Ghosh *et al. Journal of Statistical Distributions and Applications* (2015) 2:2
DOI 10.1186/s40488-015-0026-8

Journal of Statistical Distributions and Applications
a SpringerOpen Journal

RESEARCH Open Access

Failure time regression with continuous informative auxiliary covariates

Lipika Ghosh[1], Jiancheng Jiang[1*], Yanqing Sun[1] and Haibo Zhou[2]

*Correspondence:
jjiang1@uncc.edu
[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
Full list of author information is available at the end of the article

Abstract

In this paper we use Cox's regression model to fit failure time data with continuous informative auxiliary variables in the presence of a validation subsample. We first estimate the induced relative risk function by kernel smoothing based on the validation subsample, and then improve the estimation by utilizing the information on the incomplete observations from non-validation subsample and the auxiliary observations from the primary sample. Asymptotic normality of the proposed estimator is derived. The proposed method allows one to robustly model the failure time data with an informative multivariate auxiliary covariate. Comparison of the proposed approach with several existing methods is made via simulations. Two real datasets are analyzed to illustrate the proposed method.

Mathematics Subject Classification (MSC): 62G07, 62G20

Keywords: Auxiliary covariates; Censoring; Estimated partial likelihood; Local linear smoothing; Validation

1 Introduction

In epidemiologic studies, the exposure variable vector X is often too difficult or too expensive to measure on the full cohort, whereas an auxiliary variable vector W for X can be easily measured for all subjects in the study cohort. For example, in a large scale nutritional study, the PIN Study (Savitz et al. 2001), it would be prohibitively expensive to obtain the exact dietary iron intake on each individual recruited. Instead, a self administered quantitative food questionnaire is conducted on all subjects where a crude assessment of iron intake is obtained. The true exposure, the blood serrum ferritin concentration, is only assayed for a validation set consisting of a small subset of the full study cohort. Although the true covariates are missing for most individuals, the existence of some surrogates or auxiliary measurements conveys information about X and serves as common proxy measure. Utilizing the available auxiliary information to improve the efficiency of the effects estimation and in turns to increase the power of the study is critical for the success of the studies. In this paper, we study censored failure time regression with a continuous auxiliary covariate vector.

A variety of authors have contributed their work to this field. Related works include Prentice (1982), Pepe et al. (1989), Lin and Ying (1993), Hughes (1993), Lipsitz and Ibrahim (1996), Zhou and Wang (2000), Fan and Wang (2009), Liu et al. (2010), etc. In particular, Prentice (1982) introduced a partial likelihood estimator based on the induced

relative risk function. This method was further developed by Pepe et al. (1989) using parametric modeling. Zhou and Pepe (1995) proposed an estimated partial likelihood method for discrete auxiliary covariates to relax the parametric assumptions on the frequency of events and the underlying distributions of covariates. This method was extended by Zhou and Wang (2000) to deal with continuous auxiliary variables, based on the Nadaraya-Watson kernel smoother method (Nadaraya 1964; Watson, 1964). Fan and Wang (2009), Liu et al. (2010) used the same approach for multivariate failure time data with auxiliary covariates. While Zhou and Wang's (2000) approach is useful in certain situations, there are some restrictions on it. First, the approach is effective only when the auxiliary variable W is of low dimension so that the "curse of dimensionality" in nonparametric smoothing can be avoided. Secondly, it requires that, conditionally on X, W provides no additional information about the hazard of failure; that is, all of the effects of W on failure and censoring are mediated through X, which is somewhat restricted since W may not be a true surrogate and depends on the failure given X.

Further, this method does not fully utilize the observations in the non-validation subsample and hence cannot be efficient in certain situations.

We here propose a new method to deal with the above problems associated with the method in Zhou and Wang (2000). The proposed method allows W to be multivariate and to be informative in the sense that, conditional X, it may provide additional information on the hazard of failure. We first estimate the induced relative risk function with a kernel smoother based on the validation sample, and then improve the estimation by utilizing the information on the incomplete observations from the non-validation subsample. In addition, the local linear smoother (see for example in Fan and Gijbels 1996) is employed to enhance the performance of the kernel smoother in Zhou and Wang (2000) at the boundary regions. Our method will be expected to improve the efficiency of the estimator of Zhou and Wang (2000) in various situations, for example, when auxiliary variable W is informative or not very informative about X (see also the simulation results). Asymptotic normality of our estimator is derived.

The proposed methodology can be extended to model multivariate failure time data with auxiliary covariates by following the method in Fan and Wang (2009) or Liu et al. (2010).

The paper is organized as follows. In Section 2, we introduce the hazards models. In Section 3 we introduce our new estimation approach to predicting the induced relative risk for individuals in non-validation subsample based on the kernel smoother. In Section 4 we concentrate on the asymptotic properties of the proposed estimators. We conduct simulations in Section 5 to compare the efficiencies of different estimating methods. In Section 6 we apply the proposed methodology to two real datasets.

2 Cox's proportional hazards models

To facilitate exposition, we here employ the notations in Zhou and Wang (2000). Suppose that there are n independent individuals in a study cohort. Let $\{X_i(t), Z_i(t)\}$ denote the covariate vector for the ith subject at time t ($i = 1, \cdots, n$). Assume that $X_i(\cdot)$ is only observed in the validation subsample which is chosen at the baseline under the ignorable missing mechanism condition (Rubin 1976). Let $Z_i(\cdot)$ be the remaining covariate vector that is always observed, and $W_i(\cdot)$ the informative auxiliary variables for $X_i(\cdot)$. Let η_i be an indicator variable with $\eta_i = 1$ if the ith individual is in the validation set and 0 if

in the nonvalidation set. Put $V = \{i : \eta_i = 1\}$ and $\bar{V} = \{i : \eta_i = 0\}$. We assume that individuals in the validation subsample are randomly selected and hence representative. Then observed data for the ith subject is $\{S_i, \delta_i, Z_i(\cdot), W_i(\cdot), X_i(\cdot)\}$ if $\eta_i = 1$, and $\{S_i, \delta_i, Z_i(\cdot), W_i(\cdot)\}$ if $\eta_i = 0$, where S_i is the observed event time for the ith subject, which is the minimum of the potential failure time T_i and the censoring time C_i, and δ_i is the indicator of censoring. We consider the following conditional hazard rate function of failure (Cox 1972)

$$\begin{aligned}\lambda\{t; X_i(t), Z_i(t)\} &\equiv \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t}\Pr\{t \le T_i < t + \Delta t | T_i \ge t, X_i(t), Z_i(t)\}\right] \\ &= \lambda_0(t)\exp\left\{\beta_1' X_i(t) + \beta_2' Z_i(t)\right\}, \end{aligned} \tag{2.1}$$

where $\lambda_0(\cdot) \ge 0$ is the unspecified base-line hazard and $\beta = \left(\beta_1', \beta_2'\right)'$ is the relative risk parameter vector to be estimated.

For model (2.1), the relative risk functions are $\gamma_i(\beta, t) \equiv \exp\left\{\beta_1' X_i(t) + \beta_2' Z_i(t)\right\}$, and the partial likelihood function for the parameters β is

$$PL(\beta) = \prod_{i \in V \cup \bar{V}} \left\{\frac{\gamma_i(\beta, S_i)}{\sum_{j \in \mathcal{R}(S_i)} \gamma_j(\beta, S_i)}\right\}^{\delta_i}, \tag{2.2}$$

where $\mathcal{R}(S_i)$ is the risk set at time S_i. However, for $i \in \bar{V}$, the true variate $X_i(t)$ is not observed, and hence the corresponding relative risk function $\gamma_i(\beta, t)$ is not available and has to be imputed.

Zhou and Wang (2000) used the conditional expectation

$$\exp\left\{\beta_2' Z_i(t)\right\} E\left[\exp\left\{\beta_1' X_i(t)\right\} | S_i \ge t, Z_i(t), W_i(t)\right] \tag{2.3}$$

for the imputation of $\gamma_i(\beta, t)$ $(i \in \bar{V})$. Based on data in V, they obtained the Nadaraya-Watson kernel estimator (Nadaraya 1964; Watson 1964) of the above imputation and replaced $\gamma_i(\beta, t)$ for $i \in \bar{V}$ in (2.2) by the kernel estimator, which leads to the estimated partial likelihood. Under the assumption that W is not informative, that is, all of the effects of W on failure and censoring are mediated through X, so that

$$\begin{aligned}\lambda\{t; X_i(t), Z_i(t), W_i(t)\} &\equiv \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t} Pr\{t \le T_i < t + \Delta t | T_i \ge t, X_i(t), Z_i(t), W_i(t)\}\right] \\ &= \lim_{\Delta t \downarrow 0}\left[\frac{1}{\Delta t} Pr\{t \le T_i < t + \Delta t | T_i \ge t, X_i(t), Z_i(t)\}\right] \\ &= \lambda_0(t)\exp\left\{\beta_1' X_i(t) + \beta_2' Z_i(t)\right\} \\ &\equiv \lambda\{t; X_i(t), Z_i(t)\},\end{aligned}$$

they derived the consistency and asymptotic normality of the estimation. However, if W is informative, their method will generally be biased (see also Section 5). In addition, since this method directly used information in the auxiliary covariate W and estimated the conditional expectation (2.3), it may encounter the so-called "curse of dimensionality" if W is of higher dimension. For the present study, we propose a new method for imputation of the relative risk function. The information in W will be used in a new way. This leads to a new estimated partial likelihood.

3 Estimated partial likelihood with a local smoother

In this section, we introduce our method to estimate the parameters in model (2.1) based on maximizing the estimated partial likelihood.

3.1 Local smoother for the relative risk function

Instead of (2.3), we use the conditional expectation of $\gamma_i(\beta, t)$,

$$\phi_i(\beta, t) = \exp\{\beta_2' Z_i(t) E\left[\exp\left\{\beta_1' X_i(t)\right\} | S_i \geq t, Z_i(t)\right],$$

as imputation of $\gamma_i(\beta, t)$. Let d be the dimension of Z and let

$$r_i(\beta, t) = \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\phi_i(\beta, t) \tag{3.4}$$

be the induced risk function. Put $\zeta_i(\beta_1, t) = \exp\left(\beta_1' X_i(t)\right)$, and $\nu_i(\beta_1, t) = E[\zeta_i(\beta_1, t)|S_i \geq t, Z_i(t)]$. To use the partial likelihood (2.2), we need to estimate $\phi_i(\beta, t)$ or equivalently $\nu_i(\beta_1, t)$ for $i \in \bar{V}$. Using the local linear smoother (see for example, Fan and Gijbels 1996) leads to the following (functional) estimators of $\nu_j(\beta_1, t)$ for $j \in \bar{V}$

$$\hat{\nu}_j(\beta_1, t) = \sum_{i \in V} \omega_i(t, Z_j(t); h)\zeta_i(\beta_1, t), \tag{3.5}$$

where h is the bandwidth,

$$\omega_i(t, Z_j(t); h) = \frac{\{s_2 - (Z_i(t) - Z_j(t))s_1\}Y_i(t)K_h(Z_i(t) - Z_j(t))}{\sum_{i \in V}\{s_2 - (Z_i(t) - Z_j(t))s_1\}Y_i(t)K_h(Z_i(t) - Z_j(t))},$$

$Y_i(t) = I_{[S_i \geq t]}$ is the at-risk indicator, $s_k = \sum_{i \in V}(Z_i(t) - Z_j(t))^k Y_i(t)K_h(Z_i(t) - Z_j(t))$, and $K_h(\cdot) = h^{-d}K(\cdot/h)$ for a d-variate kernel function $K(\cdot)$.

The above estimation of the relative risk function was similarly used in Zhou and Wang (2000) for a nonparametric smoothing problem on the estimation of $E[\gamma_i(\beta, t)|S_i \geq t, Z_i(t), W_i(t)]$, where the "curse of dimensionality" problem can happen if W is multivariate. Note that this estimation method uses only the complete observations in V and neglects the important information on incomplete observations in $\bar{V}$. It follows that this approach cannot be expected to be efficient in certain situations. In addition, it is required in Zhou and Wang (2000) that, conditional on X, the auxiliary variable W provides no additional information on the the hazard of failure. This requirement may not hold if W is not a genuine surrogate of X. In the following, we propose an improved estimation approach which utilizes information from W and observations in $\bar{V}$ and does not impose the requirement. Moreover, the proposed method allows one to model the failure time data with informative multivariate auxiliary variable W without "curse of dimensionality". Note that even for one dimensional Z and W, the method in Zhou and Wang (2000) requires a two-dimensional smoother while the new method needs only one-dimensional smoothing. To have a performance comparable with that of a one-dimensional nonparametric smoother using $M_1 = 50$ data points, we need about $M = M_1^{1.2} = 109$ data points for a 2-dimensional nonparameteric smoother. Hence the loss of efficiency due to highly dimensional smoothing is large and increasing exponentially fast (see page 317 of Fan and Yao 2003).

3.2 Improved estimation of the relative risk function and the estimated partial likelihood

Recall that W is a vector of auxiliary variables for X and is hence correlated with X. Let $\xi_i(\alpha,t) = \exp(\alpha' W_i(t))$, where α is a parameter vector to be chosen. Considering the conditional expectation of $\psi_i(\alpha,t) = E[\xi_i(\alpha,t)|S_i \geq t, Z_i(t)]$, then we can estimate $\psi_i(\alpha,t)$ by running local linear smoothing based on the data in V:

$$\hat{\psi}_j(\alpha,t) = \sum_{i\in V} \omega_i(t, Z_j(t); h)\xi_i(\alpha,t). \tag{3.6}$$

The following result depicts asymptotic correlation of $\hat{\nu}_j(\beta_1,t)$ and $\hat{\psi}_j(\alpha,t)$.

Proposition 3.1. *Suppose that the conditions in Appendix 1 hold. Given* $(S_j \geq t, Z_j(t))$, $\sqrt{nh^d}[\,(\hat{\nu}_j(\beta_1,t) - \nu_j(\beta_1,t)), (\hat{\psi}_j(\alpha,t) - \psi_j(\alpha,t))]$ *is jointly asymptotically normal with mean zero and covariance matrix*

$$\Sigma = \nu_0(K)p^{-1}(Z_j)\begin{bmatrix} \sigma_1^2(Z_j,t) & \rho_\alpha^*(Z_j,t)\sigma_1(Z_j,t)\sigma_2(Z_j,t) \\ \rho_\alpha^*(Z_j,t)\sigma_1(Z_j,t)\sigma_2(Z_j,t) & \sigma_2^2(Z_j,t) \end{bmatrix},$$

where $\nu_0(K) = \int K^2(u)du$, $\sigma_1^2(Z_j,t) = Var[\zeta_j|S_j \geq t, Z_j]$, $\sigma_2^2(Z_j,t) = Var[\xi_j|S_j \geq t, Z_j]$, *and* $p(\cdot)$ *is the density function of* Z.

By the distribution theory for multivariate normal variates, the conditional distribution of $\sqrt{nh^d}[\,\hat{\nu}_j(\beta_1,t) - \nu_j(\beta_1,t)]$ given $\sqrt{nh^d}[\,\hat{\psi}_j(\alpha,t) - \psi_j(\alpha,t)]$ is asymptotically normal with mean

$$\rho_\alpha^*(Z_j,t)\frac{\sigma_1(Z_j,t)}{\sigma_2(Z_j,t)}\sqrt{nh^d}[\,\hat{\psi}_j(\alpha,t) - \psi_j(\alpha,t)]\,.$$

The conditional mean can be estimated by substituting consistent estimators based on the validation sample for $\rho_\alpha^*(Z_j,t)$, $\sigma_1(Z_j,t)$ and $\sigma_2(Z_j,t)$, and by replacing $\psi_j(\alpha,t)$ with the primary sample based estimator

$$\bar{\psi}_j(\alpha,t) = \sum_{i\in V\cup\bar{V}} \bar{\omega}_i(t, Z_j(t); h)\xi_i(\alpha,t), \tag{3.7}$$

where

$$\bar{\omega}_i(t, Z_j(t); h) = \frac{\left\{\bar{s}_2 - (Z_i(t) - Z_j(t))\bar{s}_1\right\} Y_i(t)K_h(Z_i(t) - Z_j(t))}{\sum_{i\in V\cup\bar{V}}\left\{\bar{s}_2 - (Z_i(t) - Z_j(t))\bar{s}_1\right\} Y_i(t)K_h(Z_i(t) - Z_j(t))}$$

and $\bar{s}_k = \sum_{i\in V\cup\bar{V}}\left(Z_i(t) - Z_j(t)\right)^k Y_i(t)K_h\left(Z_i(t) - Z_j(t)\right)$.

By equating $\sqrt{n_v h^d}[\,\hat{\nu}_j(\beta_1,t) - \nu_j(\beta_1,t)]$ with its estimated conditional mean and solving for $\nu_j(\beta_1,t)$, we obtain an improved (functional) estimate

$$\bar{\nu}_j\,(\beta_1,t) = \hat{\nu}_j\,(\beta_1,t) - \hat{\rho}_\alpha^*\left(Z_j,t\right)\frac{\hat{\sigma}_1\left(Z_j,t\right)}{\hat{\sigma}_2(Z_j,t)}\left[\hat{\psi}_j\,(\alpha,t) - \bar{\psi}_j\,(\alpha,t)\right]. \tag{3.8}$$

The updated estimator $\bar{\nu}_j(\beta_1,t)$ is doomed to be more accurate than $\hat{\nu}_j(\beta_1,t)$ in (3.5), since it has used the information from W and observations in $\bar{V}$. Even though the information about W may not be utilized in a very efficient way as in Zhou and Wang's (2000) estimator when W is not informative, it is the price we have to pay for achieving robustness against informative W. Note that $\bar{\nu}_j$ depends on α which is related to efficiency of the estimator. Intuitively, one should choose α to maximize the conditional correlation coefficient between ζ_j and ξ_j, given $(S_j \geq t, Z_j)$, which is evident from the following result.

Proposition 3.2. *Assume that the conditions in Appendix 1 hold. Given* $(S_j \geq t, Z_j(t))$, *then we have*

$$\sqrt{nh^d}\left[\bar{\nu}_j(\beta_1, t) - \nu_j(\beta_1, t)\right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega),$$

where $\Omega(Z_j, t) = \sigma_1^2(Z_j, t)\left[1 - (1-\rho)\rho_\alpha^{*2}(Z_j, t)\right]\nu_0(K)p^{-1}(Z_j)$.

When $\rho_\alpha^* = 0$, i.e. the relative risk contributed by W is not correlated to that contributed by X, given $(S \geq t, Z)$, the estimator $\bar{\nu}_j$ is asymptotically equivalent to $\hat{\nu}_j$ in (3.5).

In general, $\rho_\alpha^* > 0$. Hence, by Propositions 3.1 and 3.2, $\bar{\nu}_j$ is more efficient than $\hat{\nu}_j$. Note that the proposed estimator is consistent for any α.

The above estimation method for $\nu_j(\beta, t)$ was similarly used in Chen and Chen (2000) for estimating parameters in a parametric regression model. Our estimation can be regarded as an extension of their estimation approach in nonparametric regression. In addition, we do not need a working model to specify the regression relationship between the surrogate and the covariate, and hence there is no risk of misspecification of the working model.

For each given value of β, with the estimator $\bar{\nu}_j(\beta, t)$, we can estimate the induced relative risk $r_i(\beta, t)$ in (3.4) by

$$\hat{r}_i(\beta, t) = \eta_i \gamma_i(\beta, t) + (1 - \eta_i)\bar{\phi}_i(\beta, t), \tag{3.9}$$

where $\bar{\phi}_i(\beta, t) = \bar{\nu}_i(\beta_1, t)\exp\left\{\beta_2' Z_i(t)\right\}$. Then the parameters β can be estimated by maximizing the following estimated partial likelihood function (EPL):

$$EPL(\beta) = \prod_{i=1}^{n}\left\{\frac{\hat{r}_i(\beta, S_i)}{\sum_{j\in\mathcal{R}(S_i)}\hat{r}_j(\beta, S_i)}\right\}^{\delta_i}. \tag{3.10}$$

We denote $\hat{\beta}_{EPL} = \arg\max_\beta EPL(\beta)$.

For an extreme case with $W \approx Z$, Zhou and Wang's imputation for (2.3) approximately becomes $\hat{\phi}_i(\beta, t) = \hat{\nu}_i(\beta, t)\exp(\beta' Z_i(t))$ and uses a two dimensional smoother, which is inferior to the improved estimator $\bar{\phi}_i(\beta, t)$, and hence by the definition of $\hat{\beta}_{EPL}$, our estimator is superior to Zhou and Wang's. However, it is generally difficult to compare these two estimators. In our estimation of the induced relative risk, we used an improved estimator $\bar{\phi}_j(\beta, t)$ for $j \in \bar{V}$. The "curse of dimensionality" problem in Zhou and Wang (2000) can be avoided for a multivariate W. Our approach would at least be useful in cases where the number of variables in Z which are correlated with the missing covariate X is low, whereas the exposure variables of interest and their auxiliary variables may be multivariate.

An alternative to $\hat{\beta}_{EPL}$ is to maximize (3.10) but with $\hat{r}_i(\beta, t)$ replaced by $\tilde{r}_i(\beta, t) = \eta_i\gamma_i(\beta, t) + (1-\eta_i)\hat{\phi}_i(\beta, t)$, where $\hat{\phi}_i(\beta, t) = \hat{\nu}_i(\beta, t)\exp(\beta' Z_i(t))$. We denote the resulting estimator by $\hat{\beta}_V$, which does not use the information on W in $\bar{V}$. Intuitively, $\hat{\beta}_{EPL}$ should be better than $\hat{\beta}_V$, but this is not true in general, since comparison of the asymptotic results in Theorems 4.1 and 4.2 below could not lead to a dominated estimator. However, in small validation ratio settings, $\hat{\beta}_V$ is not expected to perform well, since it uses only the observations in the validation set for smoothing.

4 Asymptotic behaviors

Let n_v be the subsample size of the validation set and let ρ be the limit of ratio of validation observations, $\lim_{n\to\infty} n_v/n$. Assume that $\rho \in (0,1]$. Define $s^{(0)}(\beta,t) = E\left[Y_i(t)r_i(\beta,t)\right]$, $s^{(1)}(\beta,t) = (\partial/\partial\beta)s^{(0)}(\beta,t)$, $s^{(2)}(\beta,t) = (\partial/\partial\beta^\tau)s^{(1)}(\beta,t)$.

For any matrix A, we use $A^{\otimes 2}$ to denote matrix AA^τ.

Let $N_i(t) = I_{[S_i<t,\delta_i=1]}$ and

$$M_i(t) = N_i(t) - \int_0^t Y_i(u)r_i(\beta_0,u)\lambda_0(u)du.$$

Define the filters $\mathcal{F}_i(t) = \sigma\left\{N_i(u), Y_i(u+), X_i(u), Z_i(u) : 0 \le u \le t\right\}$.

The censoring time is assumed to be independent of the failure time conditioning on the true covariates in model (2.1), that is,

$$P\{t \le T < t+\Delta t | T \ge t, C \ge t, X(t), Z(t)\} = P\{t \le T < t+\Delta t | T \ge t, X(t), Z(t)\}, \tag{4.11}$$

which is different from that in Zhou and Wang (2000) where it is assumed

$$P\{t \le T < t+\Delta t | T \ge t, C \ge t, W(t), Z(t)\} = P\{t \le T < t+\Delta t | T \ge t, W(t), Z(t)\}.$$

Then, under the independent censoring assumption (4.11),

$M_i(t)$ is a mean zero martingale with repsect to $\mathcal{F}_i(t)$ (Kalbfleisch and Prentice 1980; Fleming and Harrington 1991).

In addition, the cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(w)\,dw$ can be consistently estimated as

$$\hat{\Lambda}_0(t) = \int_0^t \left[\sum_{i=1}^{n_v} Y_i(u)r_i(\hat{\beta}_{EPL},u)\right]^{-1} \sum_{i=1}^{n_v} dN_i(u).$$

Without loss of generality, we assume that $t \in [0,1]$. Put $\Delta(\phi_i)(u) = \phi_i^{(1)}(u)/\phi_i(u) - s^{(1)}/s^{(0)}$, $\Delta(\gamma_i)(u) = \gamma_i^{(1)}(u)/\gamma_i(u) - s^{(1)}/s^{(0)}$, $Q_i = \int_0^1 \Delta(\phi_i)(u)Y_i(u)[\gamma_i(\beta_0,u) - \phi_i(\beta_0,u)]\,\lambda_0(u)du$, $Q_i^* = \int_0^1 \Delta(\phi_i)(u)Y_i(u)\theta_i(u;\alpha)\lambda_0(u)du$, where $\phi_1^{(1)}(\beta,u) = (\partial/\partial\beta)r_i(\beta,u)$, and

$$\theta_i(u;\alpha) = [\xi_i(\alpha,u) - \psi_i(\alpha,u)]\exp(\beta_2' Z_i(u))\rho_\alpha^*(Z_i,u)\sigma_1(Z_i,u)/\sigma_2(Z_i,u).$$

The following theorem shows that $\hat{\beta}_{EPL}$ is asymptotically normal.

Theorem 4.1. *Suppose that Condition (A) in Appendix 1 holds. Then $\hat{\beta}_{EPL}$ is consistent estimator for β and satisfies*

$$\sqrt{n}(\hat{\beta}_{EPL} - \beta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Omega),$$

where $\Omega = I^{-1}(\beta_0)\Sigma(\beta_0)I^{-1}(\beta_0)$ with $\Sigma(\beta_0) = (1-\rho)\Sigma_1(\beta_0) + \rho\Sigma_2(\beta_0)$,

$$I(\beta_0) = -E\left[\int_0^1 \left(\frac{r_i^{(2)}(\beta_0,u)}{r_i^{(0)}(\beta_0,u)} - \left\{\frac{r_i^{(1)}(\beta_0,u)}{r_i^{(0)}(\beta_0,u)}\right\}^{\otimes 2} - \frac{s^{(2)}(\beta_0,u)}{s^{(0)}(\beta_0,u)} - \left\{\frac{s^{(1)}(\beta_0,u)}{s^{(0)}(\beta_0,u)}\right\}^{\otimes 2}\right) dN_i(t)\right],$$

$$\Sigma_1(\beta_0) = E\left[\int_0^1 \Delta(\phi_i)(u)dM_i(u) - (1-\rho)Q_i^*\right]^{\otimes 2},$$

$$\Sigma_2(\beta_0) = E\left[\int_0^1 \Delta(\gamma_i)(u)dM_i(u) - \frac{1-\rho}{\rho}\left\{Q_i - (1-\rho)Q_i^*\right\}\right]^{\otimes 2}.$$

Remark 4.1. It is interesting to note that $Q_i^* \approx Q_i$ when the auxiliary W approximates X, and hence the second term in the expectation of $\Sigma_2(\beta)$ approximates to $(1-\rho)Q_i$. Therefore, a small ρ will not result in a big $\Sigma_2(\beta_0)$. Theoretically, when $W_i = Z_i$, $Q_i^* = 0$ and the above asymptotic variance formula shares the same formula as that for the estimator in Zhou and Wang (2000) as exactly expected. However, in practice where $W_i \approx Z_i$, since Zhou and Wang (2000) used a higher dimensional smoother than us, our estimator would have better efficiency for finite samples.

When $\rho_\alpha^* = 0$, $\hat{\beta}_{EPL}$ is asymptotically equivalent to the complete-case estimator based on only the validation set V. This is also expected, since the auxiliary variable W_i contains no information on X_i at this setting. From Theorem 4.1, the asymptotic covariance matrix of $\hat{\beta}_{EPL}$ is of sandwich form, which can consistently be estimated by $\hat{\Omega}_0 = \hat{I}^{-1}(\beta_0)\hat{\Sigma}(\beta_0)\hat{I}^{-1}(\beta_0)$, where $\hat{I}(\beta)$ and $\hat{\Sigma}(\beta)$ are the corresponding empirical estimates. Specifically,

$$\hat{I}(\beta) = -n^{-1}\sum_{i=1}^{n}\int_0^1 \left(\frac{\hat{r}_i^{(2)}(\beta,u)}{\hat{r}_i^{(0)}(\beta,u)} - \left\{\frac{\hat{r}_i^{(1)}(\beta,u)}{r_i^{(0)}(\beta,u)}\right\}^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,u)}{\hat{S}^{(0)}(\beta,u)} + \left\{\frac{\hat{S}^{(1)}(\beta,u)}{\hat{S}^{(0)}(\beta,u)}\right\}^{\otimes 2}\right) dN_i(t),$$

$$\hat{\Sigma}_1(\beta) = (n-n_v)^{-1}\sum_{i\in\bar{V}}\left\{\int_0^1 \Delta(\hat{\phi}_i)(t)\left[dN_i(t) - Y_i(t)\bar{\phi}_i(\hat{\beta}_{EPL},t)\,d\hat{\Lambda}_0(t)\right] - (1-\hat{\rho})\hat{Q}_i^*\right\}^{\otimes 2},$$

$$\hat{\Sigma}_2(\beta) = n_v^{-1}\sum_{i\in V}\left\{\int_0^1 \Delta(\hat{\gamma}_i)(t)\left[dN_i(t) - Y_i(t)r_i(\hat{\beta}_{EPL},t)\,d\hat{\Lambda}_0(t)\right] - \frac{1-\hat{\rho}}{\hat{\rho}}\left[\hat{Q}_i - (1-\hat{\rho})\hat{Q}_i^*\right]\right\}^{\otimes 2},$$

where $\hat{\rho} = n_v/n$,

$$\hat{Q}_i = \int_0^1 \Delta(\hat{\phi}_i)(t)Y_i(t)\left[\hat{r}_i(\beta,t) - \hat{\phi}_i(\beta,t)\right] d\hat{\Lambda}_0(t),$$

$$\hat{Q}_i^* = \int_0^1 \Delta(\hat{\phi}_i)(t)Y_i(t)\hat{\theta}_i(t;\alpha)\,d\hat{\Lambda}_0(t),$$

$$\Delta(\hat{\phi}_i)(t) = \hat{\phi}_i^{(1)}(\beta,t)/\hat{\phi}_i(\beta,t) - \hat{S}^{(1)}(\beta,t)/\hat{S}^{(0)}(\beta,t),$$

$$\Delta(\hat{\gamma}_i)(t) = \hat{\gamma}_i^{(1)}(\beta,t)/\hat{\gamma}_i(\beta,t) - \hat{S}^{(1)}(\beta,t)/\hat{S}^{(0)}(\beta,t),$$

$$\hat{\theta}_i(t;\alpha) = \left[\xi_i(\alpha,t) - \bar{\psi}_i(\alpha,t)\right]\exp\left(\hat{\beta}_2^\tau Z_i(t)\right)\hat{\rho}_\alpha^*(Z_i,t)\hat{\sigma}_1(Z_i,u)/\hat{\sigma}_2(Z_i,t).$$

In summary, a constant variance estimator for $\hat{\beta}_{EPL}$ can be obtained by replacing the population quantities in the asymptotic covariance matrix $\Sigma(\beta_0)$ with their corresponding sample averages as in Zhou and Wang (2000). Hence, the asymptotic confidence intervals for β can also be constructed.

The following theorem demonstrates the asymptotic normality of $\hat{\beta}_V$.

Theorem 4.2. *Under the same conditions as in Theorem 4.1, the estimator* $\hat{\beta}_V$ *shares the same asymptotic distribution as* $\hat{\beta}_{EPL}$ *but with* $\Sigma_1(\beta_0)$ *and* $\Sigma_2(\beta_0)$ *replaced by* $\Sigma_{1V}(\beta_0)$ *and* $\Sigma_{2V}(\beta_0)$*, respectively, where* $\Sigma_{1V}(\beta_0) = E\left[\int_0^1 \Delta(\phi_i)(u)dM_i(u)\right]^{\otimes 2}$*, and* $\Sigma_{2V}(\beta_0) = E\left[\int_0^1 \Delta(\gamma_i)(u)dM_i(u) - \frac{1-\rho}{\rho}Q_i\right]^{\otimes 2}$.

4.1 Choice of the parameter vector α

The choice of α affects efficiency of $\hat{\beta}_{EPL}$, although the estimator is $\sqrt{n}$-consistent for any α. In this paper, we choose α by minimizing the variance of the estimator $\hat{\beta}_{EPL}$. Given initial value of β and α, one can estimate α by minimizing the trace of $\hat{\Omega}(\alpha)$.

Once the value of α is known, maximization of $EPL(\beta)$ can be solved via Newton-Raphson iterations. Repeating this procedure, one can find a solution to the optimization problem (3.10). To reduce the burden of computation in practice, one can employ a consistent naive estimator of β as initial value, for example the estimator of β based on only the validation sample which is easy to implement because it involves only a simple fit for the usual Cox's model. In our experience, using the naive estimator as the initial value the iterations converge in a few steps.

4.2 Choice of the bandwidth parameter

As for the bandwidths, they affect the estimator $\hat{\beta}_{EPL}$, which is true in any nonparametric smoothing problems. Fortunately, the proposed estimator $\hat{\beta}_{EPL}$ is effective for a large range of bandwidths (see Condition (6) in Appendix 1). Similar to that in Zhou and Wang (2000), we employed here the empirical bandwidth $h = (h_1, h_2)'$ with $h_1 = 2\hat{\sigma}_Z n^{-1/3}$ and $h_2 = 2\hat{\sigma}_W n^{-1/3}$, where $\hat{\sigma}_Z$ and $\hat{\sigma}_W$ are respectively the sample standard deviations of Z and W, which satisfy the bandwidth conditions required in this paper.

5 Simulations

In this section, we conduct finite-sample simulations[a] The aims of the simulations are three-fold: one is to examine the small sample behavior of $\hat{\beta}_{EPL}$, another is to compare the performance of our estimator with some existing estimators under various situations, and the third and the most important is to illustrate that the proposed estimation allows for an informative auxiliary vector W. The covariates (X, Z) are generated from the following transformation to create correlation:

$$\begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 0.0 \\ 0.5 & 1 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}, \tag{5.12}$$

where U_i's are independent and identically distributed as $U(0, 2)$. The failure time T conditional on covariate X is from an exponential distribution with hazard function

$$\lambda(t; X) = \lambda \exp(\beta_1 X + \beta_2 Z),$$

where λ is the baseline constant hazard. We only consider the case $\lambda = 1$. Then

$$f(t; X, Z) = \exp(\beta_1 X + \beta_2 Z) \exp(-t \exp(\beta_1 X + \beta_2 Z)).$$

The auxiliary variable W is generated from

$$W = X + \gamma \log(T) + e, \tag{5.13}$$

where $e \sim \mathcal{N}(0, \sigma^2)$ and σ^2 is the parameter controlling the strength of the association between X and W. We consider the settings with $\gamma = 0$ and 2. Model (5.13) with $\gamma = 2$ allows one to explore the effectiveness of the proposed method with an informative surrogate W. For $\gamma = 0$, it also allows us to compare the performance of the newly proposed method and that in Zhou and Wang (2000). We do simulations for $\sigma = 0.2$ and 0.8. The censoring variable is uniformly distributed and independent of the failure time. The validation set is randomly selected with $P(\eta_i = 1) = 0.5$.

We choose the Gaussian kernel function with the bandwidths $\left(h_1 = 2\hat{\sigma}_Z n^{-1/3}, h_2 = 2\hat{\sigma}_W n^{-1/3}\right)$ which satisfy the bandwidth conditions in Theorem 4.1, where $\hat{\sigma}_Z$ and $\hat{\sigma}_W$ are the sample standard deviations of Z and W respectively. In the following tables, $\beta_0 = [\log(2), 0.5]'$ denotes the true value of the parameter to be estimated, *se* is the standard error of $\hat{\beta}_{EPL}$ from simulation, *mean*($\hat{se}$) denotes the mean of the estimated standard errors and *cp* denotes the 95% coverage probability.

The methods we considered are the newly proposed estimated partial likehood estimation ($\hat{\beta}_{EPL}$)and its conterpart ($\hat{\beta}_{ZW}$) in Zhou and Wang (2000), the estimator ($\hat{\beta}_V$) which does not use the information on W in $\bar{V}$, the complete-case Cox regression analysis ($\hat{\beta}_{CC}$) which uses only the validation subsample, the Cox regression with W substituted for the missing X ($\hat{\beta}_N$), and the full data Cox regression ($\hat{\beta}_F$) which assumes that X is available for all n subjects in the study.

Tables 1 and 2 summarizes the results obtained from the simulation. Note that $\hat{\beta}_F$, $\hat{\beta}_{CC}$ and $\hat{\beta}_V$ in Table 2 are the same in Table 1, so we do not report them in Table 2. For finite sample sizes, the mean, median, standard error (*se*) and 95% confidence intervals of the estimator are calculated based on 500 independent runs. We observe that $\hat{\beta}_F$ is the best estimator but it is not always obtainable for practical studies. The estimator $\hat{\beta}_N$ is biased. In all the situations $\hat{\beta}_{EPL}$ is observed to be a consistent estimator of true β_0. The estimates obtained from Zhou and Wang (2000) method are biased when $\gamma = 2$. Also,we notice that the bias in their estimates increases when σ increases. Efficiency of $\hat{\beta}_{EPL}$ relative to the complete case estimator $\hat{\beta}_{CC}$ is approximately the same for $\beta_1 = \log(2)$ but much higher for $\beta_2 = 0.5$. For $\gamma = 0$ the estimator $\hat{\beta}_{ZW}$ is more efficient than the proposed estimator for smaller values of σ but as the correlation between the exposure and auxiliary variable decreases the efficiency becomes closer (see the values of *se* for $n = 300$). Also, we notice that $\hat{\beta}_{EPL}$ has less bias than $\hat{\beta}_V$ for different values of n, but they are still comparable and even in some cases $\hat{\beta}_V$ is better in terms of the standard deviation. For $\gamma = 2$, our method stays almost equally efficient as σ increases, but $\hat{\beta}_{ZW}$ fails because of its large bias and low coverage probability (*cp*). Note that, when $\gamma = 2$, W is an informative auxiliary variable about the failure time and is not very informative about X.

We also performed simulations to see the effect of validation ratio and different bandwidths on the estimation. The proposed estimator $\hat{\beta}_{EPL}$ works well for smaller validation percentages and is not very sensitive to the bandwidth selection. In particular, $\hat{\beta}_{EPL}$ is better than $\hat{\beta}_V$ when the validation ratio is 0.25, which is evidenced in Table 3. We also conducted simulations with smaller validation ratios. Our experience indicates that, as the validation ratio gets as small as 0.2, $\hat{\beta}_V$ is very bad but $\hat{\beta}_{EPL}$ still works.

We conclude that, the proposed partial likelihood estimator can be used to make inference for β under various situations. In particular, the estimator is consistent and efficient when the auxiliary variable is informative about the hazard rate of failure time while Zhou and Wang's estimator fails.

Table 1 Comparison of simulation results with $\sigma = 0.2$ and validation fraction 0.5

n		$\gamma = 0$						$\gamma = 2$		
		$\hat{\beta}_F$	$\hat{\beta}_{CC}$	$\hat{\beta}_N$	$\hat{\beta}_V$	$\hat{\beta}_{ZW}$	$\hat{\beta}_{EPL}$	$\hat{\beta}_N$	$\hat{\beta}_{ZW}$	$\hat{\beta}_{EPL}$
50% *censoring*										
100	*mean* $-\beta_0$	0.018	0.021	-0.034	-0.036	-0.045	0.006	-0.988	-0.434	0.031
		0.004	0.030	0.022	0.035	0.013	0.006	0.169	0.174	0.010
	median $-\beta_0$	0.013	-0.001	-0.045	-0.047	0.033	-0.033	-0.983	-0.484	0.021
		-0.015	0.023	0.002	0.019	-0.001	0.011	0.160	0.172	0.008
	se	0.323	0.439	0.302	0.410	0.346	0.457	0.084	0.574	0.459
		0.281	0.391	0.276	0.307	0.283	0.312	0.237	0.324	0.338
	mean($\hat{se}$)	0.292	0.429	0.280	0.402	0.329	0.414	0.068	0.373	0.411
		0.258	0.382	0.257	0.278	0.257	0.296	0.244	0.274	0.288
	cp	0.916	0.956	0.924	0.938	0.948	0.946	0.0	0.670	0.946
		0.930	0.960	0.934	0.922	0.914	0.924	0.920	0.892	0.916
300	*mean* $-\beta_0$	-0.019	0.026	-0.068	-0.039	-0.007	-0.021	-0.994	-0.630	0.005
		0.006	0.017	0.024	0.013	0.010	0.002	0.166	0.202	-0.007
	median $-\beta_0$	-0.028	0.024	0.077	-0.042	-0.012	-0.018	-0.991	-0.635	-0.002
		0.007	0.001	0.028	0.021	0.018	0.002	0.169	0.199	0.003
	se	0.161	0.234	0.158	0.225	0.176	0.238	0.048	0.246	0.243
		0.146	0.217	0.146	0.162	0.150	0.163	0.127	0.137	0.166
	mean($\hat{se}$)	0.164	0.233	0.158	0.227	0.177	0.222	0.039	0.170	0.231
		0.146	0.209	0.145	0.159	0.147	0.155	0.137	0.125	0.159
	cp	0.944	0.942	0.928	0.948	0.950	0.936	0.0	0.108	0.940
		0.956	0.942	0.956	0.948	0.944	0.938	0.796	0.630	0.940
20% *censoring*										
100	*mean* $-\beta_0$	0.021	0.020	-0.031	-0.041	0.044	0.002	-1.003	-0.455	0.023
		0.001	0.014	0.018	0.036	0.013	0.005	0.155	0.163	0.011
	median $-\beta_0$	0.016	0.014	-0.029	-0.048	0.038	-0.013	-1.000	-0.466	0.003
		-0.008	-0.001	0.011	0.029	0.005	0.003	0.151	0.159	0.005
	se	0.248	0.339	0.234	0.322	0.272	0.364	0.071	0.467	0.360
		0.211	0.305	0.210	0.224	0.212	0.229	0.180	0.241	0.235
	mean($\hat{se}$)	0.232	0.340	0.223	0.306	0.263	0.313	0.062	0.318	0.315
		0.205	0.302	0.204	0.217	0.204	0.213	0.195	0.214	0.215
	cp	0.936	0.956	0.934	0.912	0.966	0.894	0.0	0.550	0.904
		0.938	0.956	0.938	0.934	0.952	0.936	0.924	0.862	0.928
300	*mean* $-\beta_0$	-0.007	-0.009	-0.056	-0.032	-0.006	-0.011	-1.001	-0.617	0.015
		-0.001	0.008	0.016	0.008	0.004	-0.004	0.152	0.023	-0.012
	median $-\beta_0$	-0.019	-0.016	-0.064	-0.040	-0.006	-0.022	-1.001	-0.613	0.001
		-0.002	0.002	0.011	0.007	0.003	-0.005	0.154	0.024	-0.016
	se	0.131	0.190	0.127	0.177	0.141	0.194	0.044	0.304	0.195
		0.116	0.164	0.116	0.126	0.119	0.127	0.100	0.178	0.135
	mean($\hat{se}$)	0.131	0.187	0.260	0.179	0.142	0.179	0.034	0.219	0.179
		0.116	0.166	0.150	0.125	0.117	0.122	0.110	0.161	0.124
	cp	0.948	0.960	0.928	0.952	0.966	0.926	0.0	0.244	0.926
		0.952	0.954	0.952	0.954	0.952	0.944	0.748	0.070	0.934

6 Real data analysis

6.1 Primary Biliary Cirrhosis data

We apply the proposed approach to the data from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. A total of 424 PBC patients, referred to Mayo Clinic during that ten-year interval, met eligibility criteria for

Table 2 Comparison of simulation results with $\sigma = 0.8$ and validation fraction 0.5

n		$\gamma = 0$			$\gamma = 2$		
		$\hat{\beta}_N$	$\hat{\beta}_{ZW}$	$\hat{\beta}_{EPL}$	$\hat{\beta}_{ZW}$	$\hat{\beta}_{ZW}$	$\hat{\beta}_{EPL}$
50% *censoring*							
100	$mean - \beta_0$	-0.369	-0.045	0.034	-0.961	-0.325	0.027
		0.139	0.055	0.008	0.177	0.141	0.009
	$median - \beta_0$	-0.381	-0.052	0.021	-0.955	-0.368	0.010
		0.028	0.051	0.008	0.164	0.140	0.002
	se	0.206	0.374	0.457	0.076	0.549	0.450
		0.260	0.287	0.338	0.243	0.325	0.332
	$mean(\hat{se})$	0.196	0.256	0.414	0.064	0.374	0.414
		0.249	0.262	0.288	0.244	0.271	.293
	cp	0.504	0.940	0.934	0.0	0.721	0.940
		0.914	0.932	0.916	0.904	0.888	0.920
300	$mean - \beta_0$	-0.392	-0.056	0.012	-0.965	-0.399	0.012
		0.139	0.033	-0.011	0.175	0.147	-0.009
	$median - \beta_0$	-0.392	-0.055	0.004	-0.963	-0.395	0.004
		0.139	0.044	-0.004	0.176	0.145	-0.002
	se	0.114	0.198	0.255	0.044	0.325	0.254
		0.142	0.156	0.170	0.129	0.180	0.171
	$mean(\hat{se})$	0.108	0.223	0.227	0.036	0.213	0.228
		0.140	0.157	0.159	0.137	0.157	0.158
	cp	0.068	0.932	0.932	0.0	0.520	0.932
		0.830	0.946	0.934	0.770	0.808	0.936
20% *censoring*							
100	$mean - \beta_0$	-0.368	-0.046	0.024	-0.969	-0.328	0.022
		0.126	0.052	0.019	0.163	0.128	0.021
	$median - \beta_0$	-0.372	-0.052	0.020	-0.966	-0.354	0.016
		0.122	0.053	0.020	0.168	0.124	0.015
	se	0.165	0.272	0.360	0.064	0.450	0.348
		0.202	0.213	0.240	0.186	0.238	0.238
	$mean(\hat{se})$	0.156	0.263	0.306	0.057	0.291	0.321
		0.198	0.207	0.211	0.195	0.212	0.222
	cp	0.352	0.966	0.912	0.0	0.658	0.916
		0.918	0.942	0.928	0.910	0.878	0.924
300	$mean - \beta_0$	-0.390	-0.048	0.013	-0.972	-0.388	0.021
		0.123	0.026	-0.011	0.161	0.127	-0.015
	$median - \beta_0$	-0.393	-0.058	-0.004	-0.972	-0.399	0.013
		0.123	0.028	-0.017	0.164	0.128	-0.016
	se	0.092	0.159	0.195	0.039	0.262	0.200
		0.115	0.123	0.131	0.102	0.138	0.133
	$mean(\hat{se})$	0.086	0.172	0.186	0.033	0.168	0.179
		0.112	0.123	0.126	0.110	0.122	0.124
	cp	0.018	0.954	0.924	0.0	0.412	0.924
		0.792	0.944	0.934	0.716	0.784	0.942

the randomized placebo controlled trial of the drug D-penicillamine. The first 312 cases in the data set participated in the randomized trial and contain largely complete data. The additional 112 cases did not participate in the clinical trial, but agreed to have basic measurements recorded and to be followed for survival. Six of those cases were lost to

Table 3 Comparison of simulation results with $\beta = [ln(2)\ \ 0.5]'$, 50% censoring, $\sigma = 0.2$, and validation fraction 0.25

	$n = 100$		$n = 200$	
	$\hat{\beta}_V$	$\hat{\beta}_{EPL}$	$\hat{\beta}_V$	$\hat{\beta}_{EPL}$
$mean - \beta_0$	-0.142	0.056	-0.087	0.049
	0.107	0.018	0.056	0.001
$median - \beta_0$	-0.152	0.035	-0.125	0.011
	0.091	-0.002	0.065	0.002
se	0.506	0.565	0.405	0.417
	0.329	0.320	0.234	0.232
$mean(\hat{se})$	0.513	0.618	0.380	0.410
	0.306	0.333	0.220	0.224
cp	0.944	0.936	0.928	0.924
	0.934	0.954	0.934	0.942

follow-up shortly after diagnosis, so the data here are on an additional 106 cases as well as the 312 randomized participants.

A clinical background description and a more extended discussion for the trial and the covariates recorded can be found in Dickson et al. (1989) and Markus et al. (1989). The variables involved in our specify analysis include id: case number; days: number of days between registration and the earlier of death, transplantation, or study analysis time; status: status of censoring; bili: serum bilirubin (in mg/dl); chol: serum cholesterol (inmg/dl) and Age: age in days.

In this analysis, we are particularly interested in the effect of patients' serum cholesterol and age on the survival of the patients. This type of failure time data can be modeled by the Cox proportional hazards models with an unknown baseline hazard function. However, about 31% outcomes of cholesterol were missing in this data set. Removing those observations may lead to biased estimates and standard errors. We noted that the outcomes of serum bilirubin were completely obtained with no missing values. Preliminary analysis showed that there is a significant correlation between serum cholesterol and bilirubin. Also, intuitively bilirubin has some additional effect on the hazard of failure and we would like to use that information efficiently. To illustrate this effect, we performed a complete Cox regression analysis for two different situations. We take the logarithmic transformation of bilirubin for our study.

In Table 4, we observe that the coefficient and standard error estimates are quite different for both the situations and the 95% confidence intervals for the coefficient of age are nonoverlapping. We can conclude that serum bilirubin has some additional effect on the hazard of failure. Hence, our proposed method can be applied to this dataset considering serum bilirubin as the informative auxiliary covariate.

Table 4 Regression analysis of primary biliary cirrhosis (PBC) data study

	Method	Variable	Parameter	Standard error	95% Confidence interval
$logbili < 1.6$	CC	logchol	0.271	0.393	(-0.499, 1.040)
		age	0.055	0.012	(0.031, 0.079)
$logbili \geq 1.6$	ZW	logchol	-0.635	0.345	(-1.312, 0.042)
		age	-0.005	0.016	(-0.037, 0.027)

Table 5 displays the analysis results based on the Cox's regression for the complete data (CC), the method proposed by Zhou and Wang (2000) (ZW) and the proposed method (EPL). The CC method uses only 284 complete-case observations and the other two methods use all 418 observations. Variable "logchol" denotes the logarithm of cholesterol. The estimates of the coefficients and their standard errors are given in the table.

The regression analysis confirms that both serum cholesterol and age are significantly related to the time to event. For estimating the effect of serum cholesterol and age, there is a reasonable efficiency gain by using the two methods based on partial likelihood approach over the complete case Cox regression analysis. But there is a discrepancy between the estimates from complete data and Zhou and Wangs estimate which could be due to the fact that the latter method does not consider the additional effect contributed by the auxiliary covariate. In our simulation we observed that the standard error of the estimates were underestimated in Zhou and Wangs method when auxiliary variable was informative. In the real data analysis also the standard error estimate for serum cholesterol is underestimated. Moreover, the standard error estimates in our method is comparable to Zhou and Wangs method whereas the calculation time is much less compared to their method.

6.2 Serrum Ferritin Concentration in relation to preterm delivery study

We apply the proposed approach to the data on iron intake in relation to preterm delivery study from the University of North Carolina Hospitals at Chapel Hill. A total of 1520 women were included in the study. 17 of these women were lost to follow up. So the data consist of 1503 individuals among which 270 individuals had their serrum ferritin concentration (*FERRITIN*) measured with an immunometric assay. However a crude score for dietary iron intake (*DTFE*) was collected using a dietary food frequency questionnaire for all the individuals.

A clinical background description and a more extended discussion for the trial and the covariates recorded can be found in Savitz et al. (2001). The variables involved in our specfic analysis include (i) id: case number; (ii) Gestation Time: The number of weeks from pregnancy to delivery; (iii) *DTFE*: Dietary iron intake(in 100mg/dl); (iv) *Ferritin*: Serum Ferritin (in 100 mg/dl); and (v) *Age*: age in years. By using the notations in the proposed method, X is *Ferritin*, W is *DTFE*, and Z is *Age*.

In this analysis, we are particularly interested in the effect of patients' serum ferritin and age on the delivery of the patients. This type of failure time data can be modeled by the Cox proportional hazards models.

Table 5 Regression analysis of primary biliary cirrhosis (PBC) data

Method	Variable	Estimates of parameters	Standard error	95% Confidence interval
CC	logchol	0.853	0.214	(0.432, 1.273)
	age	0.048	0.010	(0.029, 0.067)
ZW	logchol	1.142	0.154	(0.840, 1.444)
	age	0.047	0.007	(0.033, 0.061)
EPL	logchol	0.851	0.215	(0.429, 1.273)
	age	0.044	0.007	(0.029, 0.058)

However, outcomes for serum ferritin were missing in this data set. Removing those observations can lead to biased or inefficient estimates. We noted that the outcomes of dietary iron intake were completely obtained with no missing values.

Table 6 displays the analysis results based on the the CC method, the ZW method, and the proposed EPL method. The CC method used only 270 complete-case observations and the other two methods used all 1503 observations.

The regression analysis using the new method confirms that both serum ferritin and age are significantly related to the time to event. For estimating the effect of serum ferritin and age, there is also a reasonable efficiency gain by using the two methods based on partial likelihood approach over the complete case cox regression analysis. The estimate of serrum ferritin is lower by the EPL method. The estimate is significantly different from zero with p-value 0.020. In contrast, the p-value from CC method in estimation of serrum ferritin is 0.06.

7 Conclusion

We have introduced an EPL estimation method for Cox's models with informative auxiliary covariates and established asymptotic normality of our estimator. The proposed proposed methodology allows for multivariate auxiliary covariates W without suffering the curse of dimensionality.

We used the same bandwidth as suggested by Zhou and Wang (2000) in our estimation. Though it performs reasonably well, one can develop a bandwidth selection criteria like generalized cross-validation for an improved estimation. It is desirable to increase the efficiency of the estimation. In future, we can consider the optimization of α or introduce some weight structure in the score equation to achieve robustness. Further, it is worthy extending our approach to model multivariate failure time.

Endnote

[a] All numerical results in this paper are obtained using the software MATLAB and the codes are available (Additional file 1).

Appendix 1: Condition (A)

For the risk function $r_i(\beta,t)$ (as well as for $\hat{r}_i(\beta,t)$, $\gamma_i(\beta,t)$, $\hat{\phi}_i(\beta,t)$ and $\phi_i(\beta,t)$), we denote by $r_i^{(j)}(\beta,t)$ the jth derivative of $r_i(\beta,t)$ with respect to β, $j = 0,1,2$, where $r_i^{(0)}$ means the function itself. Define $S^{(0)}(\beta,t) = n^{-1}\sum_{i=1}^{n} Y_i(t)r_i(\beta,t), S^{(1)}(\beta,t) = (\partial/\partial\beta)S^{(0)}(\beta,t), S^{(2)}(\beta,t) = (\partial/\partial\beta)S^{(1)}(\beta,t), s^{(0)}(\beta,t) = E[Y_i(t)r_i(\beta,t)], s^{(1)}(\beta,t) = (\partial/\partial\beta)s^{(0)}(\beta,t), s^{(2)}(\beta,t) = (\partial/\partial\beta)s^{(1)}(\beta,t)$. Let $\Delta(\phi)(u) = \frac{\phi^{(1)}(\beta,u)}{\phi(\beta,u)} - \frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}$,

Table 6 Regression analysis of Iron intake in relation to preterm delivery study

Method	Variables	Estimates of parameters	Standard error	*p*-value	Hazard ratio
CC	ferritin	0.2451	0.1306	0.060	1.278
	age	0.009	0.0108	0.402	1.009
ZW	ferritin	0.2236	0.076	0.004	1.251
	age	0.0102	0.0043	0.018	1.010
EPL	ferritin	0.1797	0.0771	0.020	1.197
	age	0.0159	0.0036	0.000	1.016

$$I(\beta) = -\int_0^1 \left[\frac{r^{(2)}(\beta,u)}{r^{(0)}(\beta,u)} - \left\{ \frac{r^{(1)}(\beta,u)}{r^{(0)}(\beta,u)} \right\}^{\otimes 2} - \frac{s^{(2)}(\beta,u)}{s^{(0)}(\beta,u)} + \left\{ \frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)} \right\}^{\otimes 2} \right] s^{(0)}(\beta,u)\, dN(u),$$

$$Q = \int_0^1 \Delta(\phi)(u) Y(u) [\, \gamma(\beta,u) - \phi(\beta,u)]\, \lambda_0(u) du,$$

$$Q^* = \int_0^1 \Delta(\phi)(u) Y(u) \theta(Z,u;\alpha) \lambda_0(u) du.$$

The following conditions are needed for the theoretical results in the paper:

(1) $\int_0^1 \lambda_0(s) ds < \infty$.

(2) $Pr(Y(1) = 1|V) > 0$ for any validation set V.

(3) There exists an open subset $\mathcal{B}$, containing the true value β_0 of β, of the Euclidean space $\mathcal{R}^p$. In addition, $r_i^{(2)}(\beta,t)$ with elements $(\partial^2/\partial\beta_i\partial\beta_j) r(\beta,t)$ exists and is continuous on $\mathcal{B}$ for each $t \in [0,1]$, uniform in t, and $\phi(\beta,t)$ is bounded away from 0 on $\mathcal{B}\times[\,0,1]$. Furthermore, $I(\beta)$ is positive definite.

(4)

$$E\left\{ \sup_{\mathcal{B}\times[0,1]} |Y(t) r^{(j)}(\beta,t)| \right\} < \infty, \quad j = 0,1,2,$$

$$E\left\{ \sup_{\mathcal{B}\times[0,1]} |Y(t) \left(\frac{r^{(1)}(\beta,t)}{r(\beta,t)} \right)^{\otimes 2j} r(\beta,t)| \right\} < \infty, \quad j = 1,2,$$

$$E\left\{ \sup_{\mathcal{B}\times[0,1]} |Y(t) \left(\frac{r^{(2)}(\beta,t)}{r(\beta,t)} \right)^{\otimes j} r(\beta,t)| \right\} < \infty, \quad j = 1,2.$$

Also observe that, $s^{(0)}(\beta,t) = E[Y(t) r(\beta,t)] = E[Y(t) r^*(\beta,t)]$.

(5) Let $F_{Y(t),Z}$ be the joint distribution of $(Y(t),Z)$, and $f(t,z) = (\partial/\partial z) F_{Y(t),z}(1,z)$. For each $t \in [0,1]$, both $f(t,z)$ and $\phi(\beta,t)$ have the 2nd continuous derivative almost everywhere.

(6) $h \to 0$, $nh^{d+4} \to 0$ and $nh^d (\log n)^{-2} \to \infty$, as $n \to \infty$.

Appendix 2: Technical Proofs

Proof of Proposition 3.1. The argument employed here is similar to that for Theorem 1 of Jiang et al. (2011). Note that $\hat{\nu}_j - \nu_j = \sum_{i\in V} \omega_i(\nu_i - \nu_j) + \sum_{i\in V} \omega_i(\zeta_i - \nu_i)$. By standard nonparametric regression techniques (see for example Härdle 1990; Fan and Gijbels 1996), it can be shown that the first term above contributes to bias and is $O_p(h^2)$, which is of order $o_p(1/\sqrt{nh^d})$, if one uses an undersmoothing bandwidth such that $nh^{d+4} \to 0$, so that $\hat{\nu}_j - \nu_j = \sum_{i\in V} \omega_i(\zeta_i - \nu_i) + o_p(1/\sqrt{nh^d})$. Similarly, $\hat{\psi}_j - \psi_j = \sum_{i\in V} \omega_i(\xi_i - \psi_i) + o_p(1/\sqrt{nh^d})$. Then the asymptotic normality can be obtained by using the Cramé-Wald device and directly computing the asymptotic mean and variance (see, for example the Lemma 6.3 in Jiang and Mack 2001).

Proof of Proposition 3.2. Note that from (3.8)

$$\sqrt{nh^d}[\, \bar{\nu}_j - \nu_j] = \sqrt{nh^d}[\, \hat{\nu}_j - \nu_j] - \rho^*(Z_j,t) \frac{\sigma_1(Z_j,t)}{\sigma_2(Z_j,t)} \sqrt{nh^d}[\, (\hat{\psi}_j - \psi_j) + (\bar{\psi}_j - \psi_j)]\, (1 + o_p(1)).$$

The asymptotic normality of $\sqrt{nh^d}(\bar{v}_j - v_j)$ is obtained by the Slutsky's theorem and the asymptotic normality of $\sqrt{nh^d}(\hat{v}_j - v_j)$, $\sqrt{nh^d}(\hat{\psi}_j - \psi_j)$ and $\sqrt{nh^d}(\bar{\psi}_j - \psi_j)$.

Lemma 7.1. *Under Condition (A),*

$$\sup_{\beta\in\mathcal{B}} ||n^{-1}\partial\hat{U}(\beta,1)/\partial\beta - (-I(\beta))|| \xrightarrow{P} 0.$$

Proof. By simple algebra, we have

$$n^{-1}\partial\hat{U}(\beta,1)/\partial\beta = \int_0^1 n^{-1}\sum_{i=1}^{n}\left[\frac{\hat{r}_i^{(2)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)} - \left(\frac{\hat{r}_i^{(1)}(\beta,t)}{\hat{r}_i^{(0)}(\beta,t)}\right)^{\otimes 2} - \frac{\hat{S}^{(2)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)} + \left(\frac{\hat{S}^{(1)}(\beta,t)}{\hat{S}^{(0)}(\beta,t)}\right)^{\otimes 2}\right] dN_i(t).$$

Note that for $j = 0,\ldots,2$

$$\sup_{\mathcal{B}\times[0,1]} ||\hat{r}_i^{(j)}(\beta,t) - r_i^{(j)}(\beta,t)|| \xrightarrow{P} 0, \tag{7.14}$$

and

$$\sup_{\mathcal{B}\times[0,1]} ||\hat{S}_i^{(j)}(\beta,t) - s_i^{(j)}(\beta,t)|| \xrightarrow{P} 0, \tag{7.15}$$

uniformly for $i = 1,\ldots,n$ if $h \to 0$ and $nh^d \to \infty$. It follows that

$$\begin{aligned} n^{-1}\partial\hat{U}(\beta,1)/\partial\beta &= \int_0^1 n^{-1}\sum_{i=1}^{n}\left[\frac{r_i^{(2)}(\beta,t)}{r_i^{(0)}(\beta,t)} - \left(\frac{r_i^{(1)}(\beta,t)}{r_i^{(0)}(\beta,t)}\right)^{\otimes 2} - \frac{s^{(2)}(\beta,t)}{s^{(0)}(\beta,t)} + \left(\frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}\right)^{\otimes 2}\right] dN_i(t) + o_p(1) \\ &= -I(\beta) + o_p(1), \end{aligned}$$

uniformly in $\beta \in \mathcal{B}$. □

Proof of Theorem 4.1. The proof is argued in the framework of the multivariate counting processes, the martingale theory, and the techniques commonly used in nonparametric regression. Following the same routine as in Zhou and Wang (2000), the consistency of $\hat{\beta}_{EPL}$ can be derived by using the Inverse Function Theorem (Rudin 1964; Andersen and Gill, 1982) and the argument by Foutz (1977). In the following, we give only the asymptotic normality in Theorem 4.1. The main techniques we employed are Taylor's expansion of the score function corresponding to the estimated likelihood function (3.10), Lenglart inequality, the martingale central limit theorem (see e.g. Fleming and Harrington 1991), and nonparametric regression techniques.

By using counting process notation, the score function corresponding to the estimated partial likelihood function (3.10) at time point t can be written as

$$\hat{U}(\beta,t) = \sum_{i=1}^{n}\int_0^t \Delta(\hat{r}_i)(\beta,u)dM_i(u) + \sum_{i=1}^{n}\int_0^t \Delta(\hat{r}_i)(\beta,u)r_i(\beta_0,u)Y_i(u)\lambda_0(u)du, \tag{7.16}$$

where

$$\Delta(\hat{r}_i)(u) = \frac{\hat{r}_i^{(1)}(\beta,u)}{\hat{r}_i(\beta,u)} - \frac{\sum_{i=1}^n Y_i(u)\hat{r}_i^{1)}(\beta,u)}{\sum_{i=1}^n Y_i(u)\hat{r}_i(\beta,u)}.$$

By (7.16), $\hat{\beta}_{EPL}$ solves the equation $\hat{U}(\beta,1)=0$. By Taylor's expansion, one gets

$$n^{-1/2}\hat{U}(\beta,1) = -n^{-1}\frac{\partial \hat{U}(\beta_*,1)}{\partial\beta}\sqrt{n}(\hat{\beta}_{EPL}-\beta_0), \tag{7.17}$$

where β_* is between $\hat{\beta}_{EPL}$ and β_0. By Lemma 7.1 and consistency of $\hat{\beta}_{EPL}$,

$$-n^{-1}\frac{\partial \hat{U}(\beta_*,1)}{\partial\beta} \xrightarrow{P} I(\beta_0).$$

Therefore, to prove the asymptotic normality in the theorem it suffices to show that $n^{-1/2}\hat{U}(\beta,1)$ is asymptotically normal with mean 0 and variance $\Sigma(\beta_0) = (1-\rho)\Sigma_1(\beta_0)+\rho\Sigma_2(\beta_0)$, which is evidenced in Lemma 7.4 below.

Proof of Theorem 4.2. Using similar arguments to Theorem 4.1, we establish the result.

Lemma 7.2. *Under Condition (A),*

$$n^{-1/2}\sum_{i=1}^n\int_0^1\left(\hat{r}_i^{(k)}(\beta,w)-r_i^{(k)}(\beta,w)\right)^2 Y_i(w)r_i(\beta,w)\lambda_0(w)\,dw \xrightarrow{p} 0,\quad k=0,1$$

$$n^{-1/2}\sum_{i=1}^n\int_0^1\left(\hat{S}^{(k)}(\beta,w)-S^{(k)}(\beta,w)\right)^2 Y_i(w)r_i(\beta,w)\lambda_0(w)\,dw \xrightarrow{p} 0,\quad k=0,1.$$

Proof. The result can be obtained by following the same argument as that for Lemma 2.4 of Zhou and Wang (1999).

□

Lemma 7.3. *Under Condition (A), the second term of $\hat{U}(\beta,t)$ in (7.16) admits the following decomposition*

$$\begin{aligned} &n^{-1/2}\sum_{i=1}^n\int_0^1\Delta(\hat{r}_i)(\beta,w)Y_i(w)r_i(\beta_0,w)\lambda_0(w)\,dw \\ =\ &-n^{-1/2}\sum_{i=1}^n\int_0^1\Delta(r_i)(\beta,w)Y_i(w)[\hat{r}_i(\beta,w)-r_i(\beta,w)]\,\lambda_0(w)dw+o_p(1). \end{aligned}$$

Proof. The proof uses the same argument as that for Lemma 2.5 of Zhou and Wang (1999). By the Taylor expansion

$$\begin{aligned} f(x,y) =\ & f(x_0,y_0)+\frac{\partial f(x,y)}{\partial x}\Big|_{x_0,y_0}(x-x_0) \\ &+\frac{\partial f(x,y)}{\partial y}\Big|_{x_0,y_0}(y-y_0)+O\left((x-x_0)^2+(y-y_0)^2\right), \end{aligned}$$

if $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y^2}$, and $\frac{\partial^2 f}{\partial x\partial y}$ are finite. Then

$$\frac{\hat{r}^{(1)}}{\hat{r}} = \frac{\hat{r}^{(1)}}{r}-\frac{r^{(1)}(\hat{r}-r)}{r^2}+O\left[(\hat{r}-r)^2+\left(\hat{r}^{(1)}-r^{(1)}\right)^2\right]$$

$$\frac{\hat{S}^{(1)}}{\hat{S}^{(0)}} = \frac{\hat{S}^{(1)}}{S^{(0)}} - \frac{S^{(1)}\left(\hat{S} - S^{(0)}\right)}{S^{(0)2}} + O\left[\left(\hat{S} - S^{(0)}\right)^2 + \left(\hat{S}^{(1)} - S^{(1)}\right)^2\right].$$

Note that $\sum_i \Delta\hat{r}_i(u)\hat{r}_i(u)Y_i(u) = 0$. It follows that the left side of the result in the lemma can be expressed as

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta,w)Y_i(w)r_i(\beta,w)\lambda_0(w)\,dw \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(\hat{r}_i)(\beta,w)Y_i(w)[\,\hat{r}_i(\beta,w) - r_i(\beta,w)]\,\lambda_0(w)\,dw \\ = & -n^{-1/2}\sum_{i=1}^{n}\int_0^1 \Delta(r_i)(\beta,w)Y_i(w)[\hat{r}_i(\beta,w) - r_i(\beta,w)]\,\lambda_0(w)\,dw + o_p(1), \end{aligned}$$

where the last equality is from Lemma 7.2. Therefore the result holds. □

Lemma 7.4. *Under Condition (A),*

$$n^{-1/2}\hat{U}(\beta,1) \xrightarrow{\mathcal{L}} N\big(0,(1-\rho)\Sigma_1(\beta) + \rho\Sigma_2(\beta)\big).$$

Proof. Note that $\hat{r}_i - r_i = (1-\eta_i)\left(\bar{\phi}_i - \phi_i\right)$. Applying the first order approximation $x/y = x_0/y_0 + (x-x_0)/y_0 - (y-y_0)x_0/y_0^2 + O\left((x-x_0)^2 + (y-y_0)^2\right)$ to $\hat{r}^{(1)}/\hat{r}$ and $\hat{S}^{(1)}/\hat{S}^{(0)}$ around $(r^{(1)}, r)$ and $(s^{(1)}, s^{(0)})$, respectively, and by Lemma 7.3 the second term of $n^{-1/2}\hat{U}(\beta,1)$ in (7.16) becomes

$$\begin{aligned} & \Delta(r_i)(\beta,w)Y_i(w)[\hat{r}_i(\beta,w) - r_i(\beta,w)]\,\lambda_0(w)\,dw + o_p(1) \\ = & -n^{-1/2}\sum_{j\in\bar{V}}\int_0^1 \left(\bar{\phi}_j - \phi_j\right)\Delta(\phi_j)(u)Y_j(u)\lambda_0(u)du + o_p(1) \\ = & I_{n1} + o_p(1). \end{aligned} \tag{7.18}$$

Note that $\hat{\phi}_j(\beta,t) = \hat{v}_j(\beta_1,t)\exp\left\{\beta_2' Z_j(t)\right\}$. Since

$$\begin{aligned} \bar{\phi}_j - \phi_j = & (\hat{\phi}_j - \phi_j) - \exp\left\{\beta_2' Z_j(u)\right\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\left(\hat{\psi}_j - \bar{\psi}_j\right)(1+o_p(1)) \\ = & \sum_{i\in V}\omega_i(\gamma_i - \phi_j) - \exp\{\beta_2' Z_j(u)\}\left[\sum_{i\in V}\omega_i(\xi_i - \psi_j)\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\right. \\ & \left. - \sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i - \psi_j)\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\right](1+o_p(1)) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ = & \sum_{i\in V}\omega_i\left[(\gamma_i - \phi_j) - \exp\left\{\beta_2' Z_j(u)\right\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\xi_i - \psi_j)\right](1+o_p(1)) \\ & + \sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i - \psi_j)\exp\left\{\beta_2' Z_j(u)\right\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(1+o_p(1)) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

the first term in (7.18) can be rewritten as

$$
\begin{aligned}
I_{n1} &= -n^{-1/2}\sum_{j\in\bar{V}}\int_0^1 \Delta(\phi_j)(u)Y_j(u)\lambda_0(u)\\
&\times\left\{\sum_{i\in V}\omega_i\left[(\gamma_i-\phi_j)-\exp\left\{\beta_2'Z_j(u)\right\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\xi_i-\psi_j)\right]\right.\\
&\left.+\sum_{i\in V\cup\bar{V}}\bar{\omega}_i(\xi_i-\psi_j)\exp\left\{\beta_2'Z_j(u)\right\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\right\}du(1+o_p(1))+o_p(1)\\
&\equiv J_{n1}+J_{n2}+o_p(1).
\end{aligned}
$$

Note that

$$n_v^{-1}\sum_{i\in V}Y_i(t)K_h(Z_i-Z_j)=f(t,Z_j)(1+o_p(1)),$$

$$n^{-1}\sum_{i\in V\cup\bar{V}}Y_i(t)K_h(Z_i-Z_j)=f(t,Z_j)(1+o_p(1)),$$

$$\omega_i(t,Z_j;h)=f^{-1}(t,Z_j)(1+o_p(1))\,n_v^{-1}Y_i(t)K_h(Z_i-Z_j),$$

$$\bar{\omega}_i(t,Z_j;h)=f^{-1}(t,Z_j)(1+o_p(1))\,n^{-1}Y_i(t)K_h(Z_i-Z_j),$$

uniformly for $j=1,\ldots,n$. Then

$$
\begin{aligned}
J_{n1} &= -\frac{1}{\sqrt{n}}\sum_{j\in\bar{V}}\int_0^1\Delta(\phi_j)(u)Y_j(u)\lambda_0(u)f^{-1}(u,Z_j)\times\\
&\quad\frac{1}{n_v}\sum_{i\in V}Y_i(u)K_h(Z_i-Z_j)\left[(\gamma_i-\phi_j)-\exp\{\beta_2'Z_j(u)\}\right.\\
&\quad\left.\times\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}(\xi_i-\psi_j)\right]du+o_p(1)\\
&= -\frac{1}{\sqrt{n}}\frac{n-n_v}{n_v}\sum_{i\in V}\left[Q_i-Q_i^*\right]+o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
J_{n2} &= -\frac{1}{\sqrt{n}}\sum_{j\in\bar{V}}\int_0^1\Delta(\phi_j)(u)Y_j(u)\lambda_0(u)\exp\{\beta_2'Z_j(u)\}\rho_\alpha^*(Z_j,u)\frac{\sigma_1(Z_j,u)}{\sigma_2(Z_j,u)}\\
&\quad\times\frac{1}{n}\sum_{i\in V\cup\bar{V}}Y_i(u)K_h(Z_i-Z_j)(\xi_i-\psi_j)f^{-1}(u,Z_j)du+o_p(1)\\
&= -\frac{1}{\sqrt{n}}\frac{n-n_v}{n}\sum_{i\in V\cup\bar{V}}Q_i^*+o_p(1).
\end{aligned}
$$

Therefore, the second term of $n^{-1/2}\hat{U}(\beta,1)$ in (7.16) equals

$$-\frac{1}{\sqrt{n}}\frac{n-n_v}{n_v}\sum_{i\in V}[Q_i-Q_i^*]-\frac{1}{\sqrt{n}}\frac{n-n_v}{n}\sum_{i\in V\cup\bar{V}}Q_i^*+o_p(1).$$

Hence, $n^{-1/2}\hat{U}(\beta,1)$ can be expressed as

$$n^{-1/2}\sum_{i\in\bar{V}}\left[\int_0^1\left\{\frac{\phi_i^{(1)}(\beta,u)}{\phi_i(\beta,u)}-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right\}dM_i(s)-\frac{n-n_v}{n}Q_i^*\right]+o_p(1)$$

$$+\ n^{-1/2}\sum_{i\in V}\left[\int_0^1\left\{\frac{r_i^{(1)}(\beta,u)}{r_i(\beta,u)}-\frac{s^{(1)}(\beta,u)}{s^{(0)}(\beta,u)}\right\}dM_i(s)-\frac{n-n_v}{n_v}(Q_i-Q_i^*\frac{n-n_v}{n})\right].$$

For the 1st and 3rd terms above, each of them is a sum of independently distributed terms with mean zero from the nonvalidation and validation subsamples, respectively. The 1st term converges weakly to a gaussian process with covariance $(1-\rho)\Sigma_1(\beta_0)$. The 3rd term is asymptotically normal with mean zero and variance $\rho\Sigma_2(\beta_0)$. By independence of the two terms, $n^{-1/2}\hat{U}(\beta,1)\xrightarrow{P}N(0,\Sigma(\beta))$ with $\Sigma(\beta)=(1-\rho)\Sigma_1(\beta)+\rho\Sigma_2(\beta)$. □

Additional file

Additional file 1: The codes for numerical results in the paper.

Acknowledgements
The work was partially supported by NSF grant DMS-0906482 and NSFC grant 71361010. The research of Yanqing Sun was partially supported by the National Institutes of Health NIAID [grant number R37 AI054165] and by the National Science Foundation [grant number DMS-1208978]. The authors would also like to thank Dr. Savitz for making the data on serrum ferritin concentration in relation to preterm delivery study available for the application.

Author details
[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA.
[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

Received: 28 February 2014 Accepted: 14 November 2014
Published online: 20 February 2015

References

Andersen, PK: Gill: Cox's regression model for counting processes: a large sample study. Lifetime Data Anal. **10**, 1100–1120 (1982)
Chen, Y-H, Chen, R: A unified approach to regression analysis under double-sampling designs. J. R. Stat. Soc. B. **62**, 449–460 (2000)
Cox, DR: Regression models and life-tables (with discussion). J. R. Stat. Soc. B. **34**, 187–220 (1972)
Dickson, ER, Grambsch, PM, Fleming, TR, Fisher, LD, Langworthy, A: Prognosis in Primary Biliary Cirrhosis: Model for Decision Making. Hepatology. **10**, 1–7 (1989)
Fan, J, Gijbels, I: Local Polynomial Modeling and Its Applications. Chapman and Hall, London (1996)
Fan, J, Yao, Q: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York (2003)
Fan, Z, Wang, X: Marginal hazards model for multivariate failure time data with auxiliary covariates. J. Nonparametric Stat. **21**, 771–786 (2009)
Fleming, TR: Harrington, DP Counting Process and Survival Analysis. Wiley, New York (1991)
Foutz, RV: On the unique consistent solution to the likelihood equations. J. Am. Stat. Assoc. **72**, 147–148 (1977)
Härdle, W: Applied Nonparametric Regression. Cambridge University Press, London (1990)
Hughes, MD: Regression dilution in the proportional hazards model. Biometrics. **49**, 1056–1066 (1993)
Jiang, X, Jiang, J, Liu, Y: Nonparametric regression under double-sampling designs. J. Syst. Sci. Complex. **24**, 167–175 (2011)
Jiang, J, Mack, YP: Robust local polynomial regression for dependent data. Stat. Sinica. **11**, 705–722 (2001)
Kalbfleisch, JD, Prentice, RL: The Statistical Analysis of Failure Time Data. Wiley, New York (1980)
Lin, DY, Ying, Z: Cox regression with incomplete covariate measurements. J. Am. Stat. Assoc. **88**, 1341–1349 (1993)
Lipsitz, S, Ibrahim, JG: Using the E-M algorithm for survival data with incomplete categorical covariates. Lifetime Data Anal. **2**, 5–14 (1996)
Liu, Y, Wu, Y, Zhou, H: Multivariate failure times regression with a continuous auxiliary covariate. J. Multivariate Anal. **101**, 679–691 (2010)
Markus, BH, Dickson, ER, Grambsch, PM, Fleming, TR, Mazzaferro, V, Klintmalm, GB, Wiesner, RH, Van Thiel, DH, Starzl, TE: Efficiency of liver transplantation in patients with primary biliary cirrhosis. N. Engl. J. Med. **320**, 1709–1713 (1989)
Nadaraya, EA: On estimating regression. Theory Probab. Appl. **10**, 186–190 (1964)
Pepe, MS, Self, SG, Prentice, RL: Further results on covariate measurement errors in cohort studies with time to response data. Statist. Med. **8**, 1167–1178 (1989)
Prentice, RL: Covariate measurement errors and parameter estimation in a failure time regression model. Biomtrika. **69**, 331–342 (1982)
Rubin, DB: Inference and missing data. Biomtrika. **63**, 581–592 (1976)
Rudin, W: Principle of Mathematical Analysis. McGraw-Hill Book Co., New York (1964)

Savitz, DA, Dole, N, Jr Terry, JW, Zhou, H, Jr Thorp, JM: Smoking and pregnancy outcome among African-American and white women in central North Carolina. Epidemiology. **12**, 636–642 (2001)
Watson, GS: Smooth regression analysis. Sankhya A. **26**, 359–372 (1964)
Zhou, H, Pepe, MS: Auxiliary covariate data in failure time regression analysis. Biomtrika. **82**, 139–149 (1995)
Zhou, H, Wang, C-Y: Some asymptotic results for using kernel smoother with covariate measurement error problem in survival analysis, Vol. 2200. University of North Carolina, Chapill Hill (1999)
Zhou, H, Wang, C-Y: Failure time regression with continuous covariates measured with error. J. R. Stat. Soc. B. **62**, 657–665 (2000)