Ann Inst Stat Math (2018) 70:523–551
https://doi.org/10.1007/s10463-017-0600-6

Versatile estimation in censored single-index hazards regression

Chin-Tsang Chiang[1] · **Shao-Hsuan Wang**[1] · **Ming-Yueh Huang**[1]

Received: 11 April 2016 / Revised: 26 October 2016 / Published online: 17 February 2017

Abstract One attractive advantage of the presented single-index hazards regression is that it can take into account possibly time-dependent covariates. In such a model formulation, the main theme of this research is to develop a theoretically valid and practically feasible estimation procedure for the index coefficients and the induced survival function. In particular, compared with the existing pseudo-likelihood approaches, our one proposes an automatic bandwidth selection and suppresses an influence of outliers. By making an effective use of the considered versatile survival process, we further reduce a substantial finite-sample bias in the Chambless-Diao type estimator of the most popular time-dependent accuracy summary. The asymptotic properties of estimators and data-driven bandwidths are also established under some suitable conditions. It is found in simulations that the proposed estimators and inference procedures exhibit quite satisfactory performances. Moreover, the general applicability of our methodology is illustrated by two empirical data.

Keywords Accuracy measure · Conditional survival function · Cross-validation · Kaplan–Meier estimator · Pseudo-integrated least squares estimator · Pseudo-maximum likelihood estimator · Single-index hazards model · U-statistic

Electronic supplementary material The online version of this article (doi:10.1007/s10463-017-0600-6) contains supplementary material, which is available to authorized users.

✉ Chin-Tsang Chiang
chiangct@ntu.edu.tw

1 Institute of Applied Mathematical Sciences, National Taiwan University, Taipei 10617, Taiwan, ROC

1 Introduction

In survival analysis, a significant amount of research has been devoted to characterizing the relationship between a continuous failure time T and some interesting covariates $Z_t = (Z_{1t}, \ldots, Z_{dt})^\top$, which might be time independent or predictable time dependent. Basically, hazard rates can capture all information in the survival process and lay a proper foundation to formulate the influence of possibly time-dependent covariates. In view of these facts, we propose an appealing single-index hazards (SIH) model:

$$\lambda_T(t|z_{H_t}) = \lambda(t, z_{t\beta_0}), \quad t \in [0, \tau], \tag{1}$$

where $\lambda_T(t|z_{H_t})$ stands for the hazard function of T given a realization z_{H_t} of the entire covariate history $Z_{H_t} = \{Z_u : u \in [0, t]\}$ up to time t, $\lambda(t, \upsilon)$ is an unknown bivariate function, $z_{t\beta} = z_{1t} + z_{2t}\beta_2 + \cdots + z_{dt}\beta_d$, $\beta_0 = (\beta_{02}, \ldots, \beta_{0d})^T$ is the true coefficient vector, and τ is the time of last follow-up or the end of the study period. One attractive feature of the SIH model is to resolve the limitation of time independent covariates in some semiparametric survival models. For the sake of identifiability and simplicity, the coefficient of a significant covariate, say Z_{1t}, is assumed to be one. As a result, the coefficient β_{0k} is interpreted as the relative effect of Z_{kt}, compared to Z_{1t}, on the hazard function, $k = 2, \ldots, d$.

In the past decades, the Cox's proportional hazards (PH) model (Cox 1972) and the accelerated failure time (AFT) model have been widely used and extensively studied in the analysis of survival data. To reduce the risk of misspecification in parametric models and the difficulty of explanation in a fully nonparametric one, Khan and Tamer (2007) considered a generalized AFT model for time-independent covariates and developed the partial rank estimation for the index coefficients. In fact, all of these models are particular forms of the SIH model with a monotonic dependence on the linear predictor. In the context of risk-specific SIH models for several types of failures, Gørgens (2004, 2006) modified the weighted average derivative estimation (WADE) to propose a new estimation for the setup of censored survival data. However, the derived estimator will suffer from the so called "curse of dimensionality," especially when the number of covariate increases. Furthermore, an intensive calculation task for multiple integrals is usually required and the bandwidth selection for the estimator is still unsettled. Recently, Bouaziz and Lopez (2010) introduced a pseudo-likelihood approach to avoid the computational complexity in the generalized WADE. By taking into account a more general and realistic censoring mechanism, Strzalkowska-Kominiak and Cao (2013, 2014) proposed other pseudo-maximum likelihood estimators (PMLEs). As evidenced by our numerical studies, both of the PMLEs might introduce rather large finite-sample variances, which are partially caused by some degree of subjective trimming function and bandwidths. To overcome the limitation of classical AFT models, which are only applicable for time-dependent predictors, Zeng and Lin (2007) considered an extended model of Cox and Oakes (1984) and developed a more efficient kernel-smoothed profile likelihood estimation. As one can see, the induced hazard function is different from our model formulation in (1).

By incorporating the counting process representation of a failure time into suitable estimating equations, an alternative approach is developed to estimate the index coefficients β_0. Same with some novel semiparametric approaches, the proposed estimator is also $\sqrt{n}$-consistent, asymptotically normally distributed, and independent of the dimension of covariates. Further, the involved bandwidth, which is treated as a tuning parameter, is estimated by the solution of a pseudo-cross-validation estimating equation. It is shown that the chosen bandwidth is valid only when a fourth-order or higher-order kernel function is adopted. In practice, the induced survival function of the SIH model can be directly estimated by the Kaplan–Meier type estimator in the constructed estimating equations. Different from the bandwidth selector for the estimator of β_0, another data-driven criterion is introduced to estimate the optimal bandwidth of the survival function estimator in terms of a certain asymptotic integrated mean squared error. Basically, our estimation and inference procedures rely on a very general censoring assumption and can be easily extended to survival data with competing risks. Without the limitation of continuous covariates in the WADE, the developed approach is applicable to survival data with some discrete and/or continuous covariates provided that the continuity condition of Ichimura (1993) is satisfied. Based on the defined pseudo-residual process, the test rules of Chiang and Huang (2012) can be further carried over to check the model correctness.

In light of the equality between the conditional distribution of T and the conditional expectation of the underlying counting process of T on the baseline marker values Z, modeling for the distribution function (or survival function) can be regarded as assessing the discriminability of Z on cumulative cases versus dynamic controls, which are defined by Heagerty et al. (2000). Under this classification framework and a strict increase in $\lambda(t, \upsilon)$ in υ, the single-index Z_{β_0} can be shown to have the highest receiver operating characteristic (ROC) curve among all composite markers (cf. McIntosh and Pepe 2002) at each time point t. In the perspective of optimality for classification and prediction, the area under the time-dependent ROC curve (AUC) of Z_{β_0}, thus, provides a concise and informative summary. Different from the classical rank-based measures such as the concordance index and the rank correlation, it displays the time-dependent changes in discrimination. By making an effective use of our versatile survival process, a new estimation is proposed to estimate such an accuracy measure. Basically, this approach can substantially reduce a finite-sample bias in the Chambless-Diao estimator. To achieve the $\sqrt{n}$-consistence of the introduced estimators, we also establish a cross-validation criterion for bandwidth selection.

The rest of this article is organized as follows. In Sect. 2, we outline a new pseudo-estimation for the index coefficients and the conditional survival function. Some feasible criteria for bandwidth selection and the asymptotic properties of the proposed estimators are also established in this section. Section 3 focuses on studying estimation and inference for the time-dependent AUC. In Sect. 4, we investigate the finite-sample performances of the proposed estimators and make comparisons with the existing competitors through simulations. The Mayo primary biliary cirrhosis (PBC) data and the British Columbia Vital Statistics (BCVS) data are further used in Sect. 5 to illustrate the usefulness of our methodology. Moreover, the last section provides the concluding remarks and presents future directions of this research. As for the technical lemmas and the proofs of the main theorems, they are relegated to Supplementary Materials (I) and (II).

2 Estimation and inference for the SIH model

Let $X = \min\{T, C\}$ and $\delta = I(X = T)$ with C being the censoring time. For ease of presentation, $\widetilde{N}(t)$, $N(t)$, $V(t)$, and $Y(t)$ are used to stand for the underlying counting process $I(T \le t)$, the observed counting process $I(X \le t, \delta = 1)$, the available vital status $(1 - I(X \le t, \delta = 0))$, and the at-risk process $I(X \ge t)$, respectively. Further, we define $K^{(m)}(u)$ as the mth derivative of a generic function $K(u)$, $\widehat{E}^{-k}$ as a generic estimator $\widehat{E}$ with the kth unit being deleted, and $S(t, z_{H_t}, \beta) = \prod_{\{0<u\le t\}}(1 - E[dN(u)|Z_{u\beta} = z_{u\beta}]/E[Y(u)|Z_{u\beta} = z_{u\beta}])$, where $\prod$ represents the infinite product. With censored survival data of the form $\{(X_i, \delta_i, Z_{iX_i}) : 1 \le i \le n\}$, the estimation and inference are proposed under the assumption of conditionally independent censoring:

A1. $E[\mathrm{d}\widetilde{N}(t)|Z_{H_t}, \widetilde{N}(t^-), I(C \le t)] = E[\mathrm{d}\widetilde{N}(t)|Z_{H_t}, \widetilde{N}(t^-)] \quad \forall t.$

By assumption A1 and the definition of $S(t, z_{H_t}, \beta)$, it is straightforward to have $S(t, z_{H_t}, \beta_0) = \prod_{\{0<u\le t\}}(1 - \lambda(u, z_{u\beta_0})\mathrm{d}u)$ $(= \exp(-\int_0^t \lambda(u, z_{u\beta_0})\mathrm{d}u))$.

2.1 Background

In the spirit of Buckley and James (1979), one has

$$\begin{aligned} E[\widetilde{N}(t)|X, \delta, Z_{H_t}] &= E[(V(t) + (1 - V(t))\widetilde{N}(t)|X, \delta, Z_{H_t}] \\ &= V(t)\widetilde{N}(t) + (1 - V(t))\frac{S(X, Z_{H_t}, \beta_0) - S(t, Z_{H_t}, \beta_0)}{S(X, Z_{H_t}, \beta_0)}. \end{aligned} \tag{2}$$

By substituting $S(t, Z_{H_t}, \beta)$ for $S(t, Z_{H_t}, \beta_0)$, the following counting process is defined:

$$N(t; S, \beta) = V(t)\widetilde{N}(t) + (1 - V(t))\frac{S(X, Z_{H_t}, \beta) - S(t, Z_{H_t}, \beta)}{S(X, Z_{H_t}, \beta)}. \tag{3}$$

It follows from (2)–(3) that $E[N(t; S, \beta_0)|Z_{H_t}] = E[\widetilde{N}(t)|Z_{H_t}] = 1 - S(t, Z_{H_t}, \beta_0)$. The essential idea behind our approach is to fully use the distinctive features of this estimable counting process $N(t; S, \beta_0)$. By an alternative expression $(\delta + (1-\delta)I(X > t))$ of $V(t)$, we can also derive that $N(t; S, \beta_0) = \delta\widetilde{N}(t) + (1 - \delta)E[\widetilde{N}(t)|X, \delta = 0, Z_{H_t}]$, whereas this result is not easily explained in terms of imputation for missing status of $\widetilde{N}(t)$.

Given $\beta = (\beta_2, \ldots, \beta_d)^\top$ in the compact parameter space $\mathcal{B} \subseteq \mathbb{R}^{d-1}$, $S(t, z_{H_t}, \beta)$ can be reasonably estimated by the Kaplan–Meier type estimator

$$\widehat{S}_h(t, z_{H_t}, \beta) = \prod_{0<u\le t}(1 - d_u\widehat{\Lambda}_h(u, z_{H_u}, \beta)), \tag{4}$$

where $\widehat{\Lambda}_h(t, z_{H_t}, \beta) = \int_0^t \frac{d_u \widehat{H}_{1,h}(u, z_u, \beta)}{\widehat{H}_{0,h}(u, z_u, \beta)} \triangleq \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n K_{q,h}(Z_{iu\beta} - z_{u\beta}) dN_i(u)}{\frac{1}{n}\sum_{i=1}^n K_{q,h}(Z_{iu\beta} - z_{u\beta}) Y_i(u)}$, $q = 2, 4, \ldots$, and $d_t \widehat{\Lambda}_h(t, z_{H_t}, \beta)$ is naturally interpreted as 0 for $\widehat{H}_{0,h}(t, z_{H_t}, \beta) \leq 0$. In (4), a qth-order kernel function $K_{q,h}(u) = K_q(u/h)/h$ is adopted with h being a positive-valued bandwidth and $K_q(\upsilon)$ being symmetric about zero with bounded variation and satisfying

$$\int K_q(\upsilon)\mathrm{d}\upsilon = 1, \int \upsilon^k K_q(\upsilon)\mathrm{d}\upsilon = 0, k = 1, \ldots, (q-1),$$
$$\int |\upsilon|^q K_q(\upsilon)\mathrm{d}\upsilon < \infty, \text{ and } \int K_q^{(2)}(\upsilon)\mathrm{d}\upsilon = 0.$$

Basically, the constraint on $K_q^{(2)}(\upsilon)$ is made to assure that the information matrix of the proposed pseudo-estimating equations converges in probability to a nonsingular constant matrix at each time point t. An example of such a kernel function is given by $K_q(\upsilon) = \alpha_q(\upsilon) \cdot \{15(1-\upsilon^2)^2/16\} I(|\upsilon| \leq 1)$, where $\alpha_q(\upsilon)$ is the $(q/2-1)$-th-order polynomial in υ^2, $q = 2, 4, \ldots$.

Let $\mathcal{Z}_t \subseteq \mathbb{R}^d$ be the union of a finite number of open convex sets such that $\widehat{H}_{0,h}(t, z_t, \beta)$ is far away from zero for $z_t \in \mathcal{Z}_t$ and $t \in [0, \tau]$. Under some suitable conditions, it shall be shown that $\widehat{S}_h(t, z_{H_t}, \beta)$ converges uniformly to a parameter function $S^{[0]}(t, z_{H_t}, \beta)$. Moreover, $S^{[0]}(t, z_{H_t}, \beta_0) = S(t, z_{H_t}, \beta_0)$ and

$$E\left[\left\{N(t; S, \beta_0) - (1 - S^{[0]}(t, Z_{H_t}, \beta))\right\}^2\right] \\ \geq E\left[\{N(t; S, \beta_0) - (1 - S(t, Z_{H_t}, \beta_0))\}^2\right] \quad \forall t \in [0, \tau] \tag{5}$$

with the equality holding if and only if $\beta = \beta_0$. Coupled with the invariance of β_0 with respect to time, an available information of $N(t; S, \beta_0)$ can be appropriately integrated over $[0, \tau]$. In application, τ is usually chosen as the maximum observed time $X_{(n)} = \max_{1 \leq i \leq n}\{X_i\}$.

2.2 Estimators of β_0 and $S(t, z_{H_t}, \beta_0)$

By mimicking the score function of the pseudo-sum of integrated squares

$$\mathrm{PSS}(\beta) = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{N_i(t; S, \beta_0) - (1 - \widehat{S}_h(t, Z_{iH_t}, \beta))\right\}^2 \mathrm{d}W_{in}(t), \tag{6}$$

the pseudo-integrated least squares estimator (PILSE) $\widehat{\beta}$ of β_0 is presented as a solution of the estimating equations:

$$\mathrm{PS}_h(\beta) \triangleq \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{N_i(t; \widehat{S}_h, \beta)\right.$$

$$-(1-\widehat{S}_h(t, Z_{iH_t}, \beta))\big\} \partial_\beta \widehat{S}_h(t, Z_{iH_t}, \beta)\mathrm{d}W_{in}(t)=0, \quad (7)$$

where $W_{in}(t)$ is a nonnegative and nondecreasing weight function, which uniformly converges to a Z_{iH_t}-measurable function $W_i(t)$ and satisfies assumption A5 in Sect. 2.3, $i = 1, \ldots, n$. Regardless of time-dependent or time-independent covariates, β_0 can be estimated by the estimating equations $\sum_{i=1}^{n}\{N_i(t; \widehat{S}_h, \beta) - (1-\widehat{S}_h(t, Z_{iH_t}, \beta))\}\partial_\beta \widehat{S}_h(t, Z_{iH_t}, \beta)/n = 0$ for each t. The above estimating equations further integrate available information over $[0, \tau]$. Different from the pseudo-maximum likelihood estimation, the proposed estimation avoids the effect of outliers because the possible values of $N_i(t; \widehat{S}_h, \beta)$'s and $\widehat{S}_h(t, Z_{iH_t}, \beta)$'s are bounded within the interval [0, 1]. Moreover, the difficulty in estimating the right-tail of distribution can be suppressed by an appropriate choice of $\mathrm{d}W_n(t)$. To deal with a subject with $Z_t \notin \mathcal{Z}_t$ for some t, the corresponding trimming function can also be absorbed into $\mathrm{d}W_n(t)$. For $q > 2$ and a suitable constraint on h (assumption A4 in Sect. 2.3), the PILSE $\widehat{\beta}$ can be shown to be $\sqrt{n}$-consistent with the asymptotic distribution being independent of q. As for complete failure time data with time-independent covariates, the estimating equations in (7) will be simplified to those proposed by Chiang and Huang (2012).

By extending the bandwidth selection procedure of Härdle et al. (1993), the bandwidth estimator $\widehat{h}$ is chosen for h with $(\widehat{\beta}, \widehat{h})$ simultaneously satisfying $\mathrm{PS}_h(\beta) = 0$ and

$$\mathrm{PS}_\beta(h) \triangleq \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \Big\{N_i\left(t; \widehat{S}_h^{-i}, \beta\right) - \left(1-\widehat{S}_h^{-i}\left(t, Z_{iH_t}, \beta\right)\right)\Big\} \partial_h \widehat{S}_h^{-i}(t, Z_{iH_t}\beta)\mathrm{d}W_{in}(t) = 0. \quad (8)$$

In S1 of Supplementary Material (II), $\widehat{h}$ is further demonstrated to be $O_{\mathrm{p}}(n^{-1/(1+2q)})$. Unfortunately, $\widehat{\beta}$ cannot achieve the $\sqrt{n}$-consistency when a second-order kernel function is adopted in (7), (8). To avoid the numerical instability caused by the use of higher-order kernel functions, a fourth-order kernel function $K_4(\upsilon)$ is suggested in practical implementation.

For the computation of $(\widehat{\beta}, \widehat{h})$, an iterative scoring procedure is implemented by the following steps:

Step 1. Set $\widehat{\beta}^{(1)}$ and $\widehat{h}^{(1)} \propto n^{-1/(2q+1)}$ with $\partial_\beta \mathrm{PS}_{\widehat{h}^{(1)}}(\widehat{\beta}^{(1)})$ being nonsingular and $\partial_h \mathrm{PS}_{\widehat{\beta}^{(1)}}(\widehat{h}^{(1)}) \neq 0$.

Step 2. Compute

$$\begin{pmatrix}\widehat{\beta}^{(k+1)}\\ \widehat{h}^{(k+1)}\end{pmatrix} = \begin{pmatrix}\widehat{\beta}^{(k)}\\ \widehat{h}^{(k)}\end{pmatrix} - \begin{pmatrix}(\partial_\beta \mathrm{PS}_{\widehat{h}^{*(k)}}(\widehat{\beta}^{(k)}))^{-1}\mathrm{PS}_{\widehat{h}^{(k)}}(\widehat{\beta}^{(k)})\\ (\partial_h \mathrm{PS}_{\widehat{\beta}^{(k)}}(\widehat{h}^{(k)}) + E^{(k)})^{-1}\mathrm{PS}_{\widehat{\beta}^{(k)}}(\widehat{h}^{(k)})\end{pmatrix}, k = 1, \ldots, \quad (9)$$

where $h^{*(k)} = h^{(k)}n^{\varepsilon_{1k}}$ with ε_{1k} being close to zero in the interval $1/(2q+1) - (1/5, 1/4q)$ such that $\partial_\beta PS_{\widehat{h}^{*(k)}}(\widehat{\beta}^{(k)})$ is nonsingular, and $E^{(k)} = 0$

if $\partial_h PS_{\widehat{\beta}^{(k)}}(\widehat{h}^{(k)}) \neq 0$ and $E^{(k)} = n^{-(2q-2)/(2q+1)+\varepsilon_{2k}}$ for some $\varepsilon_{2k} > 0$ otherwise.

Step 3. Repeat Step 2 until $\|\widehat{\beta}^{(k+1)} - \widehat{\beta}^{(k)}\|/\|\widehat{\beta}^{(k)}\| < \varepsilon_{\beta_0}$ and $|h^{(k+1)} - h^{(k)}|/|h^{(k)}| < \varepsilon_{h_0}$ for some pre-chosen small values ε_{β_0} and ε_{h_0}, where $\|\cdot\|$ stands for the Euclidean norm of a vector.

As in the context of generalized estimating equations, the minimum asymptotic variance of $\widehat{\beta}$ can be achieved by taking into account the conditional variance-covariance function $\sigma(t_1, t_2)$ of $N(t; S, \beta_0)$,

$$\begin{aligned} & S\left(t_m, Z_{H_{t_m}}, \beta_0\right)\left(1 - S\left(t_M, Z_{H_{t_M}}, \beta_0\right)\right) \\ & \quad + E\left[(V(t_m) - V(t_M))\frac{S\left(X, Z_{H_{t_M}}, \beta_0\right) - S\left(t_M, Z_{H_{t_M}}, \beta_0\right)}{S\left(X, Z_{H_{t_M}}, \beta_0\right)}|Z_{H_{t_M}}\right] \\ & \quad - E\left[(1 - V(t_m))S\left(t_M, Z_{H_{t_M}}, \beta_0\right)\frac{S\left(X, Z_{H_{t_m}}, \beta_0\right) - S\left(t_m, Z_{H_{t_m}}, \beta_0\right)}{S^2\left(X, Z_{H_{t_m}}, \beta_0\right)}|Z_{H_{t_M}}\right], \end{aligned} \tag{10}$$

for any $t_1, t_2 \in [0, \tau]$ with $t_m = \min\{t_1, t_2\}$ and $t_M = \max\{t_1, t_2\}$. Another improvement in efficiency is to take $\mathrm{d}W_{in}(t)/\mathrm{d}t$ as the reciprocal of a conditional variance estimator of $N_i(t; S, \beta_0), i = 1, \ldots, n$. However, it is usually complicated to estimate the above censoring distribution, especially in the presence of a high-dimensional covariate space. Currently, there is still no standard rule to choose $W_{in}(t)$'s in the estimation of β_0. The specifications of a uniform distribution over the interval of interest and an empirical distribution of X_i's are two simple ways in practical implementation. Although an estimator of the marginal survival function, say $S(t)$, is another popular candidate, more computational effort is usually required to obtain an initial estimator of β_0 and an appropriate bandwidth in the sample analogue of $E[1 - N(t; S, \beta_0)]$ $(= S(t))$. Different from the conclusion of Chiang and Huang (2012) for the case of complete failure time data, no competitive advantage can be gained by specifying this type of weight in our numerical experiments.

With specific covariates z_{H_t}, our estimation is also useful in predicting $S(t, z_{H_t}, \beta_0)$. Substituting $\widehat{\beta}$ for β, $K_2(\upsilon)$ for an arbitrary qth-order kernel function $K_q(\upsilon)$, and ς for h in (4), it is naturally estimated by the resulting estimator, which is denoted by

$$\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta}). \tag{11}$$

The bandwidth ς can be further chosen by the minimizer $\widetilde{\varsigma}$ of the following cross-validation sum of squares:

$$CV_1(\varsigma) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \left\{N_i\left(t; \widehat{S}_{\widehat{h}}^{-i}, \widehat{\beta}\right) - \left(1 - \widetilde{S}_\varsigma^{-i}\left(t, Z_{iH_t}, \widehat{\beta}\right)\right)\right\}^2 \mathrm{d}t. \tag{12}$$

The reason of using $\widehat{S}^{-i}_{\widehat{h}}(t, z_{H_t}, \widehat{\beta})$ in $N_i(t; \widehat{S}^{-i}_{\widehat{h}}, \widehat{\beta})$ is mainly to eliminate the effects of its asymptotic bias and variance in those of $\widetilde{S}^{-i}_{\varsigma}(t, Z_{iH_t}, \widehat{\beta})$, $i = 1, \ldots, n$. In the proposed estimator $\widetilde{S}_{\widetilde{\varsigma}}(t, z_{H_t}, \widehat{\beta})$, $\widetilde{\varsigma}$ is obtained by the following iterative scheme:

$$\widetilde{\varsigma}_{\text{new}} = \widetilde{\varsigma}_{\text{old}} - \left(\partial^2_{\varsigma} CV_1(\widetilde{\varsigma}_{\text{old}})\right)^{-1} \partial_{\varsigma} CV_1(\widetilde{\varsigma}_{\text{old}}). \tag{13}$$

In S2 of Supplementary Material (II), the data-driven bandwidth $\widetilde{\varsigma} = O_p(n^{-1/5})$ is shown to be optimal in terms of a certain asymptotic integrated mean squared error of $\widetilde{S}_{\varsigma}(t, z_{H_t}, \widehat{\beta})$.

Remark 1 Under assumption A1 and noninformative censoring, the proxy for the log-likelihood function of a random sample $\{(X_i, \delta_i, Z_{iH_{X_i}}) : 1 \le i \le n\}$ can be derived as follows:

$$\ell_{h_{12}}(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i \ln \widehat{\lambda}_{h_{12}}(X_i, Z_{it\beta}) + \ln \widehat{S}_{h_1}\left(X_i, Z_{iH_{X_i}}, \beta\right) \right\}, \tag{14}$$

where $h_{12} = (h_1, h_2)$ is a bandwidth vector and $\widehat{\lambda}_{h_{12}}(t, z_{t\beta}) = \int_0^{\tau} K_{q,h_2}(t - u) d_u \widehat{\Lambda}_{h_1}(u, z_{H_u}, \beta)$. Thus, the pseudo-maximum likelihood estimator (PMLE) $\bar{\beta}$, which is a maximizer of $\ell_{h_{12}}(\beta)$, is proposed as an alternative to the PILSE $\widehat{\beta}$. In contrast with the PILSE, this estimation can deal with internal time-dependent covariates. However, more effort is required in seeking an appropriate trimming sequence to trim away observations with small pseudo-hazard and survival function estimates in (14). The large sample properties of $\bar{\beta}$ are further established in S3 of Supplementary Material (II).

Remark 2 When time-independent covariates are considered, $\ell_{h_{12}}(\beta)$ will be approximately same as that of Strzalkowska-Kominiak and Cao (2014). With a second-order kernel function $K_2(\upsilon)$ in (14), the $\sqrt{n}$-consistency and asymptotic normality of $\bar{\beta}$ can be accomplished whenever $h_1 = O(n^{-\delta_1})$ and $h_2 = O(n^{-\delta_2})$ with $\delta_1 \in (1/8, 1/6)$ and $\delta_2 \in (1/4, 5/6 - 3\delta_1)$. As for the bandwidth selection, it is difficult and impractical to estimate two-separate-bandwidths in the cross-validated version of $\ell_{h_{12}}(\beta)$. Although the bandwidths $\bar{h}_1$ and $\bar{h}_2$ with $\bar{h}_1 = O_p(\bar{h}_2)$ have been recommended in the former work, the good asymptotic behavior of $\bar{\beta}$ cannot be guaranteed by such bandwidths. Even with the specification of a fourth-order kernel function $K_4(\upsilon)$ in (14), $\bar{\beta}$ is still sensitive to two-dimensional bandwidth estimators $(\check{h}_1, \check{h}_2)$ in the cross-validation counterpart of $\ell_{h_{12}}(\beta)$. In view of these drawbacks and an empirical insight of Huang and Chiang (2016), the PILSE should compare favorably with the PMLE, especially in the cases with small sample sizes or heavy censoring rates.

2.3 Inferences on $\widehat{\beta}$ and $\widetilde{S}_{\varsigma}(t, z_{H_t}, \widehat{\beta})$

For the sake of simplicity, β_0 is assumed to be an interior point of $\mathcal{B}$, and $F_C(t)$, $f_{Z_t}(z_t)$, $f_{Z_{t\beta}}(v)$, $\mathbb{P}_n$, and $\otimes$ are used to stand for the distribution function of C, the density function of Z_t, the density function of $Z_{t\beta}$, the empirical measure, and the Kronecker power, respectively. In addition, we define $\mathcal{Z}_t^{\delta_t} = \{z_t : \|z_t - z_t^*\| \le \delta_t, z_t^* \in \mathcal{Z}_t\}$ and $\mathcal{Z}_{t\beta_0}^{\delta_t} = \{z_{t\beta_0} : z_t \in \mathcal{Z}_t^{\delta_t}\}$ for given $\delta_t > 0$ and $t \in [0, \tau]$, $\mathcal{B}_n = \{\beta \in \mathcal{B} : \|\beta - \beta_0\| < b_0 n^{-1/2}\}$ for given $b_0 > 0$, $\xi(\beta) = \int_0^\tau \{N(t; S^{[0]}, \beta) - (1 - S^{[0]}(t, z_{H_t}, \beta))\} S^{[1]}(t, z_{H_t}, \beta) \mathrm{d}W(t)$, $V_1 = E[\xi^{\otimes 2}(\beta_0)]$, and $V_2 = E[\int_0^\tau \{S^{[1]}(t, Z_{H_t}, \beta_0) - (1 - V(t))\Lambda^{[1]}(t, Z_{H_t}, \beta_0)(1 - \Lambda^{[1]}(X, Z_{H_t}, \beta_0))\} S^{[1]\top}(t, Z_{H_t}, \beta_0) \mathrm{d}W(t)]$. As for the notations $\Lambda^{[\ell]}(t, z_{H_t}, \beta)$, $S^{[\ell]}(t, Z_{H_t}, \beta)$, and $H_{k_1k_2k_3}(t, z_t, \upsilon)$, their definitions are given in Supplementary Material (I).

The following conditions are further imposed for the main results:

A2. $\inf_{\{z_t \in \mathcal{Z}_t^{\delta_t}, t \in [0,\tau]\}} f_{Z_t}(z_t) > f_0$, $\inf_{\{z_{t\beta_0} \in \mathcal{Z}_{t\beta_0}^{\delta_t}, t \in [0,\tau]\}} f_{Z_{t\beta_0}}(z_{t\beta_0}) > f_1$, and $\inf_{\{z_t \in \mathcal{Z}_t^{\delta_t}, t \in [0,\tau]\}} P(X \ge \tau | z_{H_\tau}) > f_\tau$ for some positive constants f_0, f_1, and $f_\tau > 0$.

A3. $\partial_\upsilon^{k_3} H_{k_1k_2k_3}(t, z_t, \upsilon)$ is Lipschitz continuous in (t, υ) with a Lipschitz constant being independent of (t, z_t, υ).

A4. $h = h_0 n^{-\delta_0}$ for $\delta_0 \in (1/4q, 1/5)$ and some positive constant h_0.

A5. $W_{in}(t) - W_i(t) = \sum_{j=1}^n \eta_{ij}(t)/n + o_p(n^{-1/2})$ uniformly in (z_{iH_t}, t), where $\eta_{ij}(t), \ldots, \eta_{in}(t)$ are conditionally independent and identically distributed with $\sup_{(z_{iH_t}, t)} |E[\eta_{ij}(t) | z_{iH_t}]| = o_p(1)$ and $\sup_{(z_{iH_t}, t)} \mathrm{Var}(\eta_{ij}(t) | z_{iH_t}) = o_p(n^{-1/2})$, $i = 1, \ldots, n$.

A6. V_2 is nonsingular.

To achieve the $\sqrt{n}$-consistency of $\widehat{\beta}$, the involved bandwidth in $\partial_\beta^k \widehat{S}_h(t, z_{H_t}, \beta)$'s should be well controlled. The reason of not using a second-order kernel function $K_2(\upsilon)$ in (7), (8) is because the bandwidth selector $\widehat{h} = O_p(n^{-1/5})$ [see S1 of Supplementary Material (II)] will lead to a violation of assumption A4. Thus, a fourth-order kernel function $K_4(\upsilon)$ is suggested in practical implementation. As for the weight function $W_n(t)$, a uniform distribution, an empirical distribution of X, an estimator of $S(t)$, and an estimator of $S(t, z_{H_t}, \beta_0)$ can be justified to satisfy assumption A5.

The consistency and asymptotic normality of $\widehat{\beta}$ are established below.

Theorem 1 *Suppose that assumptions A1–A6 are satisfied. Then,*

$$\widehat{\beta} \xrightarrow{p} \beta_0 \text{ and } \sqrt{n}(\widehat{\beta} - \beta_0) \xrightarrow{d} N\left(0, V_2^{-1} V_1 V_2^{-1}\right). \quad (15)$$

Proof See Appendix B of Supplementary Material (I). □

By virtue of Theorem 1, the asymptotic variance of $\widehat{\beta}$ can be found to rely on the choice of weight functions $W_{in}(t)$'s in (7)–(8). As shown in Sect. 2.2, the optimal weights involve the conditional distribution of C, which is usually treated as a nuisance parameter in application and is not easily modeled in practice. For other feasible weight functions, we carefully investigate their effects on the variation of $\widehat{\beta}$ through the numerical experiments.

The next theorem gives the asymptotic Gaussian process of $\widetilde{S}_\varsigma(t, z_{H_t}, \widehat{\beta})$.

Theorem 2 *Under assumptions A1-A6 and assumption* (*A7.* $\varsigma = \varsigma_0 n^{-\kappa_1}$ *for* $\kappa_1 \in (1/8, 1/3)$ *and some positive constant* ς_0), *there exists a sequence of centered Gaussian* *processes* $G_n(t, z_{H_t})$ *with continuous sample paths and variance-covariance function* $\varsigma E[\Psi_n(t_1, z_{H_{t_1}})\Psi_n(t_2, z_{H_{t_2}})]$ *such that*

$$\sup_{(z_{H_t},t)} |\sqrt{n\varsigma}\left(\widetilde{S}_\varsigma\left(t, z_{H_t}, \widehat{\beta}\right) - S\left(t, z_{H_t}, \beta_0\right) - E\left[\Psi_n\left(t, z_{H_t}\right)\right]\right) - G_n\left(t, z_{H_t}\right)| = o_p(1), \tag{16}$$

where $\sup_{(z_{H_t},t)} |E[\Psi_n(t, z_{H_t})]| = O(\varsigma^2)$ *and* $\varsigma E[\Psi_n(t_1, z_{H_{t_1}})\Psi_n(t_2, z_{H_{t_2}})]$ *converges to some nondegenerate function of* $(t_1, z_{H_{t_1}})$ *and* $(t_2, z_{H_{t_2}})$ *with*

$$\Psi_n(t, z_{H_t}) = -S(t, z_{H_t}, \beta_0) \int_0^t \frac{K_\varsigma(Z_{u\beta_0} - z_{u\beta_0})}{H_{010}(u, z_u, z_{u\beta_0})}(dN(u) - Y(u)\lambda(u, z_{u\beta_0})du).$$

Proof See Appendix B of Supplementary Material (I). □

3 Estimation and inference for the time-dependent AUC

For the discrimination of time-independent marker values Z on cumulative cases $\{T \leq t\}$ versus dynamic controls $\{T > t\}$, Z_{β_0} can be shown to have the highest time-dependent ROC curve whenever $\lambda(t, \upsilon)$ in (1) is strictly increasing in υ. In terms of the definition by Heagerty et al. (2000), its time-dependent ROC curve is the spectrum of values for the true positive rate $P(Z_{\beta_0} > \upsilon | T \leq t)$ and the false positive rate $P(Z_{\beta_0} > \upsilon | T > t)$ over varying threshold values υ. The corresponding time-dependent AUC, denoted by $\mathcal{A}(t; \beta_0)$, also has an explicit probability expression $P(Z_{i\beta_0} > Z_{j\beta_0} | \widetilde{N}_i(t) = 1, \widetilde{N}_j(t) = 0)$. By replacing $\widetilde{N}(t)$ with $\widehat{N}(t; \widetilde{S}_\varrho, \widehat{\beta})$ in $\mathcal{A}(t; \beta_0)$, a sample analogue of the probability expression is proposed as an estimator and is compared with the Chambless-Diao type estimator. Moreover, the large sample properties of these AUC estimators are established under a very general censoring assumption.

3.1 Estimation of $\mathcal{A}(t; \beta_0)$

It follows from $E[\widetilde{N}(t)|X, Z, \delta] = N(t; S, \beta_0)$ and the probability expression of $\mathcal{A}(t; \beta_0)$ that

$$\begin{aligned} \mathcal{A}(t; \beta_0) = \frac{\mathcal{A}_1(t; \beta_0)}{\mathcal{A}_0(t; \beta_0)} \text{ with } \mathcal{A}_\ell(t; \beta_0) \\ = E\left[I^\ell(Z_{i\beta_0} > Z_{j\beta_0})N_i(t; S, \beta_0)(1 - N_j(t; S, \beta_0))\right], \ell = 0, 1. \end{aligned} \tag{17}$$

Substituting $N_i(t; \widetilde{S}_\varrho, \widehat{\beta})$'s into the sample analogues of $\mathcal{A}_\ell(t; \beta_0)$'s, an estimator of $\mathcal{A}(t; \beta_0)$ is given by

$$\bar{\mathcal{A}}(t; \widehat{\beta}) = \frac{\bar{\mathcal{A}}_1(t; \widehat{\beta})}{\bar{\mathcal{A}}_0(t; \widehat{\beta})} \text{ with } \bar{\mathcal{A}}_\ell(t; \widehat{\beta})$$

$$= \frac{1}{n(n-1)} \sum_{i \neq j} I^{\ell}(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i(t; \widetilde{S}_{\varrho}, \widehat{\beta})(1 - N_j(t; \widetilde{S}_{\varrho}, \widehat{\beta})), \ell = 0, 1. \tag{18}$$

In our estimation for $\mathcal{A}(t; \beta_0)$, the estimated counting process $N(t; \widetilde{S}_{\varrho}, \widehat{\beta})$ plays an important role. Moreover, the proposed estimator takes into account a commonly acceptable censoring mechanism and avoids the curse of dimensionality. As indicated by Chambless and Diao (2006), $\mathcal{A}(t; \beta_0)$ can also be expressed as

$$\frac{E[I(Z_{i\beta_0} > Z_{j\beta_0})(1 - S(t, Z_i, \beta_0))S(t, Z_j, \beta_0)]}{E[(1 - S(t, Z_i, \beta_0))S(t, Z_j, \beta_0)]}. \tag{19}$$

Thus, $\mathcal{A}(t; \beta_0)$ can be naturally estimated by

$$\breve{\mathcal{A}}(t; \beta) = \frac{\breve{\mathcal{A}}_1(t; \widehat{\beta})}{\breve{\mathcal{A}}_0(t; \widehat{\beta})} \text{ with } \breve{\mathcal{A}}_{\ell}(t; \widehat{\beta})$$
$$= \frac{1}{n(n-1)} \sum_{i \neq j} I^{\ell}(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}})(1 - \widetilde{S}_{\varrho}(t, Z_i, \widehat{\beta}))\widetilde{S}_{\varrho}(t, Z_j, \widehat{\beta}), \ell = 0, 1. \tag{20}$$

Basically, this type of estimation is an extension of Chiang and Hung (2010) for the discriminability of a single marker. Since the underlying counting process $\widetilde{N}(t)$ is completely disregarded in (20), $\breve{\mathcal{A}}(t; \widehat{\beta})$ is heavily affected by the inherent biases of $\widetilde{S}_{\varrho}(t, Z_i, \widehat{\beta})$'s and might suffer from a substantial bias. On the other hand, the bias effect on $\bar{\mathcal{A}}(t; \widehat{\beta})$ is generally negligible whenever the random quantity $\sum_{i=1}^{n}(1 - V_i(t))$ is small. Although the asymptotic variances of $\bar{\mathcal{A}}(t; \widehat{\beta})$ and $\breve{\mathcal{A}}(t; \widehat{\beta})$ are not exactly the same, the variance of $\bar{\mathcal{A}}(t; \widehat{\beta})$ is relatively small in most cases of our numerical experiments. In light of these advantages, $\bar{\mathcal{A}}(t; \widehat{\beta})$ is preferred to $\breve{\mathcal{A}}(t; \widehat{\beta})$ for finite samples.

For the involved bandwidths in (18) and (20), neither $\widehat{h}$ nor $\widetilde{\varsigma}$ can guarantee the $\sqrt{n}$-consistency of $\bar{\mathcal{A}}(t; \widehat{\beta})$. To resolve this problem, ϱ is chosen by the minimizer $\widetilde{\varrho}$ of the following cross-validation sum of squares:

$$CV_2(\varrho) = \frac{1}{n(n-1)} \sum_{i \neq j} \int_0^{\tau} \Big\{ I(Z_{i\widehat{\beta}} > Z_{j\widehat{\beta}}) N_i\left(t; \widehat{S}_{\widehat{h}}^{-i}, \widehat{\beta}\right)$$
$$\left(1 - N_j\left(t; \widehat{S}_{\widehat{h}}^{-j}, \widehat{\beta}\right)\right) - \bar{\mathcal{A}}_1^{-ij}(t; \widehat{\beta}) \Big\}^2 \, dt. \tag{21}$$

By virtue of Lemma 3 in Appendix A of Supplementary Material (I), we further derive in S4 of Supplementary Material (II) that $CV_2(\varrho) = O_p(n^{-1/2} + \varrho^2 + (n\varrho)^{-3/4})$ and $\widetilde{\varrho} = O_p(n^{-3/11})$ satisfies the constraint in assumption A8 of the next subsection.

3.2 Asymptotic properties of $\bar{\mathcal{A}}(t; \widehat{\beta})$ and $\breve{\mathcal{A}}(t; \widehat{\beta})$

It is noted that the sets $\mathcal{Z}_t$, $\mathcal{Z}_t^{\delta_t}$, and $\mathcal{Z}_{t\beta_0}^{\delta_t}$ can be simplified to be invariant with respect to t for time-independent covariates Z. As one shall see in Appendix B of Supplementary Material (I), the uniform approximation of $\widetilde{S}_{\varrho}(t, z, \beta_0)$ and the $i.i.d.$ approximation of $(\widehat{\beta} - \beta_0)$ are essential ingredients in deriving the asymptotic behaviors of $\bar{\mathcal{A}}(t; \widehat{\beta})$ and $\breve{\mathcal{A}}(t; \widehat{\beta})$. In the succeeding discussions, some assumptions are made as follows:

A7. $\partial_\beta H_{\ell_1\ell_2 k}(t, z, z_\beta)$ is Lipschitz continuous in β with a Lipschitz constant being independent of (t, z, z_β), $\ell_1, \ell_2 = 0, 1$, $k = 0, 1, 2$.
A8. $\varrho = \varrho_0 n^{-\kappa_2}$ for $\kappa_2 \in (1/4, 1/3]$ and some positive constant ϱ_0.

Since the selected bandwidth $\widetilde{\varrho}$ satisfies the constraint in assumption A8, both of $\bar{\mathcal{A}}(t; \widehat{\beta})$ and $\breve{\mathcal{A}}(t; \widehat{\beta})$ can achieve the $\sqrt{n}$-consistency. The notations used in the following theorem are further defined in Supplementary Material (I).

Theorem 3 *Suppose that the conditions in Theorem 1 and assumptions A7-A8 are satisfied. Then,*

(i) $\sup_{\{t\in[0,\tau]\}} |\bar{\mathcal{A}}(t; \widehat{\beta}) - \mathcal{A}_t(\beta_0)| \xrightarrow{p} 0$ *and* $\sup_{\{t\in[0,\tau]\}} |\breve{\mathcal{A}}(t; \widehat{\beta}) - \mathcal{A}(t; \beta_0)| \xrightarrow{p} 0$.

(ii) $\sqrt{n}(\bar{\mathcal{A}}(t; \widehat{\beta}) - \mathcal{A}(t; \beta_0))$ *and* $\sqrt{n}(\breve{\mathcal{A}}(t; \widehat{\beta}) - \mathcal{A}_t(\beta_0))$ *converge to mean zero Gaussian processes with the respective covariance functions*

$$\nu_1(t_1, t_2) = \frac{E[(\xi_0(t_1) + \xi_1(t_1) + \xi_2(t_1)(\xi_0(t_2) + \xi_1(t_2) + \xi_2(t_2))]}{\mathcal{A}_0(t_1; \beta_0)\mathcal{A}_0(t_2; \beta_0)} \text{ and}$$

$$\nu_2(t_1, t_2) = \frac{\sum_{k=1}^{2} E[\zeta_k(t_1)\zeta_k(t_2)]}{\mathcal{A}_0(t_1; \beta_0)\mathcal{A}_0(t_2; \beta_0)}.$$

Proof See Appendix B of Supplementary Material (I). □

After some algebraic manipulations, $\nu_2(t, t)$ is readily shown to be $E[(\xi_1(t) + \xi_2(t))^2]/\mathcal{A}_0^2(t; \beta_0)$. Let $F_C(t|z)$ and $F_C(t, v)$ stand for the distributions of C given $Z = z$ and $Z_{\beta_0} = v$, respectively. When C is independent of (T, Z) or $F_C(t|z) = F_C(t, z_{\beta_0})$, $\xi_0(t) = 0$ can be derived by the fact that $E[\Lambda^{[1]}(t, Z, \beta_0)|Z_{\beta_0}] = 0$. As a result, both of the AUC estimators have the same asymptotic variance. However, neither $\bar{\mathcal{A}}(t; \widehat{\beta})$ nor $\breve{\mathcal{A}}(t; \widehat{\beta})$ is universally better than the other in terms of their asymptotic variances.

4 Monte carlo simulations

In our numerical experiments, data were repeatedly generated 1000 times with sample sizes of 75, 125, and 250 and censoring rates of 20 and 40%. To simplify our assessment for the estimators of β_0, $S(t, z_{H_t}, \beta_0)$, and $\mathcal{A}(t; \beta_0)$, the setup of time-dependent covariates was omitted in the investigation of estimators of β_0 and $S(t, z_{H_t}, \beta_0)$. The time-independent covariates $Z = (Z_1, Z_2, Z_3)^\top$ were designed to follow a trivariate normal distribution with a mean of zero, a standard deviation of one, and a pairwise correlation of 0.5. Conditioning on $Z = z$, T and C were assumed to be independent with the respective hazard rate functions

$$\lambda_T(t|z) = \frac{\exp(-z_{\beta_0} + 5)}{2(\exp(-z_{\beta_0} + 5) - t)^2} I(0 < t < \exp(-z_{\beta_0} + 5)) \tag{22}$$

$$\text{and } \lambda_C(t|z) = \frac{1}{c_0 t} I\left(\exp\left(\frac{z_1 - z_3}{c_0} + 4\right) < t < \exp\left(\frac{z_1 - z_3}{c_0} + 4 + c_0\right)\right), \tag{23}$$

where $\beta_0 = (-1, 0.2)^\top$ and c_0 was specified to produce the expected censoring rates. Instead of designing the widely applied relative risk model, additive risk model, and extended accelerated failure time model, such a hazard formulation is used to emphasize the flexibility of a general single-index hazards model. Due to the difficulty and complexity in directly estimating the asymptotic distributions of the estimators, a random weighted bootstrap technique for semiparametric models (see Kosorok 2008) was employed. Independent of survival data, exchangeable random weights were repeatedly generated 500 times based on $i.i.d.\ Gamma(4, 2)$ random variables, which have better numerical results than others, with a scale factor modification of 0.5 for the variability in the random weights. Further, simultaneous confidence bands were constructed for $S(t, z, \beta_0)$ over $\mathcal{T}_q(z_{\beta_0}) = [0, t_q(z_{\beta_0})]$ and $\mathcal{A}_t(\beta_0)$ over $\mathcal{T}_q = [0, t_q]$, where $t_q(z_{\beta_0})$ and t_q are the respective qth quantiles of $S(t, z, \beta_0)$ and $S(t)$ with $q = 0.65$, 0.75, and 0.85.

In this study, the performances of the PILSEs and the PMLEs were assessed and compared on a variety of simulation settings. To investigate the effects of weight functions $W_{in}(t)$'s on the precision of $\widehat{\beta}$, a uniform distribution, an empirical distribution of X, and an estimator $\widehat{S}(t) = \sum_{i=1}^n \widetilde{S}_\varsigma(t, Z_{iH_t}, \widehat{\beta})/n$ of the marginal survival function $S(t)$ were specified in (7)–(8) over $(0, \max\{X_i : \delta_i = 1\})$. For ease of presentation, the resulting PILSEs are represented by $\widehat{\beta}$(unif.), $\widehat{\beta}$(empi.), and $\widehat{\beta}$(surv.), respectively. As for the bandwidth ς in the summands of $\widehat{S}(t)$, it was chosen by

$$\widehat{\varsigma} = \arg\min_{\varsigma} \frac{1}{n} \sum_{i=1}^{n} \int \left(N_i\left(t; \widehat{S}_{\widehat{h}}, \widehat{\beta}\right) - \widehat{S}^{-i}(t)\right)^2 \mathrm{d}t. \tag{24}$$

As shown by Chiang et al. (2016), $\widehat{S}(t)$ has the rate of convergence $O_p(n^{-3/11})$ but the chosen bandwidths in (9) and (12) cannot attain the $\sqrt{n}$-consistency of $\widehat{S}(t)$. The PILSE $\widehat{\beta}(K_2)$ with a uniform distribution weight and a second-order kernel function in (7)–(8) was also given to illustrate a major impact on the bandwidth selection. By specifying second and fourth-order kernel functions in $\ell_{h_{12}}(\beta)$, the derived PMLEs, which are denoted by $\bar{\beta}(K_2)$ and $\bar{\beta}$ with $\bar{\beta}(K_2)$ being asymptotically equivalent to that of Strzalkowska-Kominiak and Cao (2014), were further considered in this numerical investigation. Due to the poor performance of two-separate-bandwidths, both of the PMLEs were computed with bandwidths of the form $(\widetilde{h}_1, \widetilde{h}_2) = \widetilde{h}_0(\widehat{\sigma}(X), \widehat{\sigma}(Z_{\bar{\beta}}))$, where $\widehat{\sigma}$ is the sample standard deviation. As evidenced by the simulation findings in Strzalkowska-Kominiak and Cao (2013), we studied their PMLE $\bar{\beta}^*$ because it is comparable or even better than that of Bouaziz and Lopez (2010).

It is revealed in Table 1 that $\widehat{\beta}$(unif.), $\widehat{\beta}$(empi.), and $\widehat{\beta}$(surv.) are generally comparable and outperform $\widehat{\beta}(K_2)$, $\bar{\beta}$, $\bar{\beta}(K_2)$, and $\bar{\beta}^*$. One can also observe in this table that the bias magnitudes of $\bar{\beta}$ and $\bar{\beta}(K_2)$ are much larger than those their competitors for small samples and heavy censoring rates. Due to an inappropriate bandwidth for the maximizer $\bar{\beta}(K_2)$ of $\ell_{h_{12}}(\beta)$, the performance of $\bar{\beta}$ is expected to be better than that of $\bar{\beta}(K_2)$. Compared with $\bar{\beta}$ and $\bar{\beta}(K_2)$, $\bar{\beta}^*$ has a relatively small standard deviation. For moderate samples, the averages of 1000 bootstrap standard errors and the empirical coverage probabilities are fairly close to the standard deviations of their 1000 estimates and the nominal level of 0.95 (Table 2). Conditioning on the first, second, and third quartiles of Z_{β_0}, we further display the survival functions, the mean and standard deviation curves of 1000 estimated survival functions, and the averages of 1000 bootstrap standard error curves in Figs. 1, 2, and 3 and Figures S1–S3 of Supplementary Material (II). When the sample size is moderate or the censoring rate is low, the true and estimated curves are very similar to each other. Moreover, it is found in Table 3 that the simultaneous coverage probabilities are quite near the nominal level of 0.95 for moderate samples.

The second simulation study aims at assessing the performance of $\bar{\mathcal{A}}(t;\widehat{\beta})$ and making a comparison with the Chambless-Diao type estimator $\breve{\mathcal{A}}(t;\widehat{\beta})$. For the computation of the time-dependent AUC estimators, the involved bandwidths were chosen by the minimizer $\widetilde{\varrho}$ of a cross-validation criterion in (21). In Fig. 4 and Figure S4 of Supplementary Material (II), one can observe that $\breve{\mathcal{A}}(t;\widehat{\beta})$ systematically deviates from $\mathcal{A}(t;\beta_0)$, whereas $\bar{\mathcal{A}}(t;\widehat{\beta})$ has a very small bias. Therefore, an extremely large sample size is generally expected to alleviate the bias problem in $\breve{\mathcal{A}}(t;\widehat{\beta})$. Furthermore, the variance of $\breve{\mathcal{A}}(t;\widehat{\beta})$ is slightly larger than that of $\bar{\mathcal{A}}(t;\widehat{\beta})$. As for the bootstrap standard errors, their averages are found to increase toward the standard deviations of 1000 estimates as the sample size increases or the censoring rate decreases. The simultaneous coverage probabilities of $\mathcal{A}(t;\beta_0)$, which are exhibited in Table 4, also stay around the nominal level of 0.95 for moderate samples.

5 Applications

The Mayo PBC data set was first used to illustrate the applicability of our estimation approach for the index coefficients and the survival functions of interest. In an angiography cohort study, the proposed estimator for the time-dependent AUC was further used to assess the overall classification ability of the optimal composite plasma biomarker. Based on the computed pseudo-residuals $e_{it} = N_i(t;\widehat{S}_{\widehat{h}},\widehat{\beta}) - (1 - \widetilde{S}_{\widetilde{\varsigma}}(t, Z_{iH_t},\widehat{\beta}))$, $i = 1,\ldots,n$, the test rules established by Chiang and Huang (2012) can be directly applied to examine the correctness of the SIH model.

5.1 Application to a primary biliary cirrhosis study

The data presented by Fleming and Harrington (1991) were collected from the Mayo clinical trial study. Among 418 patients, who participated in this study on PBC of

Table 1 The means (Mean) and the standard deviations (SD) of the PILSEs $\widehat{\beta}$(unif.), $\widehat{\beta}$(empi.), $\widehat{\beta}$(surv.), and $\widehat{\beta}(K_2)$, and the PMLEs $\bar{\beta}$ and $\bar{\beta}(K_2)$

(n,c.r.) (%)	β_0	$\widehat{\beta}$(unif.)		$\widehat{\beta}$(empi.)		$\widehat{\beta}$(surv.)		$\widehat{\beta}(K_2)$		$\bar{\beta}$		$\bar{\beta}(K_2)$		$\bar{\beta}^*$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
(75.20)	−1.0	−0.99	0.106	−1.00	0.089	−1.00	0.082	−1.01	0.124	−1.03	0.209	−1.01	0.251	−1.00	0.140
	0.2	0.20	0.109	0.19	0.087	0.19	0.073	0.21	0.126	0.21	0.178	0.21	0.151	0.20	0.128
(75.40)	−1.0	−1.00	0.124	−1.00	0.135	−1.00	0.117	−1.00	0.164	−1.04	0.240	−1.05	0.289	−1.02	0.182
	0.2	0.20	0.115	0.20	0.141	0.20	0.139	0.24	0.167	0.21	0.222	0.25	0.354	0.23	0.171
(125.20)	−1.0	−1.00	0.083	−0.99	0.066	−0.99	0.051	−1.01	0.112	−1.02	0.161	−1.02	0.182	−1.00	0.108
	0.2	0.20	0.081	0.19	0.058	0.19	0.060	0.21	0.100	0.20	0.141	0.21	0.169	0.20	0.097
(125.40)	−1.0	−1.00	0.104	−1.00	0.116	−1.00	0.108	−1.01	0.133	−1.02	0.170	−1.02	0.188	−1.01	0.119
	0.2	0.20	0.110	0.20	0.102	0.20	0.104	0.24	0.129	0.20	0.173	0.19	0.178	0.22	0.120
(250.20)	−1.0	−1.00	0.045	−1.00	0.041	−1.00	0.036	−1.00	0.075	−1.01	0.098	−1.01	0.144	−1.00	0.062
	0.2	0.20	0.043	0.20	0.040	0.20	0.037	0.20	0.071	0.20	0.078	0.21	0.133	0.20	0.067
(250.40)	−1.0	−1.00	0.064	−0.99	0.065	−0.99	0.069	−1.00	0.097	−1.01	0.140	−1.01	0.159	−1.00	0.086
	0.2	0.20	0.073	0.20	0.080	0.20	0.085	0.23	0.087	0.19	0.121	0.20	0.141	0.21	0.083

Table 2 The standard deviations (SD) and bootstrap standard errors (B.S.E.) of $\widehat{\beta}$(unif.), the quantile intervals (Q.I.), the bootstrap confidence intervals (B.C.I.), and the coverage probabilities (C.P.)

c.r.	β_0	20%					40%				
n		SD	B.S.E	Q.I.	B.C.I	C.P.	SD	B.S.E	Q.I.	B.C.I	C.P.
75	−1.0	0.106	0.123	(−1.228, −0.780)	(−1.123, −0.738)	0.943	0.124	0.148	(−1.247, −0.763)	(−1.297, −0.704)	0.967
	0.2	0.109	0.123	(0.005, 0.434)	(−0.048, 0.445)	0.961	0.115	0.146	(0.010, 0.417)	(−0.090, 0.494)	0.961
125	−1.0	0.083	0.087	(−1.184, −0.876)	(−1.175, −0.821)	0.950	0.104	0.102	(−1.204, −0.812)	(−1.201, −0.793)	0.942
	0.2	0.081	0.086	(0.052, 0.373)	(0.012, 0.385)	0.945	0.110	0.116	(−0.020, 0.388)	(−0.049, 0.417)	0.944
250	−1.0	0.045	0.043	(−1.092, −0.905)	(−1.079, −0.906)	0.947	0.064	0.070	(−1.189, −0.906)	(−1.146, −0.868)	0.955
	0.2	0.043	0.042	(0.102, 0.289)	(0.113, 0.282)	0.951	0.073	0.077	(0.024, 0.304)	(0.053, 0.362)	0.954

the liver between 1974 and 1984, 185 died during the study period. The primary research interest focuses on investigating the influences of some prognostic factors on the number of days between registration and death. Measurements taken in our data analysis included age in days (age), albumin in gm/dl (albumin), serum bilirubin in mg/dl (bilirubin), presence of edema (edema), and prothrombin time in seconds (protime).

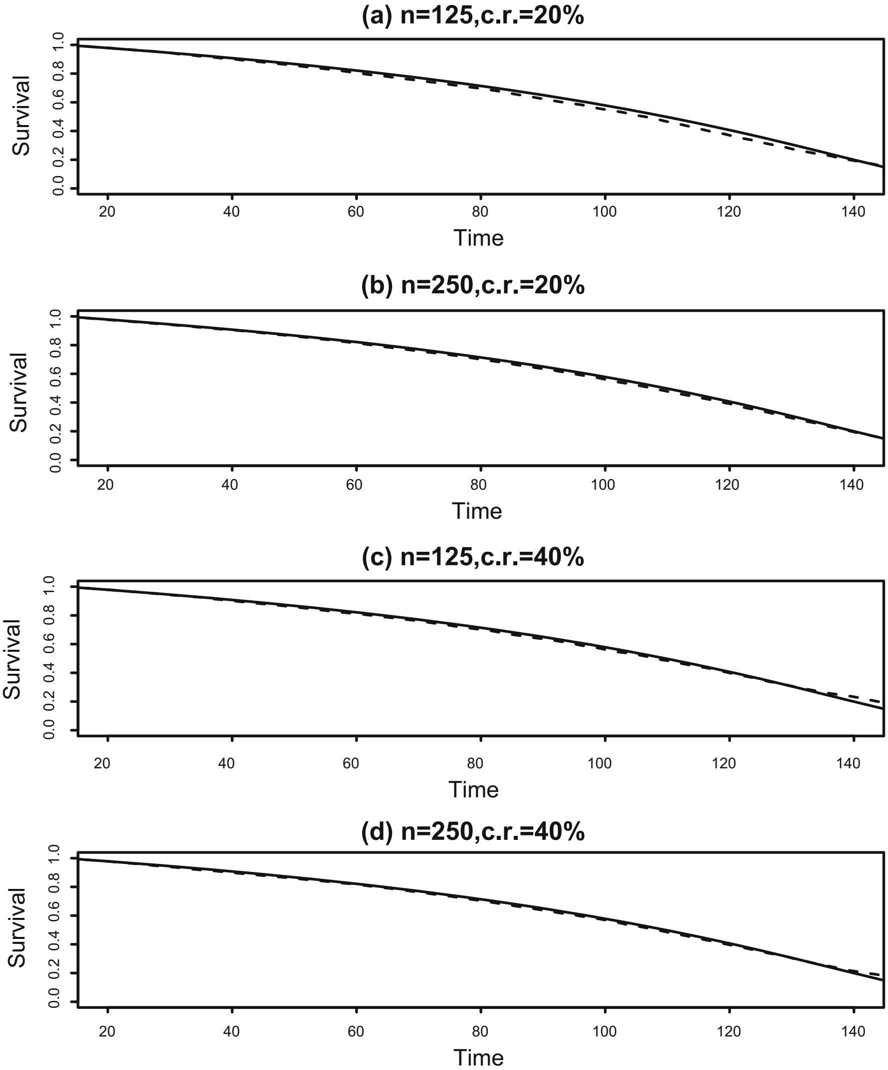

Fig. 1 **a–d** The survival functions (*solid curves*) conditioning on $Z_{\beta_0} = SI_{0.25}$ and the estimated conditional survival functions (*dashed curve*). **e**, **f** The standard deviation curves (*solid curves*) and the bootstrap standard error curves (*dashed curves*) of the estimated conditional survival functions

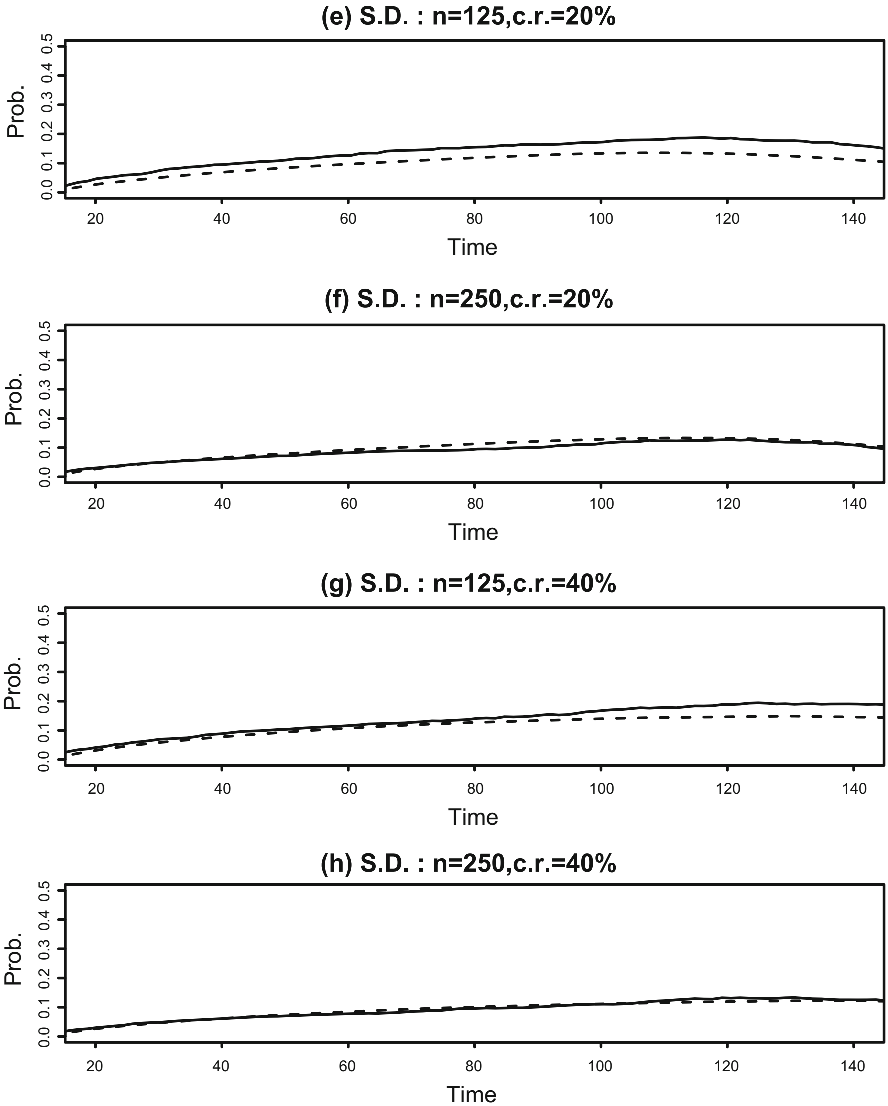

Fig. 1 continued

To characterize the relative effects of age, log(albumin), log(bilirubin), and edema, compared to log(protime), on the death time, a SIH model was fitted with the linear predictor

$$\log(\text{protime}) + \beta_{02}\text{age} + \beta_{03}\log(\text{albumin}) + \beta_{04}\log(\text{bilirubin}) + \beta_{05}\mathit{edema}.$$

The PILSE (0.026, −2.866, 0.905, 1.239) of $(\beta_{02}, \beta_{03}, \beta_{04}, \beta_{05})$ is computed with the bootstrap standard error vector (0.0081, 0.4391, 0.1754, 0.3202). Meanwhile, the bandwidth $\widehat{h} = 0.864$ is obtained from the cross-validation estimating equation in (8). The corresponding 95% bootstrap confidence intervals of the coefficients are further constructed as (0.0116, 0.0435), (−3.4035, −1.6104), (0.6101, 1.3589), and (0.5385, 1.9375). In this data analysis, the estimates and inferences are found to have a comparable degree of agreement with those of Zeng and Lin (2007) for an extended AFT model. As for the correctness of the fitted SIH model, it is ascertained by the

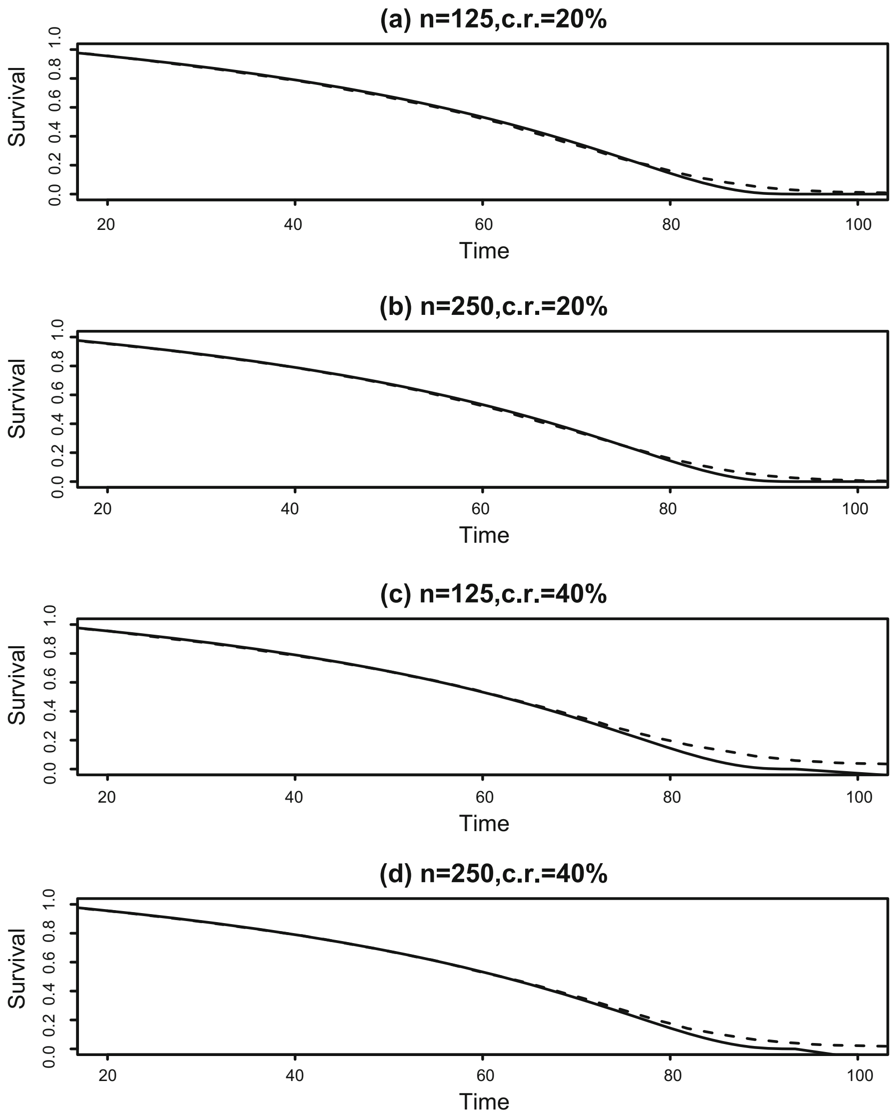

Fig. 2 **a–d** The survival functions (*solid curves*) conditioning on $Z_{\beta_0} = SI_{0.5}$ and the estimated conditional survival functions (*dashed curves*). **e–f** The standard deviation curves (*solid curves*) and the bootstrap standard error curves (*dashed curves*) of the estimated conditional survival functions

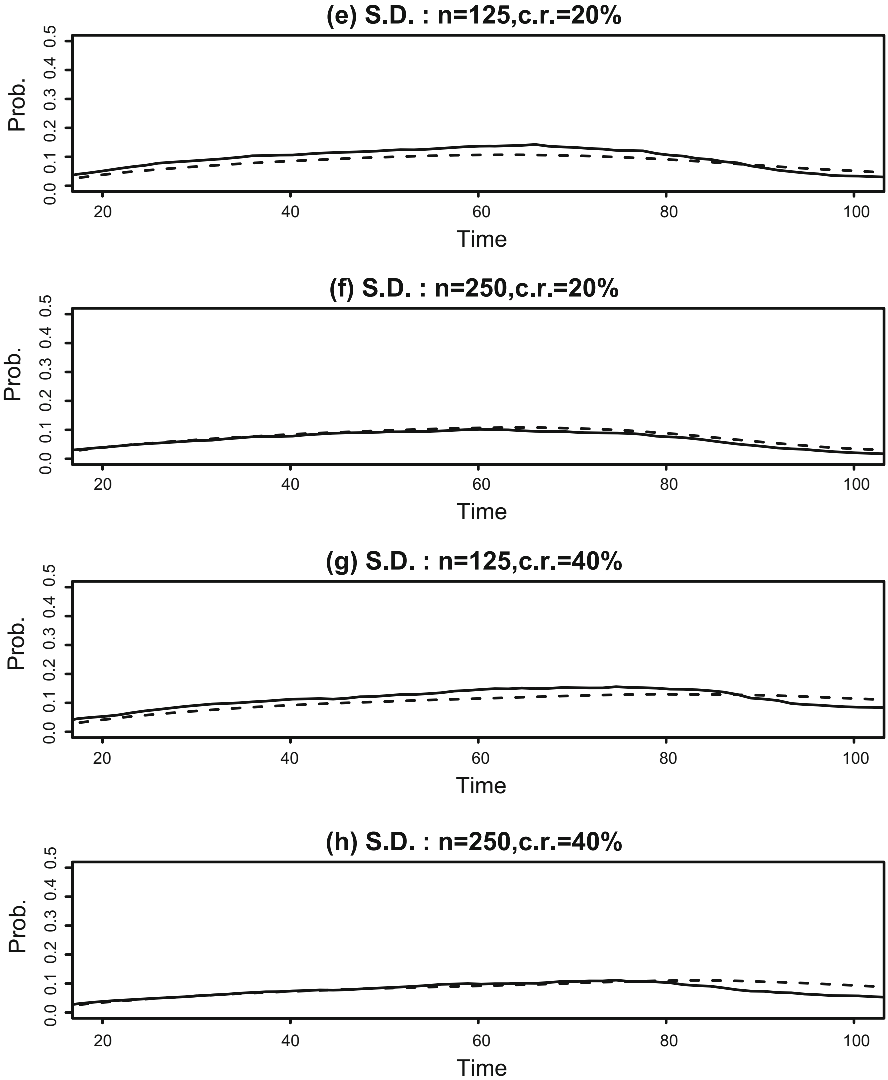

Fig. 2 continued

equality of the single-index cross-validation measure $SIC_n = 0.0996$ and the total residual sum of squares $TSS_n = 0.0996$. This conclusion is also evidenced by the F-test statistic of 0.9894 with the bootstrap p value of 0.338.

Provided that the conditional mean $E[T|Z = z]$ exists and is not constant for all possible values of z, the estimation procedure of Lopez et al. (2013) was implemented to estimate their proposed mean regression single-index model. The estimated linear predictor $log(protime) + 0.021age - 0.315log(albumin) + 0.522log(bilirubin) + 1.438edema$ appears to closely match our one, i.e. $Z_{\widehat{\beta}}$. This conclusion is further ensured by the extremely high canonical correlation of 0.992 between both of the estimated directions. As one can expect, the corresponding bootstrap standard errors (0.0294, 1.1499, 0.1797, 0.2701) of the parameter estimates are somewhat larger than ours because this estimation is rather sensitive to the presence of outliers in data. In Fig. 5a–c, the estimated survival functions, which are computed with the chosen bandwidth $\widetilde{\varsigma} = 1.271$, are presented together with their 95% simultaneous bootstrap confidence bands. These figures partly justify the qualitative structure that the survival function is a nonincreasing function of the SI at each time.

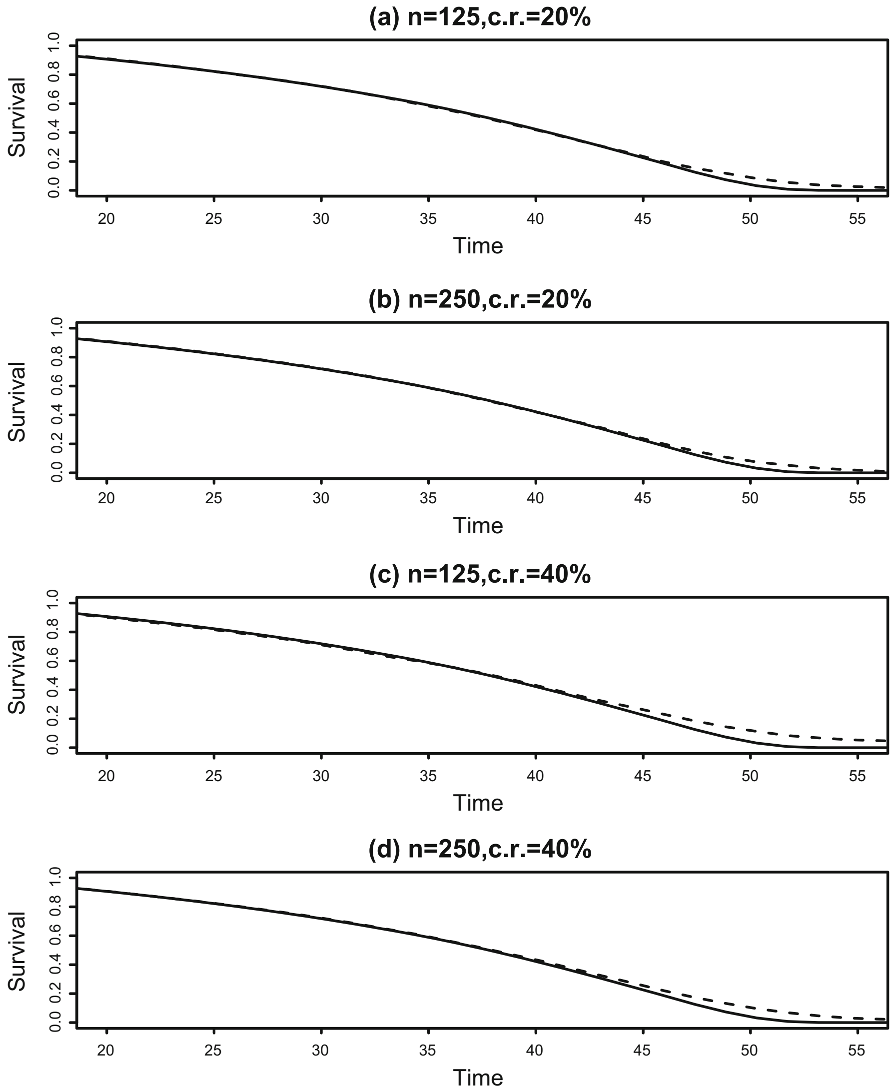

Fig. 3 **a–d** The survival functions (*solid curves*) conditioning on $Z_{\beta_0} = SI_{0.75}$ and the estimated conditional survival functions (*dashed curves*). **e**, **f** The standard deviation curves (*solid curves*) and the bootstrap standard error curves (*dashed curves*) of the estimated conditional survival functions

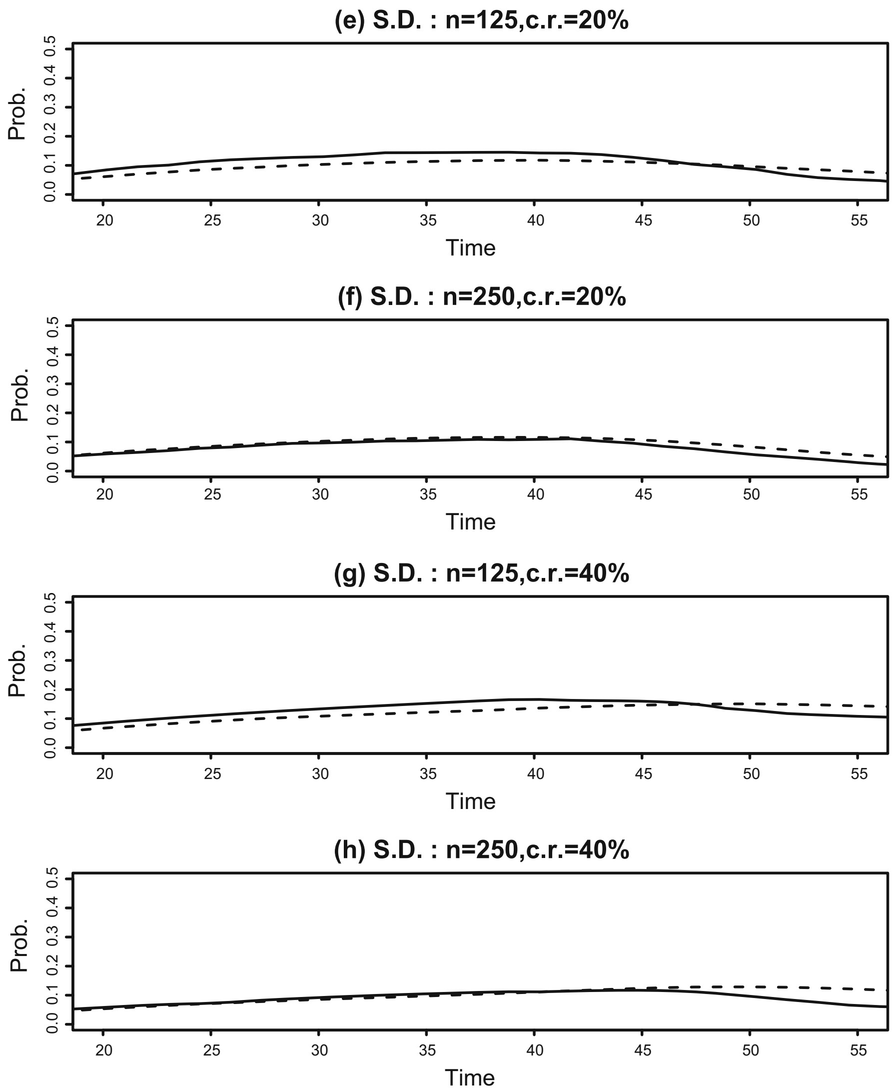

Fig. 3 continued

5.2 Application to an angiography cohort study

The second survival data were obtained from the BCVS database. In this angiography cohort study, 1050 patients were recruited between 1993 and 1995 from two Vancouver teaching hospitals for selective coronary angiography and 231 of the traceable patients died within 3500 days. In the work of Lee et al. (2006), the baseline plasma levels of C-reactive protein (CRP), serum amyloid A protein (SAA), interleukin IL-6 (IL-6), and total homocysteine (tHcy) were considered to be linked to the death.

Table 3 The simultaneous coverage probabilities of the conditional survival functions over different time periods

(n, c.r.) (%)	$Z_{\beta_0} = SI_{0.25}$			$Z_{\beta_0} = SI_{0.5}$			$Z_{\beta_0} = SI_{0.75}$		
	$\mathcal{T}_{0.85}(z_{\beta_0})$	$\mathcal{T}_{0.75}(z_{\beta_0})$	$\mathcal{T}_{0.65}(z_{\beta_0})$	$\mathcal{T}_{0.85}(z_{\beta_0})$	$\mathcal{T}_{0.75}(z_{\beta_0})$	$\mathcal{T}_{0.65}(z_{\beta_0})$	$\mathcal{T}_{0.85}(z_{\beta_0})$	$\mathcal{T}_{0.75}(z_{\beta_0})$	$\mathcal{T}_{0.65}(z_{\beta_0})$
(75.20)	0.942	0.940	0.941	0.940	0.941	0.941	0.952	0.948	0.946
(75.40)	0.973	0.971	0.970	0.968	0.968	0.970	0.972	0.967	0.965
(125.20)	0.953	0.950	0.943	0.948	0.948	0.944	0.961	0.957	0.957
(125.40)	0.961	0.955	0.946	0.962	0.954	0.958	0.953	0.953	0.955
(250.20)	0.954	0.951	0.947	0.948	0.947	0.945	0.950	0.951	0.953
(250.40)	0.957	0.956	0.956	0.954	0.953	0.952	0.953	0.951	0.946

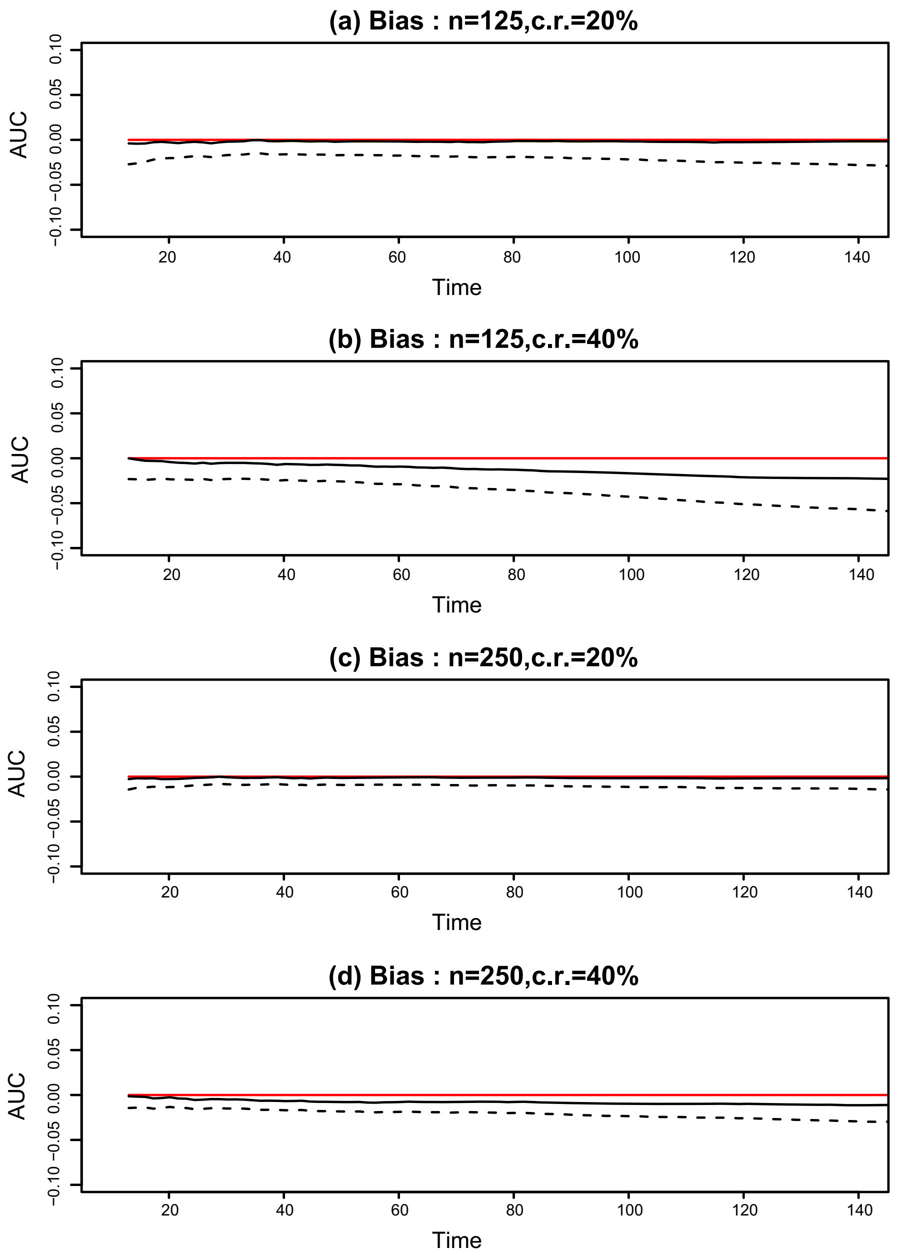

Fig. 4 **a**–**d** The bias curves (*solid* and *dashed curves*) of $\bar{A}(t;\widehat{\beta})$ and $\breve{A}(t;\widehat{\beta})$ with a *horizontal line* at zero (*gray line*). **e**–**h** The standard deviation curves (*solid curve* and *dashed curves*) of $\bar{A}(t;\widehat{\beta})$ and $\breve{A}(t;\widehat{\beta})$ and the bootstrap standard error curves (*dotted-dashed curves*) of $\bar{A}(t;\widehat{\beta})$

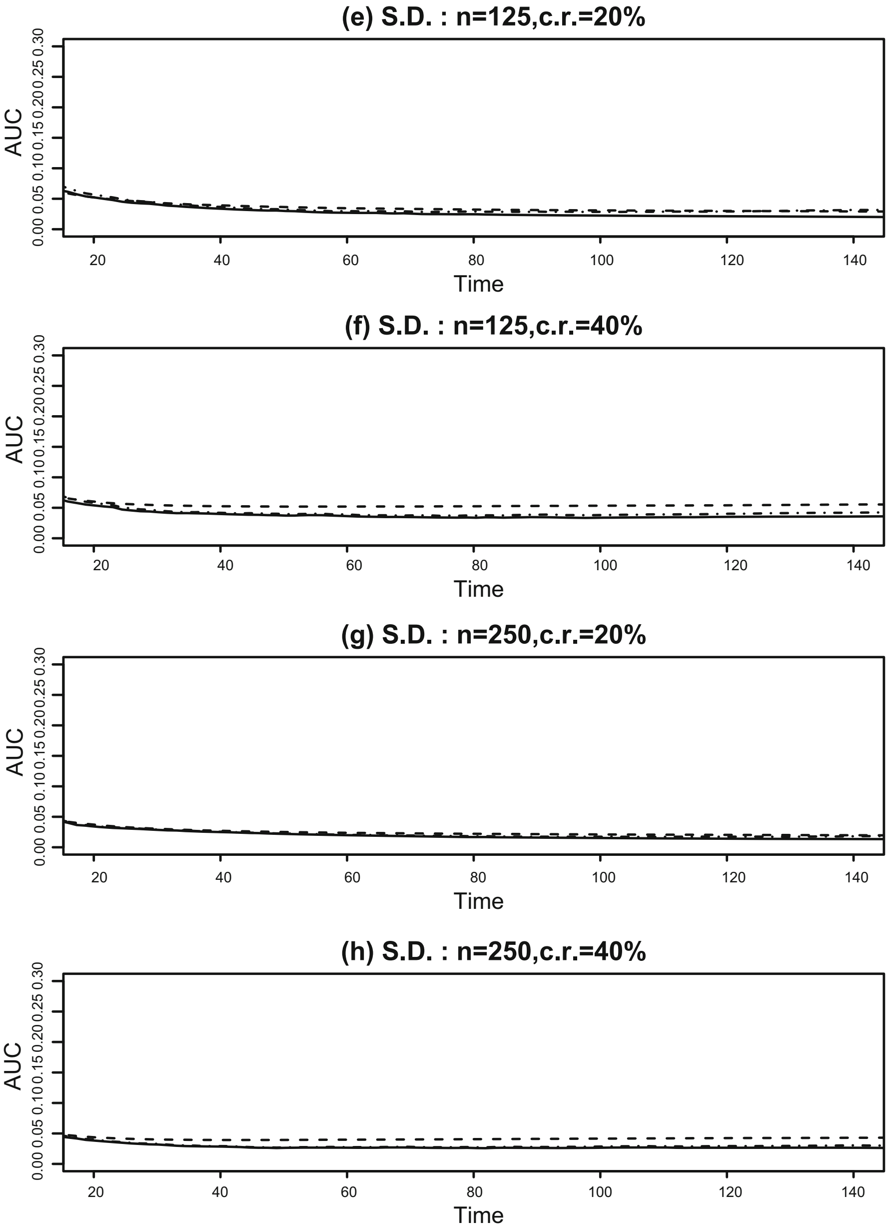

Fig. 4 continued

In our data analysis, a more flexible SIH model was employed to predict the all-cause death time with the single-index of the form

$$\mathrm{tHcy} + \beta_{02}\mathrm{CRP} + \beta_{03}(\mathrm{SAA}/100) + \beta_{04}IL\text{-}6.$$

Table 4 The simultaneous coverage probabilities of the AUC curve over different time periods

c.r.	20%			40%		
n	$\mathcal{T}_{0.85}(z_{\beta_0})$	$\mathcal{T}_{0.75}(z_{\beta_0})$	$\mathcal{T}_{0.65}(z_{\beta_0})$	$\mathcal{T}_{0.85}(z_{\beta_0})$	$\mathcal{T}_{0.75}(z_{\beta_0})$	$\mathcal{T}_{0.65}(z_{\beta_0})$
75	0.945	0.946	0.940	0.972	0.971	0.968
125	0.954	0.952	0.954	0.953	0.953	0.955
250	0.950	0.951	0.953	0.954	0.953	0.950

The index coefficient vector of the composite biomarker value is estimated by $(-0.221, 0.285, 1.124)$ with the bootstrap standard error vector $(0.0449, 0.0304, 0.0380)$. Meanwhile, the chosen bandwidth $\widehat{h} = 1.186$ is chosen by our cross-validation estimating equation. All the relative effects $(\beta_{02}, \beta_{03}, \beta_{04})$ are further detected to be significant from their 95% bootstrap confidence intervals $((-0.3642, -0.1660), (0.2004, 0.3322), (1.0511, 1.1942))$. Moreover, there is no significant evidence to reject the correctness of a fitted SIH model through the test statistic $F_n = 1$ with the bootstrap p value of 0.518 and the equality $\text{SIC}_n = \text{TSS}_n = 0.1478$. It follows from this conclusion that the time-varying logistic regression model, which was employed in the data analysis of Chiang and Huang (2009), might be questionable.

Once again, we apply the approach of Lopez et al. (2013) to estimate the index coefficients of a mean regression single-index model. The relative effects of (CRP, SAA/100, IL-6) are estimated by $(-0.324, 0.167, 1.195)$ with the bootstrap standard errors $(0.1564, 0.3185, 0.2398)$. In light of the extremely high canonical correlation 0.998, the estimated single-index and $Z_{\widehat{\beta}}$ are found to be closely matched. As mentioned in Sect. 5.1, the standard errors of their estimates are generally larger than ours. In Fig. 5d–f, the estimated survival functions, which are computed with the chosen bandwidth $\widetilde{\varsigma} = 0.690$, have a very similar shape and a slightly decreasing trend in the linear predictor. For the time-dependent AUC of the optimal composite plasma biomarker value Z_{β_0}, the proposed estimates at the selected days are all higher than those based on the time-varying logistic regression model (Fig. 6). With such a large sample size, both of $\bar{\mathcal{A}}(t; \widehat{\beta})$ and $\breve{\mathcal{A}}(t; \widehat{\beta})$ (not shown in the figure), which are computed with the chosen bandwidth $\widetilde{\varrho} = 0.849$, are almost the same.

6 Conclusion and discussion

In the context of this article, the word "versatility" mainly highlights the essential role and the diverse utility of $N(t; S, \beta_0)$ in estimation, classification, and model checking. Under a very general censoring mechanism, the features of this estimable counting process are fully considered in developing estimation procedures for the index coefficients, the induced conditional survival function, and the time-dependent AUC. For the proposed estimators, we further establish the related large sample properties and bandwidth selection procedures. It is noticeable that a chosen bandwidth for the conditional survival function estimator is an estimator for the asymptotically

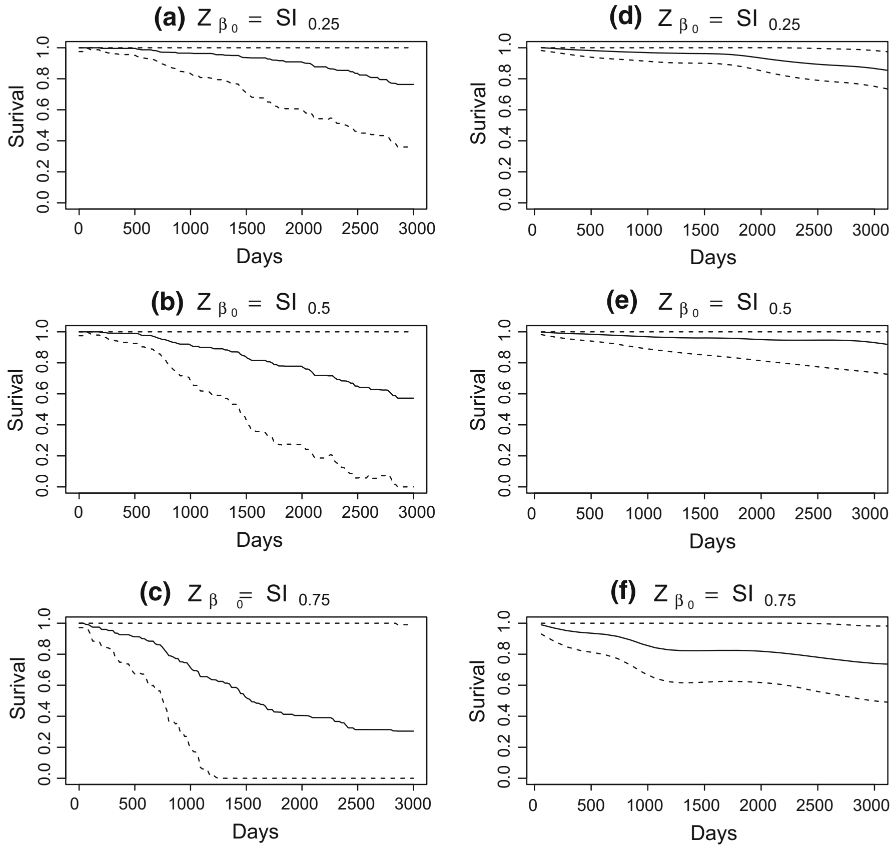

Fig. 5 The estimated conditional survival functions (*solid curves*) and the corresponding 95% bootstrap-based simultaneous confidence bands (*dashed curves*) for a primary biliary cirrhosis study (**a**–**c**) and an angiography cohort study (**d**–**f**)

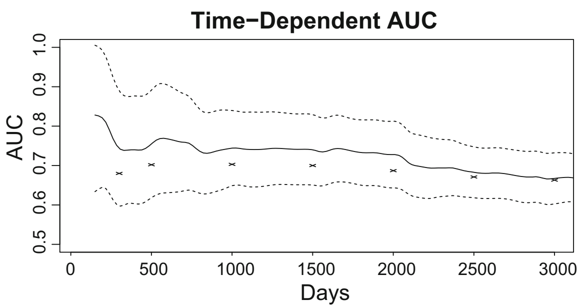

Fig. 6 The estimated AUC curve (*solid curve*) and the corresponding 95% bootstrap-based simultaneous confidence bands (*dashed curves*), and the estimated AUCs (+) based on the time-varying logistic regression at the selected days

optimal bandwidth selector but those involved bandwidths in the PILSE and the AUC estimators are only regarded as tuning parameters. In the numerical experiments, our estimators for the index coefficients and the time-dependent AUC are found to outperform their competitors (the PMLEs and the Chambless-Diao type estimator) and the presented conditional survival function estimator has a quite satisfactory performance.

As expected, an intrinsic capacity of $N(t; S, \beta_0)$ in discrimination should be a fairly good foundation for estimating the concordance index and other prospective and retrospective accuracy measures. Based on the defined pseudo-residual processes e_{it}'s in Sect. 5, we can directly apply the test rules of Chiang and Huang (2012) to check the validity of the SIH model. To deal with competing risks data, the hazards for failures of type j can be naturally formulated by generalizing model (1) as follows:

$$\lambda_{T_j}(t|z_{H_t}) = \lambda_j(t, z_{t\beta_{j0}}), \quad j = 1, \ldots, J, \tag{25}$$

where $\lambda_j(\cdot, \cdot)$'s are unknown bivariate functions and $Z_{t\beta_{j0}}$'s are type-specific linear predictors. One estimation strategy for the above model is to extend our estimation procedure with an appropriate modification. Another core theme emerged from the data analysis is to develop a more powerful test for the monotonicity of $\lambda(t, \upsilon)$ in υ. In longitudinal studies, time-dependent covariates might be intermittently collected at multiple follow-up times. It remains a challenge to either develop a new approach or a more flexible hazards model.

Acknowledgements The research of the first author was partially supported by the National Science Council Grants 99-2118-M-002-003- and 100-2118-M-002-005-MY2 (Taiwan). We would also like to thank the associate editor and a reviewer for some constructive comments on this paper.

References

Bouaziz, O., Lopez, O. (2010). Conditional density estimation in a censored single-index regression model. *Bermoulli*, *16*, 514–542.

Buckley, J., James, I. (1979). Linear regression with censored data. *Biometrika*, *66*, 429–436.

Chambless, L. E., Diao, G. (2006). Estimation of time-dependent area under the ROC curve for long-term risk prediction. *Statistics in Medicine*, *25*, 3474–3486.

Chiang, C. T., Huang, M. Y. (2012). New estimation and inference procedures for a single-index conditional distribution model. *Journal of Multivariate Analysis*, *111*, 271–285.

Chiang, C. T., Huang, S. Y. (2009). Estimation for the optimal combination of markers without modeling the censoring distribution. *Biometrics*, *65*, 152–158.

Chiang, C. T., Hung, H. (2010). Nonparametric estimation for time-dependent AUC. *Journal of Statistical Planning and Inference*, *140*, 1162–1174.

Chiang, C. T., Huang, M. Y., Wang, S. H. (2016). Bias and variance reduction in nonparametric estimation of time-dependent accuracy measures. *Statistics in Medicine*, *35*, 5247–5266.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society*, *B34*, 187–220.

Cox, D. R., Oakes, D. (1984). *Analysis of survival data*. London: Chapman and Hall.

Fleming, T. R., Harrington, D. P. (1991). *Counting processes and survival analysis*. New York: Wiley.

Gørgens, T. (2004). Average derivatives for hazard functions. *Econometric Theory*, *20*, 437–463.

Gørgens, T. (2006). Semiparametric estimation of single-index hazard functions without proportional hazards. *Journal of Econometrics*, *9*, 1–22.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, *21*, 157–178.

Heagerty, P. J., Lumley, T., Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and diagnostic marker. *Biometrics*, *56*, 337–344.

Huang, M. Y., Chiang, C. T. (2016). An effective semiparametric estimation approach for the sufficient dimension reduction model. *Journal of the American Statistical Association*. doi:10.1080/01621459

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation single-index models. *Journal of Econometrics*, *58*, 71–120.

Khan, S., Tamer, E. (2007). Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics*, *136*, 251–280.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. New York: Springer.

Lee, K. W. J., Hill, J. S., Walley, K. R., Frohlich, J. J. (2006). Relative value of multiple plasma biomarkers as risk factors for coronary artery disease and death in an angiography cohort. *Canadian Medical Association Journal*, *174*, 461–466.

Lopez, O., Patilea, V., Van Keilegom, I. (2013). Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, *19*, 721–747.

McIntosh, M. W., Pepe, M. S. (2002). Combining several screening tests: Optimality of the risk score. *Biometrics*, *58*, 657–664.

Strzalkowska-Kominiak, E., Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *Journal of Multivariate Analysis*, *114*, 74–98.

Strzalkowska-Kominiak, E., Cao, R. (2014). Beran-based approach for single-index models under censoring. *Computational Statistics*, *29*, 1243–1261.

Zeng, D., Lin, D. Y. (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association*, *102*, 1387–1396.