denotes the failure time for the subject who is susceptible to the failure event of interest and η takes values 1 or 0 indicating whether a subject is susceptible (by 1) or not. The model above is commonly referred to as the two-component mixture cure model (Farewell 1982; Kuk and Chen 1992), assuming that the population consists of an uncured group and a cured group. One advantage of model (1) is that it is intuitively attractive and gives easy interpretations. An alternative to model (1) is the so-called promotion time cure model, which formulates both cured and noncured subjects in a single survival function and focuses on the combined population (Yakovlev and Tsodikov 1996; Tsodikov 1998).

To model T^*, let (X, Z, W) denote the covariates with X being a p-dimensional vector and Z and W being scalars. We will assume that T^* follows the following varying coefficient model

$$\log H(T^*) = -X^T\beta - Z\psi(W) + \epsilon\,. \tag{2}$$

Here $H(.)$ is an unknown, strictly increasing function with $H(0) = 0$; $\psi(.)$ is also an unknown function, β denotes regression parameters, and ϵ is the error with a known continuous density function f_ϵ, which is assumed to be independent of the covariates. Note that Z denotes the covariate that, for the given values of W, has a linear relationship with the mean of the response variable $\log H(T^*)$. Also note that Z can include the intercept term, that is $Z = (1, Z^*)^T$, where Z^* is a scalar variable of interest; $\psi(W) = (\psi_1(W), \psi_2(W))^T$ is a two-dimensional unknown function. $\psi_1(W)$ is the main or baseline effect of W, and $\psi_2(W)$ is the interaction of Z^* and W. All the theorems for $Z = (1, Z^*)^T$ and $\psi(W) = (\psi_1(W), \psi_2(W))^T$ hold true. Chen and Tong (2010) considered the same model (2) for the situation without a cured group, and one can find more discussion on the model there and below. Without Z and W, model (2) is commonly referred to as the linear transformation model and has been extensively investigated for the analysis of failure time data. For the cure probability η in model (1), it will be assumed that it follows the following logistic model given covariates U.

$$P(\eta = 1|U) = \pi(U\gamma) = \frac{e^{U^T\gamma}}{1 + e^{U^T\gamma}}\,. \tag{3}$$

Here γ is a d-dimensional vector of regression parameters, and U denotes the covariates that may have effects on η and could be different from or a part of (X, Z, W). Let Λ denote the cumulative hazard function of ϵ. Then under the assumptions above, we have

$$\begin{aligned} P(T \geq t|X, Z, W, U) &= P(T \geq t, \eta = 0|X, Z, W, U) + P(T \geq t, \eta = 1|X, Z, W, U) \\ &= 1 - \pi(U^T\gamma) + \pi(U^T\gamma)\exp(-\Lambda[\log H(t) + X^T\beta + Z\psi(W)])\,. \end{aligned}$$

For the description of the observed data, suppose that there exists a right-censoring time denoted by C, assumed to be independent of T^* and η given covariates (X, Z, W). Also suppose that there exists an administrative stopping time denoted by τ and define $\tilde{T} = \min(T, \min\{C, \tau\})$ and $\delta = I(\tilde{T} \leq \min\{C, \tau\})$, which are the observed failure time and censoring indicator, respectively. Then the observed data consist of $\{O_i = (X_i, Z_i, W_i, U_i, \tilde{T}_i, \delta_i) : i = 1, \ldots, n\}$, and the i.i.d. copies of $O = (X, Z, W, U, \tilde{T}, \delta)$. In general, η is unobservable, but if $\delta = 1$, η equals one. Define $\theta = (\beta, \gamma, \psi, H)$. Then the log likelihood function of θ has the form

$$
\begin{aligned}
l_n(\theta) = \sum_{i=1}^{n} \delta_i \left(\log \pi(U_i^T \gamma) - \Lambda[V_i(\theta)] + \log \lambda[V_i(\theta)] + \log \frac{\Delta H(\tilde{T}_i)}{H(\tilde{T}_i)} \right) \\
+(1-\delta_i) \log\left(1 - \pi(U_i^T \gamma) + \pi(U_i^T \gamma) \exp(-\Lambda[V_i(\theta)]) \right), \qquad (4)
\end{aligned}
$$

where $\lambda(t) = d\Lambda(t)/dt$, $V_i(\theta) = \log H(\tilde{T}_i) + X_i^T \beta + Z_i \psi(W_i)$ and $\Delta H(t) = H(t) - H(t^-)$. Here we will restrict $H(t)$ to be a nondecreasing step function with $H(0) = 0$ and it jumps only at $t_1 < \cdots < t_m$, the observed true failure times, where $t_1 > 0, t_m < \tau$ and $m = \sum_{i=1}^n \delta_i$. Note that in the log likelihood function above, the term log(0) may occur but does not have any effect since one only needs to calculate $\Delta H(t)$ and related terms at the distinct observed event times, where $\delta_i = 1$. For completeness, we will define log(0) and also in the following, without loss of generality, we assume that W has the support on [0, 1].

3 Sieve maximum likelihood estimation

In this section we discuss estimation of the parameters $\theta = (\beta, \gamma, \psi, H)$ and present a sieve maximum likelihood estimation procedure. For this, assume that $(\beta^T, \gamma^T)^T \in \mathcal{B}$, a bounded open subset of R^{p+d}, and define

$$
\Psi_r = \{\psi : |\psi^{(l)}(w_1) - \psi^{(l)}(w_2)| \leq \mathcal{M}|w_1 - w_2|^{r-l}, \text{ for any } w_1, w_2 \in [0, 1]\},
$$

and let $\mathcal{H} = \{H : H(\cdot)$ be a nondecreasing right continuous function, $H(\tau) < \mathcal{M}\}$, where $\mathcal{M} > 0$ is a constant and $r > 1/2$. Also let $K = K_n$ be the integer part of n^v with $0 < v < 0.5$ and $\{B_i(\cdot), i = 1, \ldots, q_n\}$ denote the normalized B-spline basis functions in the space of B-spline functions of order $l+1$ with $q_n = K_n + l$ and the knots $0 = \xi_0 < \xi_1 < \cdots < \xi_{K_n - 1} < \xi_{K_n} = 1$, satisfying $\max(\xi_j - \xi_{j-1} : j = 1, \ldots, K_n) = O(n^{-v})$ (Schumaker 1981).

Define $B_n(\cdot) = \{B_1(\cdot), \ldots, B_{q_n}(\cdot)\}^T$, a vector of q_n-dimensional functions. We will define the estimator, denoted by $\hat{\theta}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\psi}_n(w) = B_n(w)\,\hat{\alpha}_n, \hat{H}_n)$, of θ as the value that maximizes the log likelihood function $l_n(\theta)$ over $\Theta = \mathcal{B} \times \Psi_r \times \mathcal{H}$ with $\psi(w) = B_n(t)\,\alpha$. For the determination of $\hat{\theta}_n$, it is apparent that the direct maximization of $l_n(\theta)$ may not be easy to do due to the large number of parameters involved. Also the score functions for β, γ and α are actually quite complex. To deal with this, we will develop an EM algorithm.

For the EM algorithm, we assume that the η_i's are known and treat them and the observed data O_i's together as pseudo-complete data. The pseudo likelihood function then has the form

$$
\begin{aligned}
L^{EM}(\theta) = \prod_{i=1}^{n} \left(\pi(U_i^T \gamma) \exp(-\Lambda(V_i^*(\theta)))\lambda(V_i^*(\theta)) \frac{\Delta H(\tilde{T}_i)}{H(\tilde{T}_i)} \right)^{\delta_i \eta_i} \\
\times \left(1 - \pi(U_i^T \gamma) \right)^{(1-\delta_i)(1-\eta_i)} \left(\pi(U_i^T \gamma) \exp(-\Lambda(V_i^*(\theta))) \right)^{\eta_i (1-\delta_i)},
\end{aligned}
$$

where $V_i^*(\theta) = \log H(\tilde{T}_i) + X_i^{*T}\theta^*$, $X^{*T} = (Z, B_n(W))$, and $\theta^{*T} = (\beta, \alpha)$. It follows that the log-likelihood function is

$$l^{EM}(\theta) = \sum_{i=1}^{n} \eta_i \log\{\pi(U_i^T\gamma)\} + (1-\delta_i)(1-\eta_i)\log\{1-\pi(U_i^T\gamma)\} + \delta_i\eta_i$$
$$\Big\{ -\Lambda(V_i^*(\theta)) + \log\lambda(V_i^*(\theta)) + \log\Delta H(\tilde{T}_i) - \log H(\tilde{T}_i)\Big\} - (1-\delta_i)\eta_i\Lambda\{V_i^*(\theta)\}\,.$$

For the E-step of the EM algorithm, it is easy to see that one needs to evaluate the conditional expectation $E(\eta_i|O_i, \theta)$ for given θ. For this, given O_i and θ, we have

$$P(\eta_i = 1|X_i, Z_i, W_i, \tilde{T}_i, \delta_i = 0, \theta) = \frac{\pi(U_i^T\gamma)\exp(-\Lambda(V_i^*(\theta)))}{\pi(U_i^T\gamma)\exp(-\Lambda(V_i^*(\theta))) + 1 - \pi(U_i^T\gamma)}$$

and $P(\eta_i = 1|X_i, Z_i, W_i, \tilde{T}_i, \delta_i = 1, \theta) = 1$.

Denote the expectation of η conditional on O, θ with respect to $l^{EM}(\theta)$ by $\tilde{l}^{EM}(\theta)$. For the M step, we need to maximize $\tilde{l}^{EM}(\theta)$. The EM algorithm for the determination of $\widehat{\theta}_n = (\widehat{\beta}_n, \widehat{\gamma}_n, \widehat{\psi}_n(w) = B_n(w)\widehat{\alpha}_n, \widehat{H}_n)$ can be summarized as follows, where $\widehat{H}_n = \sum_{i=1}^m h_j I(t \geq t_j)$, $h_j = \Delta H(t_j) = H(t_j) - H(t_{j-})$, $j = 1 \cdots m$.

Computational algorithm

- Step 0: Take 0 as the initial values for β and γ, $inv(B_n^T B_n)B_n^T\psi_n(w)$ for α and $1/n$ for h_j with n samples, where B_n is the cubic B-splines with number of knots K. Take $2\sin(2w + 0.1) + \exp(-0.5w)$ as the initial function for $\psi_n(w)$.
- Step 1: At the sth iteration, compute the expectation $E(\eta_i|O_i, \theta^{(s)})$.
- Step 2: Determine the updated estimate $\widehat{h}_j^{(s+1)}$ of h_j by solving the expectation of the first derivatives h_j of $l^{EM}(\beta^{(s)}, \gamma^{(s)}, \psi^{(s)}, H^{(s)})$ setting to be equal to zero.
- Step 3: Determine the updated estimate $\widehat{\gamma}_n^{(s+1)}$ of γ by solving the expectation of the first derivatives γ of $l^{EM}(\beta^{(s)}, \gamma^{(s)}, \psi^{(s)}, H^{(s+1)})$ setting to be equal to zero.
- Step 4: Determine the updated estimate $\widehat{\beta}_n^{(s+1)}$ and $\widehat{\alpha}_n^{(s+1)}$ of β and α by solving the expectation of the first derivatives β and α of $l^{EM}(\beta^{(s)}, \gamma^{(s+1)}, \psi^{(s)}, H^{(s+1)})$ setting to be equal to zero.
- Step 5: Repeat Steps 1–4 above until convergence.

Note that the procedure described above divides the equations into three parts to avoid performing optimization algorithms in a high dimensional situation, and some comments on the computational algorithm of nonparametric hazard function h_j can be found in Zeng and Lin (2006). For the implementation of the estimation procedure above, it is apparent that one needs to choose $l + 1$, the order of B-spline functions. In general, cubic splines ($l = 3$) are good enough to approximate unknown functions smoothly, and it is also common to use linear ($l = 1$) or quadratic ($l = 2$) splines, especially for less smooth functions. The number of knots for the B-spline is for controlling the roughness of functions and one can choose the optimal choice of knots

(K) through the BIC criterion as discussed in Sect. 6 among others. In the numerical studies below, MATLAB is used for implementing the algorithm here.

4 Asymptotic properties

Now we establish some asymptotic properties of the estimator $\hat{\theta}_n$. For this, let θ_0 denote the true value of θ and define

$$\rho(\theta_1, \theta_2) = \|\beta_1 - \beta_2\| + \|\gamma_1 - \gamma_2\| + |\psi_1 - \psi_2|_\infty + |H_1 - H_2|_\infty$$

for $\theta_j = (\beta_j, \gamma_j, \psi_j, H_j) \in \Theta$, $j = 1, 2$, where $\|\cdot\|$ denotes the Euclidean norm. First we give the consistency of $\hat{\theta}_n$. In the following, all limits are with respect to $n \to \infty$.

Theorem 1 *Suppose that the conditions (C1)–(C7) given in the "Appendix" hold, then we have*

$$\rho(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-v)/2} + n^{-rv})$$

for $0 < v < 0.5$ *and* $r > 0.5$.

The theorem above shows that the estimator $\hat{\theta}_n$ not only is consistent, but also achieves the optimal convergence rate. This is because with $v = 1/(2r + 1)$, the convergence rate of $\hat{\psi}_n(\cdot)$ is equal to $n^{-r/(2r+1)}$, the optimal global convergence rate of the nonparametric regression estimators (Stone 1980, 1982). To present the asymptotic distribution, let $\hat{\xi}_n = (\hat{\beta}_n^T, \hat{\gamma}_n^T)^T$ and ξ_0 denote the true value of $\xi = (\beta^T, \gamma^T)^T$.

Theorem 2 *Suppose that* $0.25/r < v < 0.5$ *and the conditions (C1)–(C9) given in the "Appendix" hold, then we have*

$$n^{1/2}(\hat{\xi}_n - \xi_0) = \frac{1}{\sqrt{n}} I^{-1}(\xi_0) \sum_{i=1}^{n} \dot{l}_\xi^*(\xi_0) + o_p(1) \to N(0, I^{-1}(\xi_0))$$

in distribution, where $\dot{l}_\xi^*(\xi_0)$ *and* $I(\xi_0)$ *denote the efficient score and information matrix of* ξ*, respectively, and both are given in the "Appendix".*

The proofs of the results above are sketched in the "Appendix". In addition to the asymptotic normality, the theorem above also indicates that the estimator $\hat{\xi}_n$ is asymptotically efficient. Although one can derive the information matrix $I(\xi_0)$, it would be very difficult to give a consistent estimator of $I(\xi_0)$. Thus to estimate the covariance matrix of $\hat{\xi}_n$, by following Zeng et al. (2005) and Chen and Tong (2010), we suggest to treat the problem as a parametric estimation problem and to employ the observed Fisher information matrix of all parameters. More specifically, let Σ_n denote the $(p + d + q_n + m) \times (p + d + q_n + m)$ negative Hessian matrix of $l_n(\beta, \gamma, \alpha, H)$ evaluated at the estimator $\hat{\theta}_n$. Then the covariance matrix of $\hat{\xi}_n$ can be consistently estimated by the upper-left $(p + d) \times (p + d)$ sub-matrix of Σ_n^{-1}.

5 A simulation study

This section presents some of the results obtained from the simulation study to evaluate the limited sample performance of the sieve maximum likelihood estimation procedure presented in the previous sections. In the study, we take $H(t) = t/2$ and $\psi(w) = 2\sin(2w + 0.1) + \exp(-0.5w)$ in model (2) and generate X and Z from the Bernoulli distribution with the probability of success 0.5 and the uniform distribution [0, 2], respectively. We use the cubic B-splines for $B_n(\cdot)$ with the number of knots being $1.5n^{1/3}$ and the knots chosen to be equally spaced over [0, 1]. In addition, we consider the effect of linear covariates and varying covariates on cure rate, as describe below and discuss in Sect. 7.

5.1 Linear covariate effects on the cure rate

We set $U = (1, X)^T$ and generate W, considering whether W is independent of Z. One is (a) assuming W is independent of Z and follows a uniform distribution [0, 1], and the other is (b) assuming W depends on Z; $W \sim Unif[0, 0.5]$ if $Z \leq 1$, and $W \sim Unif[0.5, 1]$ otherwise. Tables 1 and 2 present the results on estimation of β and γ with $\beta = 0$ or 1, and $\nu = 0, 0.5$ or 1. We consider γ with three different cases of $(\gamma_1, \gamma_2) : (0.6, 0.6), (-0.1, -0.1)$, or $(-0.5, -0.5)$, which represents that the treatment group ($X = 1$) has a low cure rate (23%), moderate cure rate (55%), or high cure rate (73%), respectively. In each case, we consider $n = 200$ or 400 with 1000 replications. The results include the average of the actual values subtracted from the estimated mean (Bias), estimated sample standard derivation (SSD), estimated standard error mean (ESE), and empirical 95% coverage probability (CP). The results in Table 1 are $n = 200$, and Table 2 corresponds to the case of $n = 400$.

5.2 Varying covariate effects on the cure rate

The generations of X, Z and W are the same set-up shown in Sect. 5.1. We consider varying covariate effects on the cure rate, $P(\eta = 1|U_1, U_2, W^*) = \frac{e^{U_1^T\gamma + U_2\psi^*(W^*)}}{1+e^{U_1^T\gamma + U_2\psi^*(W^*)}}$. We take $\psi^*(w^*) = 8w^*(1 - w^{*'}w^*)$, set $U_1 = (1, X)^T$, $U_2 = Z$, and generate W^* in two scenarios. The first scenario is that (c) both W and W^* are independent of Z and follow $Unif[0, 1]$. The second scenario is that (d) both W and W^* follow $Unif[0, 0.5]$ if $Z \leq 1$, then $Unif[0.5, 1]$ otherwise. We consider the similar parameter settings mentioned in Sect. 5.1 above. Based on 1000 replicates, the results are shown in Table 3, $n = 400$.

In addition, we assume that the censoring time C follows the uniform distribution $(0, \tau)$, and the stop time τ is set to $min(40, max(T^*))$. We also specify that the general form of the baseline hazard function is $\exp(t)/\{1 + \nu\exp(t)\}$, where τ represents the maximum follow-up time and ν is a constant. In the case of $\nu = 0$, model (2) gives the proportional hazard model, which gives a proportional odds model with $\nu = 1$. Figures. 1 and 2 show $\psi(\cdot)$ and $\psi^*(\cdot)$ of case (c), $n = 400$, respectively, where $(\beta, \gamma_1, \gamma_2) = (1, -0.5, -0.5)$. Figures 3 and 4 show $\psi(\cdot)$ and $\psi^*(\cdot)$ of case (d),

Table 1 Estimation of regression parameters under case (a) (b), $n = 200$

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (a) Z and W are independent												
0	− 0.007	0.193	0.185	0.95	− 0.002	0.258	0.249	0.94	− 0.007	0.311	0.303	0.94
0.6	− 0.059	0.220	0.209	0.93	− 0.023	0.218	0.210	0.94	− 0.014	0.223	0.210	0.94
0.6	− 0.011	0.328	0.315	0.95	0.012	0.342	0.319	0.94	0.016	0.339	0.319	0.93
0	− 0.018	0.269	0.235	0.92	− 0.022	0.314	0.311	0.95	0.009	0.402	0.378	0.94
− 0.1	− 0.069	0.203	0.202	0.93	− 0.010	0.205	0.201	0.95	0.002	0.204	0.202	0.94
− 0.1	0.005	0.294	0.287	0.95	0.004	0.296	0.286	0.94	− 0.010	0.287	0.286	0.95
0	− 0.022	0.306	0.294	0.94	− 0.040	0.377	0.385	0.94	− 0.045	0.492	0.466	0.95
− 0.5	− 0.069	0.220	0.209	0.93	− 0.012	0.218	0.208	0.93	− 0.017	0.217	0.208	0.94
− 0.5	0.027	0.321	0.311	0.93	− 0.006	0.305	0.309	0.96	0.035	0.323	0.308	0.94
1	0.007	0.185	0.198	0.96	0.002	0.247	0.259	0.96	0.007	0.323	0.311	0.93
0.6	− 0.071	0.216	0.208	0.93	− 0.029	0.218	0.210	0.93	− 0.012	0.220	0.211	0.95
0.6	0.044	0.303	0.317	0.96	0.011	0.348	0.318	0.93	0.027	0.325	0.320	0.95
1	0.007	0.253	0.249	0.95	0.013	0.312	0.323	0.96	− 0.028	0.386	0.389	0.95
− 0.1	− 0.080	0.203	0.203	0.94	− 0.005	0.206	0.202	0.95	− 0.010	0.211	0.202	0.95
− 0.1	0.064	0.283	0.286	0.95	0.013	0.269	0.286	0.95	0.011	0.293	0.286	0.94
1	0.005	0.331	0.309	0.94	− 0.016	0.368	0.398	0.96	− 0.070	0.472	0.477	0.94
− 0.5	− 0.054	0.210	0.209	0.94	− 0.015	0.212	0.208	0.96	− 0.007	0.203	0.208	0.95
− 0.5	0.039	0.300	0.310	0.95	0.025	0.306	0.308	0.95	0.026	0.303	0.308	0.96

Table 1 continued

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (b) Z and W are dependent												
0	− 0.010	0.211	0.186	0.93	− 0.012	0.271	0.249	0.94	− 0.016	0.325	0.303	0.94
0.6	− 0.061	0.217	0.209	0.93	− 0.024	0.231	0.210	0.93	− 0.015	0.227	0.210	0.94
0.6	0.017	0.336	0.316	0.96	0.022	0.350	0.319	0.93	0.023	0.359	0.320	0.93
0	0.002	0.271	0.235	0.93	− 0.045	0.346	0.310	0.93	− 0.009	0.434	0.380	0.93
− 0.1	− 0.050	0.215	0.203	0.93	0.014	0.221	0.202	0.95	− 0.017	0.209	0.202	0.94
− 0.1	0.009	0.307	0.287	0.93	− 0.021	0.306	0.286	0.93	0.000	0.297	0.286	0.95
0	− 0.007	0.331	0.292	0.93	− 0.050	0.418	0.385	0.93	− 0.029	0.442	0.467	0.96
− 0.5	− 0.033	0.213	0.208	0.95	− 0.006	0.221	0.208	0.95	− 0.005	0.213	0.208	0.95
− 0.5	− 0.009	0.305	0.311	0.95	0.005	0.304	0.309	0.96	0.003	0.307	0.309	0.94
1	0.022	0.197	0.198	0.94	0.018	0.243	0.259	0.96	0.014	0.308	0.311	0.96
0.6	− 0.060	0.215	0.209	0.93	− 0.007	0.222	0.210	0.94	− 0.017	0.217	0.211	0.94
0.6	0.038	0.339	0.317	0.94	− 0.006	0.333	0.319	0.93	0.026	0.342	0.319	0.94
1	0.019	0.245	0.249	0.95	0.009	0.322	0.323	0.95	− 0.019	0.384	0.388	0.96
− 0.1	− 0.061	0.204	0.202	0.96	− 0.008	0.212	0.202	0.95	− 0.022	0.204	0.202	0.95
− 0.1	0.036	0.284	0.286	0.95	0.025	0.299	0.286	0.94	0.030	0.292	0.286	0.96
1	0.019	0.308	0.311	0.96	− 0.031	0.398	0.397	0.95	− 0.028	0.474	0.479	0.96
− 0.5	− 0.069	0.203	0.209	0.95	− 0.011	0.215	0.208	0.95	− 0.008	0.209	0.208	0.96
− 0.5	0.053	0.290	0.310	0.96	0.025	0.312	0.308	0.94	0.024	0.322	0.308	0.94

Table 2 Estimation of regression parameters under case (a), (b), $n = 400$

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (a) Z and W are independent												
0	0.008	0.170	0.160	0.93	− 0.013	0.206	0.217	0.96	− 0.017	0.256	0.264	0.95
− 0.1	− 0.050	0.215	0.203	0.93	0.014	0.221	0.202	0.95	− 0.017	0.209	0.202	0.94
− 0.1	0.009	0.307	0.287	0.93	− 0.021	0.306	0.286	0.93	0.000	0.297	0.286	0.95
0	− 0.017	0.220	0.199	0.93	− 0.020	0.282	0.266	0.93	− 0.022	0.345	0.325	0.94
− 0.5	− 0.068	0.152	0.147	0.93	− 0.017	0.145	0.147	0.94	− 0.023	0.154	0.147	0.93
− 0.5	0.021	0.220	0.219	0.95	− 0.004	0.222	0.218	0.95	− 0.002	0.221	0.218	0.94
1	− 0.004	0.175	0.171	0.94	0.008	0.210	0.225	0.96	− 0.001	0.262	0.271	0.95
− 0.1	− 0.074	0.142	0.142	0.93	− 0.032	0.139	0.142	0.96	− 0.025	0.153	0.142	0.94
− 0.1	0.032	0.194	0.202	0.94	0.031	0.194	0.201	0.96	0.021	0.205	0.202	0.95
1	− 0.018	0.204	0.209	0.96	− 0.023	0.272	0.275	0.95	− 0.034	0.338	0.333	0.95
− 0.5	− 0.064	0.152	0.148	0.93	− 0.015	0.149	0.146	0.95	− 0.016	0.149	0.147	0.95
− 0.5	0.056	0.221	0.218	0.93	0.015	0.224	0.217	0.95	0.013	0.226	0.217	0.94

Table 2 continued

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (b) Z and W are dependent												
0	− 0.003	0.173	0.160	0.93	0.000	0.234	0.217	0.94	− 0.021	0.281	0.264	0.94
− 0.1	− 0.058	0.148	0.142	0.93	− 0.033	0.141	0.142	0.94	− 0.012	0.149	0.142	0.96
− 0.1	0.008	0.223	0.202	0.93	0.013	0.193	0.202	0.96	− 0.026	0.209	0.202	0.94
0	0.005	0.228	0.200	0.93	− 0.022	0.283	0.266	0.93	− 0.019	0.339	0.325	0.95
− 0.5	− 0.048	0.159	0.147	0.93	− 0.015	0.151	0.147	0.93	− 0.015	0.154	0.147	0.95
− 0.5	− 0.009	0.231	0.219	0.95	− 0.007	0.227	0.218	0.94	− 0.006	0.220	0.218	0.96
1	0.024	0.175	0.172	0.95	0.025	0.214	0.226	0.96	0.015	0.259	0.271	0.96
− 0.1	− 0.054	0.153	0.143	0.93	− 0.030	0.143	0.142	0.94	− 0.016	0.146	0.142	0.93
− 0.1	0.027	0.201	0.202	0.95	0.012	0.203	0.202	0.94	0.000	0.200	0.202	0.95
1	0.035	0.230	0.212	0.93	0.055	0.294	0.278	0.94	0.012	0.319	0.332	0.96
− 0.5	− 0.032	0.163	0.147	0.93	− 0.021	0.145	0.147	0.95	− 0.010	0.154	0.146	0.96
− 0.5	− 0.004	0.240	0.219	0.93	− 0.002	0.214	0.218	0.95	0.006	0.227	0.217	0.95

Table 3 Estimation of regression parameters under case (c), (d), $n = 400$

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (c) Z and W, and $W^*$are independent												
0	− 0.003	0.128	0.121	0.94	− 0.011	0.164	0.164	0.94	− 0.002	0.202	0.200	0.94
− 0.1	− 0.085	0.308	0.278	0.93	− 0.068	0.295	0.280	0.94	− 0.075	0.292	0.281	0.94
− 0.1	0.014	0.291	0.274	0.94	− 0.027	0.289	0.278	0.94	− 0.018	0.303	0.280	0.93
0	− 0.005	0.137	0.129	0.94	− 0.009	0.187	0.175	0.94	0.014	0.214	0.213	0.96
− 0.5	− 0.054	0.267	0.264	0.94	− 0.056	0.263	0.266	0.94	− 0.026	0.293	0.267	0.93
− 0.5	− 0.014	0.260	0.257	0.94	− 0.035	0.271	0.259	0.94	− 0.042	0.276	0.261	0.94
1	0.007	0.131	0.129	0.96	0.003	0.176	0.170	0.95	0.000	0.216	0.205	0.94
− 0.1	− 0.092	0.306	0.276	0.93	− 0.055	0.302	0.280	0.93	− 0.051	0.300	0.281	0.94
− 0.1	0.012	0.295	0.273	0.93	− 0.010	0.293	0.278	0.94	− 0.058	0.301	0.279	0.94
1	0.001	0.135	0.138	0.95	0.001	0.190	0.182	0.94	0.001	0.230	0.219	0.94
− 0.5	− 0.067	0.273	0.264	0.93	− 0.026	0.282	0.266	0.95	− 0.056	0.295	0.267	0.93
− 0.5	− 0.001	0.258	0.256	0.94	− 0.047	0.270	0.259	0.95	− 0.028	0.291	0.261	0.93

Table 3 continued

β and γ_1, γ_2	ν											
	0 Cox				0.5				1 PO			
	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP	Bias	SSD	ESE	CP
Case (d) Z and W, and $W^*$are dependent												
0	0.001	0.155	0.141	0.93	0.007	0.203	0.190	0.94	− 0.001	0.237	0.230	0.95
− 0.1	− 0.020	0.256	0.232	0.93	0.024	0.254	0.231	0.93	0.021	0.250	0.232	0.94
− 0.1	− 0.025	0.238	0.211	0.93	− 0.027	0.240	0.212	0.93	− 0.002	0.231	0.213	0.93
0	− 0.009	0.192	0.165	0.93	− 0.021	0.242	0.220	0.93	0.009	0.286	0.267	0.94
− 0.5	− 0.009	0.240	0.234	0.96	0.045	0.259	0.234	0.93	0.025	0.246	0.233	0.94
− 0.5	0.002	0.229	0.212	0.93	− 0.028	0.230	0.212	0.93	0.002	0.224	0.212	0.94
1	0.011	0.154	0.150	0.94	0.015	0.198	0.197	0.95	0.023	0.251	0.237	0.95
− 0.1	− 0.034	0.242	0.231	0.95	0.032	0.231	0.231	0.95	0.061	0.251	0.232	0.93
− 0.1	0.001	0.224	0.211	0.93	− 0.049	0.218	0.212	0.94	− 0.053	0.230	0.213	0.93
1	0.015	0.185	0.175	0.93	0.015	0.241	0.230	0.95	0.004	0.291	0.276	0.94
− 0.5	0.000	0.239	0.234	0.96	0.058	0.245	0.234	0.93	0.068	0.257	0.233	0.93
− 0.5	− 0.031	0.227	0.212	0.93	− 0.068	0.221	0.212	0.93	− 0.081	0.229	0.212	0.93

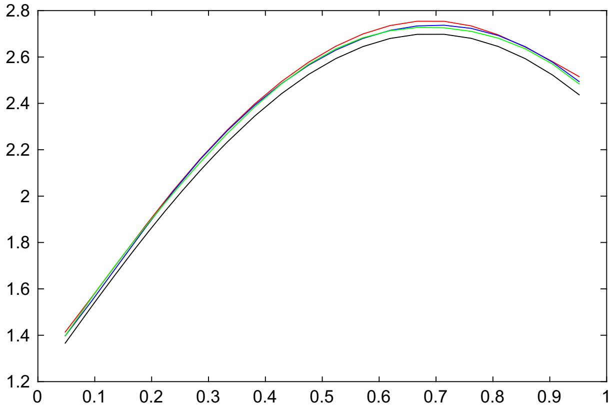


Fig. 1 The estimation of $\psi(\cdot)$ and true curve (black) with case (c). $\nu = 0$ (red); $\nu = 0.5$ (blue); $\nu = 1$ (green) (Color figure online)

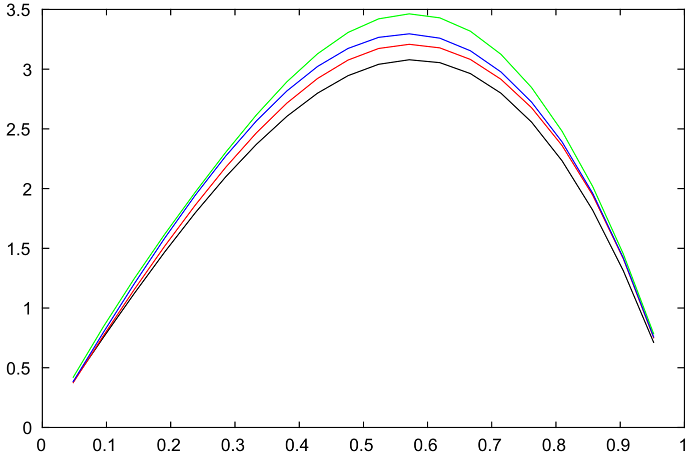


Fig. 2 The estimation of $\psi^*(\cdot)$ and true curve (black) with case (c). $\nu = 0$ (red); $\nu = 0.5$ (blue); $\nu = 1$ (green) (Color figure online)

$n = 400$, respectively, where $(\beta, \gamma_1, \gamma_2) = (1, -0.5, -0.5)$. It can be clearly seen from the tables and figures that the proposed estimator seems to be unbiased and the variance estimate seems reasonable. The normal approximation of the proposed estimator distribution works well.

6 An illustration

To illustrate the approach presented in the previous sections, we apply it to the kidney transplant data discussed by Klein and Moeschberger (2003) among others. This data consists of 863 kidney transplant patients with information on the age, gender and race of the subjects. Among them, 432 are white males, 92 are black males, 280 are

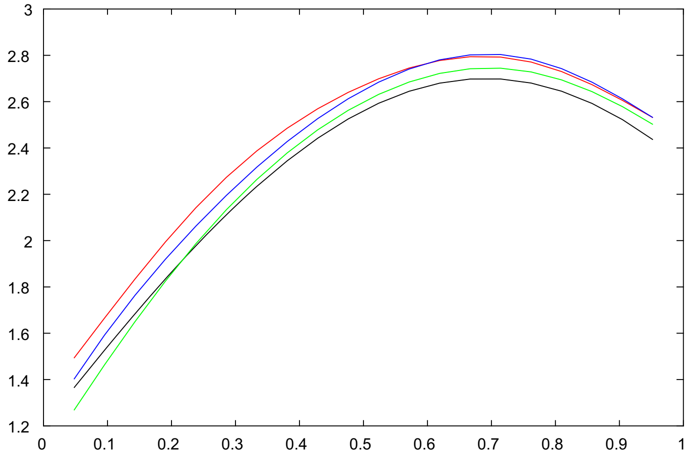


Fig. 3 The estimation of $\psi(\cdot)$ and true curve (black) with case (d). $\nu = 0$ (red); $\nu = 0.5$ (blue); $\nu = 1$ (green) (Color figure online)

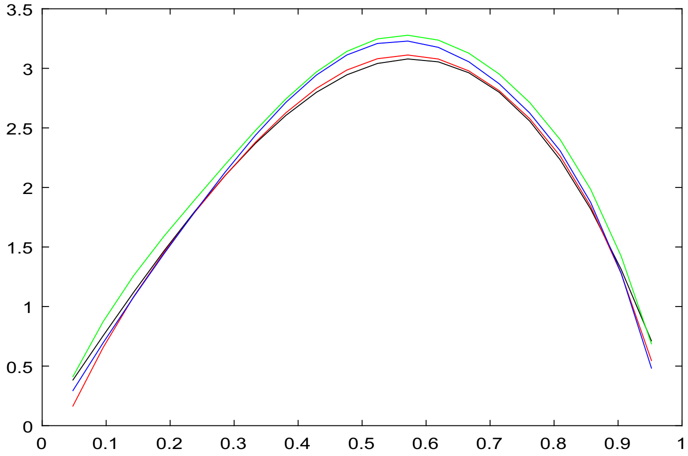
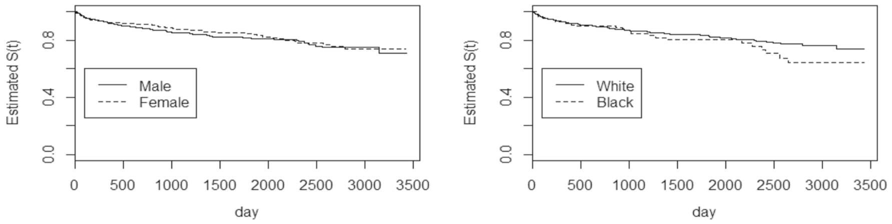


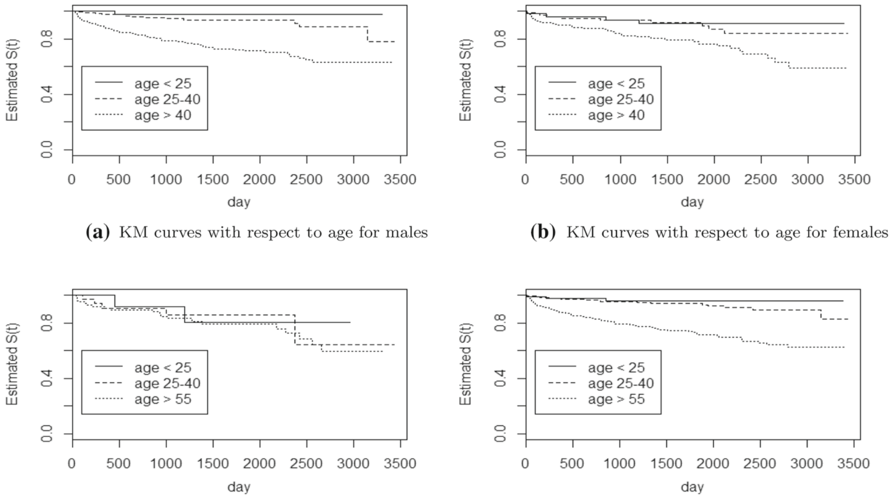
Fig. 4 The estimation of $\psi^*(\cdot)$ and true curve (black) with case (d). $\nu = 0$ (red); $\nu = 0.5$ (blue); $\nu = 1$ (green) (Color figure online)

white females, and 59 are black females, with an average age of 42.8 years and a range of 9.5 months to 74.5 years. If the patient is unavailable for follow-up on June 30, 1992 (at the end of the study) or if they are still alive, they are censored. A major feature of this data is a high right-censored rate of 84%. To illustrate this, Fig. 5a, b show Kaplan–Meier (KM) estimates of survival function from kidney transplantation to failure time, interest failure time, gender, and race, respectively. It can be seen that the two curves show a leveling away from zero at the tail, which means that there may be a nonsusceptible subgroup in the study. In the analysis below, we will focus on assessing the possible effects of age, gender and race from kidney transplantation to failure.



(a) KM curves with respect to gender **(b)** KM curves with respect to race

Fig. 5 **a** KM curves with respect to gender. **b** KM curves with respect to race



(a) KM curves with respect to age for males **(b)** KM curves with respect to age for females

(c) KM curves with respect to age for blacks **(d)** KM curves with respect to age for white

Fig. 6 **a** KM curves with respect to age for males. **b** KM curves with respect to age for females. **c** KM curves with respect to age for blacks. **d** KM curves with respect to age for white

For the analysis, we first adapt the data to the Cox model with respect to each covariate and found that both gender and race are not significant but age is. Furthermore, we look at the linear interactions between either gender to age or race to age, and none of them are significant. Figure 6a–d present the plots of the KM curves for male, female, black and white patients stratified by age (< 25, $25-40$, > 40), respectively. It can be seen that there is a cross survival curve, which means that the effects may interact non-linearly. Therefore, we should consider the varying coefficient effects. Let W be the age and Z^* be the gender or race. Based on the analysis above, we obtain the following transformation cure model with varying coefficients

$$\begin{aligned} logH(T^*) &= -\psi_1(w) - Z^*\psi_2(w) + \epsilon\,, \\ P(\eta = 1|W) &= e^{\psi_3(w)}/(1 + e^{\psi_3(w)})\,. \end{aligned}$$

Here $\psi_1(w)$ characterizes the main effect of age (W). Z^* represents gender or race, and $\psi_2(w)$ depicts the susceptible effect of female or black patients of different ages. Since the long-term survivors are closely related to age, we consider $\psi_3(w)$ as an unknown smooth curve for non-susceptibility. For the analysis we used the cubic spline ($l = 3$) and specified the baseline hazard function of ϵ as $e^t/(1 + e^t)$. For the selection of ν and the knot number K of the B-spline, we applied the BIC criterion that minimizes

$$BIC(\nu, K) = -2LLF + p(K + l) \times log(N)\,,$$

where LLF is the log likelihood function with all parameters replaced by their estimates associated with ν, K. N and p represent numbers as observations and smoothing functions, respectively. The optimal choice is given by $\nu = 0$ (Cox model) and $K = 9$, which is used throughout this section.

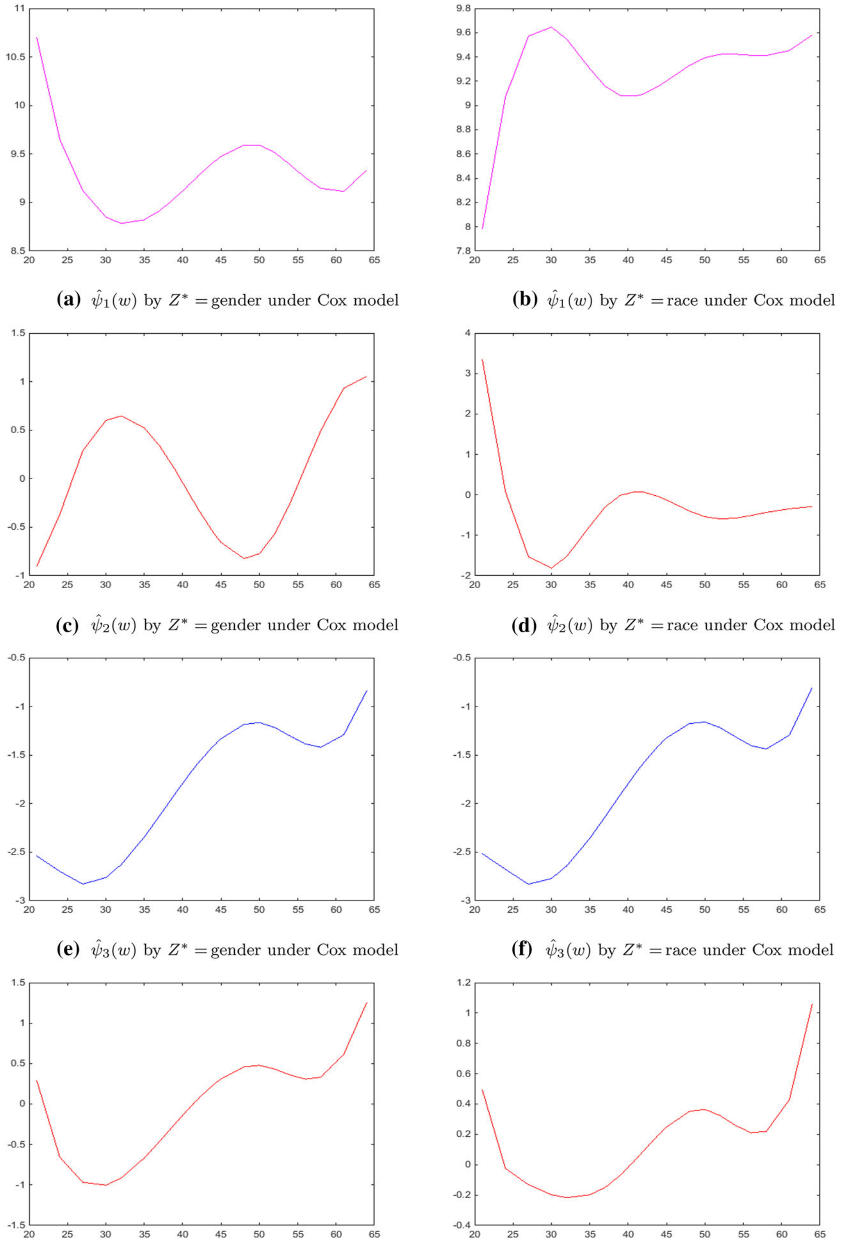
Figure 7a, b depict estimates of $\psi_1(w)$ with Z^* (gender or race), and it can be seen that the estimates are usually positive, and age generally seems to have some negative impact on the time to failure. As shown in Fig. 7c (female), the estimated $\psi_2(w)$ suggests that kidney transplantation may be beneficial for women under 25 years of age and women aged 40–55 years. Figure 7d (black) indicates that kidney transplantation is beneficial for black patients between the ages of 25 and 40. Furthermore, Fig. 7e, f also display an estimated $\psi_3(w)$, indicating that age generally has a negative impact on patients who are less susceptible to infection. In other words, younger patients have higher cure rates.

We consider the above-mentioned model ($logH(T^*) = -\psi_1(w) - Z^*\psi_2(w) + \epsilon$, $P(\eta = 1|W) = e^{\psi_3(w)}/(1 + e^{\psi_3(w)})$). Next, we consider the logistic model $P(\eta = 1|W) = e^{U^T\gamma}/(1 + e^{U^T\gamma})$ of a cure rate (U represents age), and the estimated value of γ is (−3.78, 0.47) or (- 3.70, 0.46) with Z^* female or black, respectively. In addition to the above model, we also consider the model $logH(T^*) = -Z\psi(w) + \epsilon$ of $\nu = 0$ and $K = 9$ to investigate the varying coefficient $Z\psi(w)$ without the primary influence of age (W), where Z represents gender or race. For this situation, the estimated γ is (− 14.26, 0.68) or (− 8.16, 0.41), and Z is female or black, respectively, and again the age seems to have some significant negative effects on the cure rate as above. Figure 7g, h give the estimated $\psi(w)$, where Z represents female or black, respectively. It is again proven that a kidney transplant benefits women between the ages of 22 and 40, and blacks between 25 and 40 years old.

7 Discussion and concluding remarks

This article presented a class of varying coefficient cure models that allows covariates to interact nonlinearly with susceptible and nonsusceptible subjects. For the purpose of inference, a sieve maximum likelihood estimation procedure was developed with the use of B-spline functions. The EM algorithm was also provided, and the asymptotic properties of the estimator were established. The main contribution of the proposed method is that it allows for the presence of a set of cures and possible time-varying or varying covariate effects. In addition, simulation studies have shown that the proposed method is applicable to the actual situation.

The proposed method can be seen as an extension of Chen and Tong (2010), who considered only the susceptible subgroup. It is well known that in the presence of



(a) $\hat{\psi}_1(w)$ by $Z^* =$gender under Cox model

(b) $\hat{\psi}_1(w)$ by $Z^* =$race under Cox model

(c) $\hat{\psi}_2(w)$ by $Z^* =$gender under Cox model

(d) $\hat{\psi}_2(w)$ by $Z^* =$race under Cox model

(e) $\hat{\psi}_3(w)$ by $Z^* =$gender under Cox model

(f) $\hat{\psi}_3(w)$ by $Z^* =$race under Cox model

(g) $\hat{\psi}(w)$ by $Z =$gender under Cox model

(h) $\hat{\psi}(w)$ by $Z =$race under Cox model

Fig. 7 **a** $\hat{\psi}_1(w)$ by $Z^* =$gender under Cox model. **b** $\hat{\psi}_1(w)$ by $Z^* =$race under Cox model. **c** $\hat{\psi}_2(w)$ by $Z^* =$gender under Cox model. **d** $\hat{\psi}_2(w)$ by $Z^* =$race under Cox model. **e** $\hat{\psi}_3(w)$ by $Z^* =$gender under Cox model. **f** $\hat{\psi}_3(w)$ by $Z^* =$race under Cox model. **g** $\hat{\psi}(w)$ by $Z =$gender under Cox model. **h** $\hat{\psi}(w)$ by $Z =$race under Cox model a cured subgroup, ignoring this subgroup may lead to biased estimates and lead to incorrect conclusions. On the other hand, one may take the existence of a cured subgroup into consideration. A common method of graphical inspection is to check if there is some leveling effect away from zero at the end of the survival curve. It would be helpful if some testing procedures can be developed to see if patients are cured or not.

Note that in the above, we have assumed that there exist some varying covariate effects on the failure time of interest, and correspondingly this may be true for the cure rate in practice. Actually it is straightforward to generalize the inference approach proposed above to this latter more general situation. More specifically, instead of model (3), one may consider the following varying covariate effect model

$$P(\eta = 1|U_1, U_2, W) = \frac{e^{U_1^T\gamma + U_2\psi^*(W)}}{1 + e^{U_1^T\gamma + U_2\psi^*(W)}},$$

where U_1 and U_2 are covariate vectors as U and $\psi^*(\cdot)$ is an unknown smooth function like $\psi(\cdot)$. Here also as with $\psi(\cdot)$, one may employ B-spline functions to approximate the unknown function $\psi^*(W)$ and develop an inference procedure similarly as above. Another issue that one might consider for future research is to develop methods to determine which covariates have linear effects and which covariates have non-parametric effects. In addition, it would be helpful to build some hypothesis testing procedures to check if the smoothing function is significant.

Acknowledgements The authors are grateful to Dr. Jianguo Sun for reading draft and giving detailed comments and suggestions. The authors also thank the Editor, the Associate Editor and referees for their comments that greatly improved the article. Funding was provided by Ministry of Science and Technology, Taiwan (Grant No. 102-2118-M-032-004), NSFC (Grant Nos. 11971064, 11671338).

Appendix: Proofs of the asymptotic properties of $\hat{\theta}_n$

In this appendix, we sketch the proofs for the two theorems described in Sect. 3. For this, we first define some notation and give the required regularity conditions. For given random variable X, measurable function f and probability measure P, define $Pf(X) = \int f dP$, $P_n f(X) = n^{-1}\sum_{i=1}^n f(X_i)$ and $G_n f(x) = n^{-1/2}\sum_{i=1}^n\{f(X_i) - Pf\}$, where the X_i's denote a random sample of X. Also for a function f, let $f^{(l)}$ denote its lth derivative and $\dot{f}(\cdot)$ and $\ddot{f}(\cdot)$ the first and second derivatives of f, respectively. The following regularity conditions are needed.

(C1). The true value $(\beta_0^T, \gamma_0^T)^T$ is an interior point of a known compact set $\mathcal{B}$ in R^{p+d}.

(C2). The function H_0 is strictly increasing with $H_0(0) = 0$, $\dot{H}_0(0) > 0$ and $H_0(\tau) < \mathcal{M}$.

(C3). The covariates X, W and Z are uniformly bounded.

(C4). The distribution of the error term ϵ has support R, and its hazard function $\lambda(t)$ has a continuous second derivative with $\lim_{t\downarrow 0}\lambda(\log t)/t > 0$.

(C5). Assume that $\psi_0 \in \Psi_r$ and $r > 0.5$, $0 < v < 0.5$.

(C6). The failure time T and the censoring time C are conditionally independent given the covariates (X, Z, W) with $\inf P(\eta T^* > \tau, C > \tau | X, Z, W) > 0$.

(C7). For any $t \in R$, we have $\log \lambda(t) - \Lambda(t) < 0$ and $\log \lambda(t) - \Lambda(t) \to -\infty$ as $t \to \infty$.

(C8). There exist unique $h^* \in \mathcal{T}(\mathcal{H})$ and $\psi^* \in \Psi_r$ such that for any $h \in \mathcal{T}(\mathcal{H})$ and $\psi_1 \in \Psi_r$, we have

$$E\left[\ddot{l}_{\xi H}(\theta_0)[h] - \ddot{l}_{\psi H}[\psi^*, h] - \ddot{l}_{HH}[h^*, h]\right] = 0$$

and

$$E\left[\ddot{l}_{\xi\psi}(\theta_0)[\psi] - \ddot{l}_{\psi\psi}[\psi^*, \psi_1] - \ddot{l}_{H\psi}[h^*, \psi]\right] = 0\,.$$

Here $\ddot{l}_{\xi H}(\theta_0)[h] = \{d\dot{l}_\xi(\xi_0, \psi_0, H_0 + t\int h dH_0)/dt\}|_{t=0}$ and other two quantities are defined similarly.

(C9). The information matrix

$$I(\theta_0) = E\left[\ddot{l}_{\xi\xi}(\theta_0) - \ddot{l}_{\xi\psi}(\theta_0)[\psi^*] - \ddot{l}_{\xi H}(\theta_0)[h^*]\right]$$

is nonsingular.

For a large M that may possibly depend on n, let $0 = a_0 < a_1 < \cdots < a_M = \tau$ be a sequence of points in $[0, \tau]$ such that $E\{\delta I(\tilde{T} \in R_j)\} = E(\delta)/M$, where $R_j = (a_{j-1}, a_j]$, $j = 1, \ldots, M$. For any $\theta = (\beta, \gamma, H, \psi) \in \Theta$, define the function

$$\begin{aligned} g(x, z, w, u, t, d; \theta) = d&\Bigg(\log \pi(u^T\gamma) - \Lambda[V(t;\theta)] + \log \lambda[V(t;\theta)]/H(t)\Bigg) \\ &+ d\textstyle\sum_{j=1}^M I(t \in R_j) \log[\{H(a_j) - H(a_{j-1})\}/\{E(\delta)/M\}] \\ &+ (1-d)\log\Bigg(1 - \pi(u^T\gamma) + \pi(u^T\gamma)\exp(-\Lambda[V(t;\theta)])\Bigg), \end{aligned}$$

where $V(t;\theta) = \log H(t) + x^T\beta + z\psi(w)$.

Also let $l^*(\theta) = g(X, Z, W, U, \tilde{T}, \delta; \theta)$ and for any $\theta \in \Theta$,

$$\begin{aligned} l_n^*(\theta) = &\textstyle\sum_{i=1}^n g(X_i, Z_i, W_i, U_i, \tilde{T}_i, \delta_i; \theta) \\ &+ \textstyle\sum_{i=1}^n \delta_i \sum_{j=1}^M I(\tilde{T}_i \in R_j)[\log\{E(\delta)/M\} - \log(m_{jn}/n)], \end{aligned}$$

where $m_{jn} = \sum_{i=1}^n \delta_i I(\tilde{T}_i \in R_j)$. Define

$$\begin{aligned} \hat{\theta}_n^* &= (\hat{\beta}_n^*, \hat{\gamma}_n, \hat{H}_n^*, \hat{\psi}_n^*) = \text{argmax}_{\theta \in \Theta_n} l_n^*(\theta) \quad \text{and} \\ \theta_0^* &= (\beta_0^*, \gamma_0^*, H_0^*, \psi_0^*) = \text{argmax}_{\theta \in \Theta} E\{l^*(\theta)\}. \end{aligned}$$

It follows from Corollary 4.10 of Schumaker (1981) that there exist vectors $\alpha_{0n} \in R^{q_n}$ and $\alpha_{0n}^* \in R^{q_n}$ such that

$$\sup_w |\psi_{0n}(w) - \psi_0(w)| = O(n^{-rv}) \quad \text{and} \quad \sup_w |\psi_{0n}^*(w) - \psi_0^*(w)| = O(n^{-rv}),$$

where $\psi_{0n}(w) = B_n(w)^T \alpha_{0n}$ and $\psi_{0n}^*(w) = B_n(w)^T \alpha_{0n}^*$. Denote $\theta_{0n} = (\beta_0, \gamma_0, H_0, \psi_{0n})$ and $\theta_{0n}^* = (\beta_0^*, \gamma_0^*, H_0^*, \psi_{0n}^*)$. Then

$$\rho(\theta_0, \theta_{0n}) = O(n^{-rv}) \quad \text{and} \quad \rho(\theta_0^*, \theta_{0n}^*) = O(n^{-rv}). \tag{A1}$$

The following lemmas will be used in the proof of Theorems 1 and 2.

Lemma 1 *Assume that conditions (C1)–(C6) hold, then we have that*

$$\rho(\hat{\theta}_n^*, \theta_{0n}^*) = O_p(n^{-(1-v)/2}) \quad \textit{and} \quad \rho(\hat{\theta}_n^*, \theta_0^*) = O_p(n^{-(1-v)/2} + n^{-rv}).$$

Proof Let η be a small and fixed positive number. Define a function class

$$\Gamma(\eta) = \{g(x, z, w, u, t, d; \theta) : \rho(\theta, \theta_{0n}^*) \le \eta\}.$$

Since H is nondecreasing bounded function, Theorem 2.7.5 of van der Vaart and Wellner (1996) yields that for any $0 < \xi < \eta$, the logarithm of the bracketing number $N_{[]}(\xi, \Gamma, \rho)$ for the function class $\Gamma(\eta)$ satisfy

$$\log N_{[]}(\xi, \Gamma, \rho) \le A\eta/\xi + Aq_n \log(\eta/\xi), \tag{A2}$$

by condition (C4), the monotonicity of functions $H(\cdot)$, $\Lambda\{\log(\cdot)\}$ and $\log(\cdot)$, where A is a constant not depending on n (van der Vaart and Wellner 1996). Let $J_{[]}(\eta, \Gamma, \rho) = \int_0^\eta \{1 + \log N_{[]}(\xi, \Gamma, \rho)\}^{1/2} d\xi$ be the integral entropy. It follows from (A2) that

$$J_{[]}(\eta, \Gamma, \rho) \le A_1 q_n^{1/2} \eta, \tag{A3}$$

where $A_1 > 0$ is a constant depending only on A. Then (A3) and Lemma 3.4.2 in van der Vaart and Wellner (1996, page 324) imply that for sufficiently large n,

$$\begin{aligned} &E\left(\sup_{\eta/2 \le \rho(\theta, \theta_{0n}^*) \le \eta,\ \theta \in \Theta} \left| \frac{1}{n}\{l_n^*(\theta) - l_n^*(\theta_{0n}^*)\} - [E\{l^*(\theta)\} - E\{l^*(\theta_{0n}^*)\}] \right| \right) \\ &\le n^{-1/2} J_{[]}(\eta, \Gamma, \rho)\{1 + A_2 J_{[]}(\eta, \Gamma, \rho)/(\eta^2 n^{1/2})\} \le A_3 \eta (q_n/n)^{1/2}, \end{aligned} \tag{A4}$$

where A_2 and A_3 are all constants not depending on n.

One can verify that the conditions of Lemma 3.4.1 in van der Vaart and Wellner (1996, page 322) are all satisfied. It then follows that $\rho(\hat{\theta}_n^*, \theta_{0n}^*) = O_p\{(n/q_n)^{-1/2}\} = O_p(n^{-(1-v)/2})$. Applying (A1), we have $\rho(\hat{\theta}_n^*, \theta_0^*) = O_p(n^{-(1-v)/2} + n^{-rv})$. The proof is complete. □

Lemma 2 *Assume that conditions (C1)–(C6) hold, then we have that*

$$\frac{1}{n}l_n^*(\hat{\theta}_n^*) - \mathcal{M}\max_{1\le j\le M}|\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})| \le \frac{1}{n}l_n(\hat{\theta}_n) + \frac{1}{n}\sum\nolimits_{i=1}^n \delta_i \log(n)$$
$$\le \frac{1}{n}l_n^*(\hat{\theta}_n) \ \le \ \frac{1}{n}l_n^*(\hat{\theta}_n^*), \tag{A5}$$

where $\mathcal{M} > 0$ is a constant which doesn't depend on M and n.

Proof Since $\hat{\theta}_n^*$ maximizes $l_n^*(\cdot)$, $l_n^*(\hat{\theta}_n) \le l_n^*(\hat{\theta}_n^*)$. For any $\theta \in \Theta$,

$$\frac{1}{n}l_n(\theta) + \frac{1}{n}\sum_{i=1}^n \delta_i \log(n) - \frac{1}{n}l_n^*(\theta)$$
$$= \frac{1}{n}\sum_{j=1}^M \Big[\log\Big\{\prod_{i:\tilde{T}_i \in R_j} \Delta H(\tilde{T}_i)\Big\} - \log\Big\{\frac{H(a_j) - H(a_{j-1})}{m_{jn}}\Big\}^{m_{jn}}\Big] \le 0.$$

Therefore the second inequality of (A5) holds. Define $\tilde{H}_n(t)$ as the step function with the same jump points as those of $\hat{H}_n^*$ such that, for any jump points $\tilde{T}_i \in R_j$,

$$\Delta\tilde{H}_n(\tilde{T}_i) = \{\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})\}/m_{jn}.$$

Let $\tilde{\theta}_n = (\hat{\beta}_n^*, \hat{\gamma}_n^*, \tilde{H}_n, \hat{\psi}_n^*)$. By the uniform boundedness of X, Z, W, there exists a fixed large $\mathcal{M}_0 > 0$ such that $\sup_{i,\theta\in\Theta_n} v_i(\theta) \le \mathcal{M}_0$ where $v_i(\theta) = \exp(V_i(\theta))$. By condition (C4) and (C5), a direct calculation using mean value theorem gives that

$$\begin{aligned}
&(1/n)\{l_n^*(\tilde{\theta}_n) - l_n^*(\hat{\theta}_n^*)\}\\
&\quad= \frac{1}{n}\sum_{i=1}^n \delta_i\Big(\log\Big[\frac{\lambda\{\log v_i(\tilde{\theta}_n^*)\}}{v_i(\tilde{\theta}_n^*)}\Big] - \log\Big[\frac{\lambda\{\log v_i(\hat{\theta}_n)\}}{v_i(\hat{\theta}_n)}\Big]\Big)\\
&\qquad - \frac{1}{n}\sum_{i=1}^n \delta_i\Big[-\Lambda\{\log v_i(\tilde{\theta}_n)\} - \Lambda\{\log v_i(\hat{\theta}_n)\}\Big]\\
&\qquad + \frac{1}{n}\sum_{i=1}^n (1-\delta_i)\Big[\log(1 - \pi(U_i^T\hat{\gamma}_n^*) + \pi(U_i^T\hat{\gamma}_n^*)\exp(-\Lambda\{\log v_i(\tilde{\theta}_n^*)\}))\\
&\qquad - \log(1 - \pi(U_i^T\hat{\gamma}_n^*) + \pi(U_i^T\hat{\gamma}_n^*)\exp(-\Lambda\{\log v_i(\hat{\theta}_n^*)\}))\Big]\\
&\quad\ge -A_4\frac{1}{n}\sum_{i=1}^n |\tilde{H}_n(\tilde{T}_i) - \hat{H}_n^*(\tilde{T}_i)|\exp\{X_i^T\hat{\beta}_n^* + Z_i\hat{\psi}_n^*(W_i)\}\\
&\quad\ge -A_4\sup_{1\le i\le n}\exp\{X_i^T\hat{\beta}_n^* + Z_i\hat{\psi}_n^*(W_i)\}\max_{1\le j\le M}|\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})|\\
&\quad\ge -A_4\mathcal{M}_0\max_{1\le j\le M}|\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})|,
\end{aligned}$$

where

$$A_4 = \sup_{0 \le v \le \mathcal{M}_0} \Big(|\frac{d}{dv} \log[\lambda\{\log(v)\}/v]| + |\lambda\{\log(v)\}/v| + \lambda\{\log(v)\} \Big)$$

is a constant not depending on M and n. As $\hat{\theta}_n$ is the maximizer of $l_n(\theta)$ and $l_n^*(\tilde{\theta}_n) = l_n(\tilde{\theta}_n) + \sum_{i=1}^n \delta_i \log(n)$, we have $l_n(\hat{\theta}_n) + \sum_{i=1}^n \delta_i \log(n) \ge l_n^*(\tilde{\theta}_n)$. The first inequality of (A5) follows and thus the proof is complete. □

Lemma 3 *Assume that conditions (C1)–(C6) hold and let* $M = M_n = O(n^{1-\gamma})$ *with* $\gamma < v$*, then we have that*

$$M \max_{j=1,\ldots,M} \{\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})\} = O_p(1). \tag{A6}$$

Proof Define a function of (x, z, w, u, t, d) for every given (θ, a) as

$$\begin{aligned}\phi(x, z, w, u, t, d; \theta, a) = {} & I(t \ge a) \exp\{x^T \beta + z\psi(w)\} \\ & \times \Bigg(dG[H(t) \exp\{x^T \beta + z\psi(w)\}] - d\frac{\dot{G}[v(t;\theta)]}{G[v(t;\theta)]} \\ & + (1-d)\frac{\pi(u^T\gamma)\exp(-\Lambda(\log[v(t;\theta)]))\lambda(\log[v(t;\theta)])}{1 - \pi(u^T\gamma) + \pi(u^T\gamma)\exp(-\Lambda(\log[v(t;\theta)]))} \Bigg)\end{aligned}$$

where $G(s) = \log\{\lambda(s)\}/s$ and $v(t;\theta) = H(t)\exp\{x^T\beta + z\psi(w)\}$. Similar to (A2)–(A4), one can show that the class of functions of (x, z, w, u, t, d),

$$\Big\{\phi(x, z, w, u, t, d; \theta, a) : \ \theta \in \Theta, \rho(\theta, \theta_{0n}^*) \le \eta, \ a \in [0, \tau]\Big\}$$

is a Glivenko-Cantelli class. It then follows from Lemma 1 that, as $n \to \infty$,

$$\sup_{a \in [0,\tau]} \Big| \frac{1}{n} \sum_{i=1}^n \phi(X_i, Z_i, W_i, U_i, \tilde{T}_i, \delta_i; \hat{\theta}_n^*, a) - \mu(\theta_0^*, a) \Big| \to 0 \tag{A7}$$

in probability, where

$$\mu(\theta_0^*, a) = E\{\psi(X, Z, W, U, \tilde{T}, \delta; \theta_0^*, a)\}$$

is continuous in $a \in [0, \tau]$.

Differentiate l_n^* with respect to $\Delta H(\tilde{T}_{(j)})$, where $\tilde{T}_{(j)} = \max\{\tilde{T}_i : \tilde{T}_i \in R_j, \delta_i = 1\}$, one obtains

$$\begin{aligned} \hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1}) &= (m_{jn}/n) / \left\{ (1/n) \sum_{i=1}^{n} \phi(X_i, Z_i, W_i, U_i, \tilde{T}_i, \delta_i; \hat{\theta}_n^*, \tilde{T}_{(j)}) \right\}^{-1} \\ &\leq (m_{jn}/n) / \left\{ (1/n) \sum_{i=1}^{n} \phi(X_i, Z_i, W_i, U_i, \tilde{T}_i, \delta_i; \hat{\theta}_n^*, a_j) \right\}^{-1} \end{aligned} \tag{A8}$$

The boundedness of $\hat{H}_n^* \in \mathcal{H}_n$ implies the left hand side is uniformly bounded. By taking fixed a_j, a_{j-1} one can see the denominator on the right hand side is uniformly bounded below away from 0. It then follows from (A7) and condition (C5) that $\inf_{a \in [0,\tau]} \mu(\theta_0^*, a) > 0$. Combining (A7) and (A8), we have

$$\sup_{1 \leq j \leq M} (n/m_{jn}) |\hat{H}_n^*(a_j) - \hat{H}_n^*(a_{j-1})| = O_p(1).$$

Since $E(m_{jn})/n = E(\delta)/M$, the uniform convergence of empirical distribution ensures that $\sup_{1 \leq j \leq M} |m_{jn}/n - E(\delta)/M| \to 0$. Then (A6) follows. The proof is complete. □

Given the lemmas above, we now present the proofs of Theorems 1 and 2.

Proof of Theorem 1.

Step 1. We first show that $\hat{\theta}_n$ exists and is finite. It is easy to check that

$$n^{-1} l_n(\hat{\theta}_n) < O_p(1) + P_n[\log \lambda(V(\hat{\theta}_n)) - \Lambda(V(\hat{\theta}_n))] \to -\infty$$

if there exists a $\hat{H}_n(T_i) = \infty$ such that $V_i(\hat{\theta}_n) = \infty$.
Step 2. We show that $\sup_n \hat{H}_n(\tau) < \infty$. Let $\xi_n = \hat{H}_n(\tau)$, $\bar{H}_n(t) = \hat{H}_n(t)/\xi_n$ and $\bar{V}_i(\theta) = \log \bar{H}_n(\tilde{T}_i) + X_i^T \beta + Z_i \psi_n(W_i)$. Since $\hat{\theta}_n$ maximizes $l_n(\theta)$, then

$$\begin{aligned} 0 &\leq n^{-1}[l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{\psi}_n, \hat{H}_n(\tau)) - l_n(\hat{\beta}_n, \hat{\gamma}_n, \hat{\psi}_n, \bar{H}_n(\tau))] \\ &= \frac{1}{n} \sum_{i=1}^{n} \delta_i [\log \lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n)) - \Lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n))] + O_p(1) \end{aligned}$$

yielding that

$$\frac{1}{n} \sum_{i=1}^{n} \delta_i [\log \lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n)) - \Lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n))] \geq O_p(1),$$

and thus by Condition (C7) that $\frac{1}{n} \sum_{i=1}^{n} \delta_i [\log \lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n)) - \Lambda(\log \xi_n + \bar{V}_i(\hat{\theta}_n))] I(\eta = 0, C_i \geq \tau) \geq O_p(1)$. If $\xi_n \to \infty$, then $\frac{1}{n} \sum_{i=1}^{n} \delta_i [\log \lambda(\log \xi_n +$

$\bar{V}_i(\hat{\theta}_n)) - \Lambda(\log\xi_n + \bar{V}_i(\hat{\theta}_n))]I(\tilde{T}_i \geq \tau) = \inf_i \log f(\log\xi_n + X_i\hat{\beta}_n + Z_i\hat{\psi}_n(W_i))E\delta_i I$ $(\tilde{T}_i \geq \tau)$, which tends to -∞ if $\xi_n \to \infty$.

Step 3. We prove that $\rho(\hat{\theta}_n, \theta_0) = O_p(n^{-(1-v)/2} + n^{-rv})$. Let $M = M_n = O(n^{1-\gamma})$ with $\gamma < v$. Lemma 2 gives

$$0 \leq (1/n)\{l_n^*(\hat{\theta}_n^*) - l_n^*(\hat{\theta}_n)\} \leq \mathcal{M} \max_{1\leq j\leq M_n} |\hat{H}_n^*(a_j) - \hat{H}_n(a_{j-1})|.$$

Hence Lemma 3 ensures

$$(1/n)|l_n^*(\hat{\theta}_n^*) - l_n^*(\hat{\theta}_n)| = O_p(n^{-(1-\gamma)}),$$

yielding

$$\rho(\hat{\theta}_n, \hat{\theta}_n^*) = O_p(n^{-(1-\gamma)/2}).$$

Analogously, one can show $\rho(\theta_0, \theta_0^*) = O_p(n^{-(1-\gamma)/2})$. Applying the triangle inequality and Lemma 1, we have, for $\gamma < v$,

$$\begin{aligned}\rho(\hat{\theta}_n, \theta_0) &\leq \rho(\hat{\theta}_n, \hat{\theta}_n^*) + \rho(\hat{\theta}_n^*, \theta_0^*) + \rho(\theta_0^*, \theta_0)\\ &= O_p(n^{-(1-v)/2} + n^{-rv} + n^{-(1-\gamma)/2})\\ &= O_p(n^{-(1-v)/2} + n^{-rv}).\end{aligned}$$

The proof is complete.

Proof of Theorem 2.

Denote ψ_n^* as the projection of ψ^* into the space spanned by the B-spline basis functions. According to Schumaker (1981), we have that

$$|\psi^* - \psi_n^*|_\infty = O(n^{-rv}). \tag{A9}$$

By Taylor expansion, we can have the following equations:

$$\begin{aligned}0 = P_n\dot{l}_\xi(\hat{\theta}_n) = {} & P_n\dot{l}_\xi(\theta_0) + P_n\ddot{l}_{\xi\xi}(\theta_0)(\hat{\xi}_n - \xi_0)\\ & + P_n\ddot{l}_{\xi\psi}(\theta_0)[\hat{\psi}_n - \psi_0] + P_n\ddot{l}_{\xi H}(\theta_0)[\hat{H}_n - H_0] + O_p(\rho^2(\hat{\theta}_n, \theta_0)),\end{aligned}$$

$$\begin{aligned}0 = P_n\dot{l}_\psi(\hat{\theta}_n)[\psi_n^*] = {} & P_n\dot{l}_\psi(\theta_0)[\psi_n^*] + P_n\ddot{l}_{\psi\xi}(\theta_0)[\psi_n^*](\hat{\xi}_n - \xi_0)\\ & + P_n\ddot{l}_{\psi\psi}(\theta_0)[\psi_n^*, \hat{\psi}_n - \psi_0] + P_n\ddot{l}_{\psi H}(\theta_0)[\psi_n^*, \hat{H}_n - H_0] + O_p(\rho^2(\hat{\theta}_n, \theta_0))\\ = {} & P_n\dot{l}_\psi(\theta_0)[\psi^*] + P_n\ddot{l}_{\psi\xi}(\theta_0)[\psi^*](\hat{\xi}_n - \xi_0) + P_n\ddot{l}_{\psi\psi}(\theta_0)[\psi^*, \hat{\psi}_n - \psi_0]\\ & + P_n\ddot{l}_{\psi H}(\theta_0)[\psi^*, \hat{H}_n - H_0] + O_p(\rho^2(\hat{\theta}_n, \theta_0) + |\psi_n^* - \psi^*|_\infty\rho(\hat{\theta}_n, \theta_0)),\end{aligned}$$

and

$$\begin{aligned}0 = P_n\dot{l}_H(\hat{\theta}_n)[h^*] = {} & P_n\dot{l}_H(\theta_0)[h^*] + P_n\ddot{l}_{H\xi}(\theta_0)[h^*](\hat{\xi}_n - \xi_0)\\ & + P_n\ddot{l}_{H\psi}(\theta_0)[h^*, \hat{\psi}_n - \psi_0] + P_n\ddot{l}_{HH}(\theta_0)[h^*, \hat{H}_n - H_0] + O_p(\rho^2(\hat{\theta}_n, \theta_0)).\end{aligned}$$

It follows form Theorem 1 with the conditions $0.25/r < v < 0.5$ and (A9) that $|\psi_n^* - \psi^*|_\infty \rho(\hat{\theta}_n, \theta_0) = \rho^2(\hat{\theta}_n, \theta_0) = o(n^{-1/2})$. Combing all above equalities with conditions (C8) and (C9), one obtains that

$$\begin{aligned}\hat{\xi}_n - \xi_0 = &\left[P_n(\ddot{l}_{\xi\xi}(\theta_0) - \ddot{l}_{\xi\psi}(\theta_0)[\psi^*] - \ddot{l}_{\xi H}(\theta_0)[h^*]) \right]^{-1} \\ &\times P_n\left[\dot{l}_{\xi}(\theta_0) - \dot{l}_{\psi}(\theta_0)[\psi^*] - \dot{l}_{H}(\theta_0)[h^*] \right] + o_p(n^{-1/2}),\end{aligned}$$

which reduces to the desired results by strong large number law and central limit theorem.

References

Cai Z, Fan J, Li R (2000) Efficient estimation and inferences for varying-coefficient models. J Am Stat Assoc 9:888–902

Cai J, Fan J, Jiang JC, Zhou HB (2007) Partially linear hazard regression for multivariate survival data. J Am Stat Assoc 102:538–551

Chen K, Tong X (2010) Varying coefficient transformation models with censored data. Biometrika 97:969–976

Chen C-M, Lu T-FC, Hsu C-M (2013) Association estimation for clustered failure time data with a cure fraction. Comput Stat Data Anal 57:210–222

Choi S, Huang X, Chen Y-H (2014) A class of semiparametric transformation models for survival data with a cured proportion. Lifetime Data Anal 20:369–386

Demarqui FN, Dey DK, Loschi RH, Colosimo EA (2014) Fully semiparametric Bayesian approach for modeling survival data with cure fraction. Biom J 56:198–218

Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. Ann Stat 27:1491–1518

Fan J, Lin H, Zhou Y (2006) Local partial-likelihood estimation for life time data. Ann Stat 34:290–325

Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 43:181–192

Hastie T, Tibshirani R (1990) Exploring the nature of covariate effects in the proportional hazards model. Biometrics 46:1005–1016

Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York

Klein JP, Moeschberger ML (2003) Techniques for censored and truncated data, 2nd edn. Springer, New York

Kuk AYC, Chen C-H (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79:531–541

Lu W (2010) Efficient estimation for an accelerated failure time model with a cure fraction. Stat Sin 20:661–674

Lu W, Ying Z (2004) On semiparametric transformation cure models. Biometrika 91:331–343

Schumaker LL (1981) Spline functions: basic theory. Wiley, New York

Stone CJ (1980) Optimal rates of convergence for nonparametric estimators. Ann Stat 8:1348–1360

Stone CJ (1982) Optimal global rates of convergence for nonparametric regression. Ann Stat 10:1040–1053

Tsodikov AD (1998) A proportional hazards model taking account of long-term survivors. Biometrics 54:1508–1516

van der Vaart A, Wellner JA (1996) Weak convergence and empirical processes: with applications to statistics. Springer, New York

Wang L, Du P, Liang H (2012) Two-component mixture cure rate model with spline estimated nonparametric components. Biometrics 68:726–735

Yakovlev AY, Tsodikov AD (1996) Stochastic models of tumor latency and their biostatistical applications. World Scientific, Singapore

Zeng D, Lin DY (2006) Efficient estimation of semiparametric transformation models for counting processes. Biometrika 93:627–640

Zeng D, Lin DY, Yin G (2005) Maximum likelihood estimation for the proportional odds model with random effects. J Am Stat Assoc 100:470–483

Zeng D, Ying G, Ibrahim JG (2006) Semiparametric transformation models for survival data with a cure fraction. J Am Stat Assoc 101:670–684

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.