© Board of the Foundation of the Scandinavian Journal of Statistics 2002. Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA Vol 29: 637–655, 2002

Cox Regression with Covariate Measurement Error

CHENGCHENG HU

Harvard University

D. Y. LIN

University of North Carolina

ABSTRACT. This article deals with parameter estimation in the Cox proportional hazards model when covariates are measured with error. We consider both the classical additive measurement error model and a more general model which represents the mis-measured version of the covariate as an arbitrary linear function of the true covariate plus random noise. Only moment conditions are imposed on the distributions of the covariates and measurement error. Under the assumption that the covariates are measured precisely for a validation set, we develop a class of estimating equations for the vector-valued regression parameter by correcting the partial likelihood score function. The resultant estimators are proven to be consistent and asymptotically normal with easily estimated variances. Furthermore, a corrected version of the Breslow estimator for the cumulative hazard function is developed, which is shown to be uniformly consistent and, upon proper normalization, converges weakly to a zero-mean Gaussian process. Simulation studies indicate that the asymptotic approximations work well for practical sample sizes. The situation in which replicate measurements (instead of a validation set) are available is also studied.

Key words: censoring, corrected score, mismeasured covariates, partial likelihood, proportional hazards, survival data

1. Introduction

The proportional hazards model (Cox, 1972) specifies that the cumulative hazard function for the survival time associated with possibly time-dependent covariates $\mathbf{Z}$ takes the form $\Lambda(t|\mathbf{Z}) = \int_0^t \exp(\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{Z}(s))d\Lambda_0(s)$, where $\boldsymbol{\beta}_0$ is a vector-valued regression parameter, and $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function. Suppose that we have a random sample of n subjects. For $i = 1, \ldots, n$, let T_i be the survival time, C_i be the censoring time, and $\mathbf{Z}_i(\cdot)$ be the vector of covariates. Write $\tilde{T}_i = \min(T_i, C_i)$, $\delta_i = I(T_i \le C_i)$, $N_i(t) = \delta_i I(\tilde{T}_i \le t)$, and $Y_i(t) = I(\tilde{T}_i \ge t)$, where $I(\cdot)$ is the indicator function. Suppose that C is independent of T conditional on $\mathbf{Z}$ and that the data are observed on the time interval $[0, \tau]$, where $0 < \tau < \infty$. Then a consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ can be obtained by solving the partial likelihood score equation $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$, where

$$\mathbf{U}(\beta) = \sum_{i=1}^{n} \int_0^\tau \left\{ \mathbf{Z}_i(t) - \frac{\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j(t))\mathbf{Z}_j(t)}{\sum_{j=1}^{n} Y_j(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_j(t))} \right\} dN_i(t). \tag{1}$$

The baseline cumulative hazard $\Lambda_0(t)$ can be estimated by

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s)\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_j(t))} \tag{2}$$

(Breslow, 1972).

In many applications, covariates are subject to measurement error. A common approach to deal with this problem is to use the mismeasured version of the covariate directly. It is well-known that this practice causes bias in the estimated regression parameter. Great efforts have been made in developing methods to correct this bias; see Fuller (1987) and Carroll *et al.* (1995) for excellent reviews of various methods for non-censored data. Several methods have been developed for the Cox regression with censored data. Prentice (1982) proposed using the induced partial likelihood under the restrictive assumption that the event is rare. Zhou & Pepe (1995) developed a consistent estimator for the regression parameter when the covariates are discrete. Wang *et al.* (1997) developed a regression calibration method which only gives approximately consistent estimators. Hu *et al.* (1998) described some likelihood-based methods under the classical additive error model (Carroll *et al.*, 1995, p. 8) with a known error distribution.

Stefanski (1989) and Nakamura (1990) developed the approach of corrected score functions to remove the bias induced by measurement error. For the proportional hazards model, an exact corrected score does not exist. Nakamura (1992) introduced two approximately corrected scores under the restrictive assumption of additive normal error with known covariance matrix. The properties of the resulting estimators were explored only by simulations. Buzas (1998) removed the condition of normal error, but assumed that the moment generating function of the error distribution is known. The asymptotic properties of the resulting estimator were not carefully studied. Neither Nakamura nor Buzas studied the estimation of $\Lambda_0(\cdot)$.

In this paper, we extend the work of Nakamura (1992) and Buzas (1998) to obtain a broad class of consistent estimators for the regression parameter of the Cox model when one or more covariates are measured with error. We pay primary attention to the setting in which the surrogate covariate, i.e. the mismeasured version of the covariate, is measured on all study subjects while the true covariate is ascertained on a randomly selected validation set. By using the validation set data to estimate the error distribution, we provide a practical way of calculating the parameter estimators and their variances in real applications. The distributions of the covariates and measurement error are completely unspecified. The proposed estimators are proven to be consistent and asymptotically normal with easily estimated variances. Furthermore, we develop a corrected version of the Breslow estimator for the cumulative hazard function, which is shown to be uniformly consistent and asymptotically normal. Finally, we extend the results to the situations in which replicate measurements rather than validation sets are available.

2. Classical error model with a single covariate

Suppose that there is only a single time-independent covariate. For the ith subject, let Z_i be the true covariate, and W_i be the surrogate covariate. Assume that $W_i = Z_i + \epsilon_i$, where ϵ_i is a zero-mean measurement error, which is independent of T_i, C_i and Z_i. Let $\xi_i = 1$ if the ith subject is in the validation set, and $\xi_i = 0$ otherwise. Assume that ξ_i $(i = 1, \dots, n)$ are i.i.d. with $P(\xi_i = 1) = \alpha$, and are independent of all other variables. Let $\bar{\xi}_i = 1 - \xi_i$ and $\bar{\alpha} = 1 - \alpha$. The complete data consist of i.i.d. random elements $(\tilde{T}_i, \delta_i, Z_i, W_i, \xi_i)$ $(i = 1, \dots, n)$, but we cannot observe Z_i if $\xi_i = 0$.

If we could observe all Z_is, the partial likelihood score function (1) would be $U(\beta) = \sum_{i=1}^{n} \int_0^{\tau} \{Z_i - S^{(1)}(\beta, t)/S^{(0)}(\beta, t)\}\, dN_i(t)$, where $S^{(k)}(\beta, t) = n^{-1} \sum_{j=1}^{n} Y_j(t) \exp(\beta Z_j) Z_j^k$ $(k = 0, 1, 2)$. When $\xi_i = 0$, we cannot observe the true covariate Z_i, but only the surrogate W_i. A naïve estimator of β_0 is obtained by substituting W_i for Z_i in $U(\beta)$ for all the subjects who are not in the validation set. Let $\mathscr{F}_t$ be the σ-algebra generated by all the failure, censoring and true covariate histories of all the subjects over $[0, t]$. Although $E(W_i|\mathscr{F}_\tau) = E(W_i|Z_i) = Z_i$, in general $E(\exp(\beta W_i)|\mathscr{F}_\tau) \neq \exp(\beta Z_i)$ and $E(\exp(\beta W_i)W_i|\mathscr{F}_\tau) \neq \exp(\beta Z_i)Z_i$. Thus, the naïve estimator for β_0 is not asymptotically unbiased. Our goal is to correct this bias.

Let $\eta_k(\beta) = E(\epsilon_1^k \exp(\beta\epsilon_1))$ $(k = 0, 1, 2)$. Clearly,

$$E(\exp(\beta W_i)|\mathscr{F}_\tau) = \eta_0(\beta)\exp(\beta Z_i), \tag{3}$$

$$E(\exp(\beta W_i)W_i|\mathscr{F}_\tau) = \eta_0(\beta)\exp(\beta Z_i)Z_i + \eta_1(\beta)\exp(\beta Z_i). \tag{4}$$

Thus, when the true covariate Z_i is not available, i.e. $\xi_i = 0$, we may replace $\exp(\beta Z_i)$ and $\exp(\beta Z_i)Z_i$ in $U(\beta)$ by $\eta_0^{-1}(\beta)\exp(\beta W_i)$ and $\eta_0^{-1}(\beta)\exp(\beta W_i)W_i - \eta_0^{-2}(\beta)\exp(\beta W_i)\eta_1(\beta)$, respectively. Define

$$R_i^{(0)}(\beta) = \xi_i \exp(\beta Z_i) + \bar{\xi}_i \eta_0^{-1}(\beta)\exp(\beta W_i), \tag{5}$$

$$R_i^{(1)}(\beta) = \xi_i \exp(\beta Z_i)Z_i + \bar{\xi}_i \eta_0^{-1}(\beta)\exp(\beta W_i)\{W_i - \eta_0^{-1}(\beta)\eta_1(\beta)\}. \tag{6}$$

It is easy to see that $E\{n^{-1}\sum_{i=1}^n Y_i(t)R_i^{(k)}(\beta)|\mathscr{F}_\tau\} = S^{(k)}(\beta, t)$ $(k = 0, 1)$. Thus we consider the following modification of $U(\beta)$:

$$\sum_{i=1}^n \int_0^\tau \left\{ (\xi_i Z_i + \bar{\xi}_i W_i) - \frac{\sum_{j=1}^n Y_j(t)R_j^{(1)}(\beta)}{\sum_{j=1}^n Y_j(t)R_j^{(0)}(\beta)} \right\} dN_i(t). \tag{7}$$

The above expression is not a genuine estimating function for β_0 because η_0 and η_1 are unknown. For fixed β, however, we can estimate $\eta_k(\beta)$ $(k = 0, 1, 2)$ consistently from the validation set by the moment estimators $\hat{\eta}_k(\beta) = \sum_{i=1}^n \xi_i \exp(\beta\epsilon_i)\epsilon_i^k / \sum_{i=1}^n \xi_i$ $(k = 0, 1, 2)$. This motivates us to replace $R_i^{(0)}(\beta)$ and $R_i^{(1)}(\beta)$ in (7) with $\hat{R}_i^{(0)}(\beta)$ and $\hat{R}_i^{(1)}(\beta)$, which are obtained from (5) and (6) by replacing $\eta_k(\beta)$ with $\hat{\eta}_k(\beta)$ $(k = 0, 1)$. To increase efficiency, we introduce a weight $\omega \in [0, 1]$ to down-weight the influence of the subjects in the non-validation set. Thus, we propose the following estimating function

$$U_C(\beta, \omega) = \sum_{i=1}^n (\xi_i + \omega\bar{\xi}_i) \int_0^\tau \{\xi_i Z_i + \bar{\xi}_i W_i - E_C(\beta, t, \omega)\} dN_i(t),$$

where $E_C(\beta, t, \omega) = S_C^{(1)}(\beta, t, \omega)/S_C^{(0)}(\beta, t, \omega)$, and $S_C^{(k)}(\beta, t, \omega) = n^{-1}\sum_{j=1}^n (\xi_j + \omega\bar{\xi}_j)Y_j(t)\hat{R}_j^{(k)}(\beta)$ $(k = 0, 1)$. Denote the solution of $U_C(\beta, \omega) = 0$ by $\hat{\beta}_C \equiv \hat{\beta}_C(\omega)$. In addition, by analogy to (2), we propose a corrected Breslow estimator for the cumulative hazard function $\Lambda_0(t)$:

$$\hat{\Lambda}_C(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)\hat{R}_j^{(0)}(\hat{\beta}_C)}.$$

Let $s^{(k)}(\beta, t) = E\{S^{(k)}(\beta, t)\}$ $(k = 0, 1, 2)$, $e(\beta, t) = s^{(1)}(\beta, t)/s^{(0)}(\beta, t)$, $v(\beta, t) = s^{(2)}(\beta, t)/s^{(0)}(\beta, t) - e(\beta, t)^2$, $\varphi = \int_0^\tau s^{(0)}(\beta_0, t)d\Lambda_0(t)$, and $\Gamma = \int_0^\tau v(\beta_0, t)s^{(0)}(\beta_0, t)d\Lambda_0(t)$. Also define

$$\Gamma_v = E\Bigg[\int_0^\tau \{W_1 - e(\beta_0, t)\}\{dN_1(t) - \eta_0^{-1}(\beta_0)\exp(\beta_0 W_1)Y_1(t)d\Lambda_0(t)\} + \eta_0^{-2}(\beta_0)\eta_1(\beta_0)\int_0^\tau \exp(\beta_0 W_1)Y_1(t)d\Lambda_0(t)\Bigg]^2,$$

$$\Gamma_r = (\bar{\alpha}/\alpha)^2\varphi^2\eta_0^{-4}(\beta_0)\{\eta_0^2(\beta_0)\eta_2(2\beta_0) - 2\eta_0(\beta_0)\eta_1(\beta_0)\eta_1(2\beta_0) + \eta_1^2(\beta_0)\eta_0(2\beta_0)\}.$$

Then we have the following theorem for the asymptotic distribution of $n^{-1/2}U_C(\beta_0, \omega)$:

Theorem 1
Under conditions 1–5 given in the appendix, $n^{-1/2}U_C(\beta_0,\omega) \to_d N(0,\Gamma_C)$, *where* $\Gamma_C = \alpha\Gamma + \omega^2(\bar{\alpha}\Gamma_v + \alpha\Gamma_r)$.

The proof of theorem 1 is given in the appendix. The following theorem establishes the consistency and asymptotic normality of $\hat{\beta}_C$:

Theorem 2
Under conditions 1–5, $\hat{\beta}_C(\omega)$ *exists and is unique in a neighbourhood* $\mathscr{B}$ *of* β_0 *with probability converging to one as* $n \to \infty$, *and* $\hat{\beta}_C(\omega) \to_p \beta_0$. *Furthermore,* $n^{1/2}\{\hat{\beta}_C(\omega) - \beta_0\} \to_d N(0,\Gamma_\beta)$, *where* $\Gamma_\beta = \Gamma_C/\{(\alpha + \omega\bar{\alpha})\Gamma\}^2$.

The proof of this theorem and a consistent estimator $\hat{\Gamma}_\beta$ for Γ_β are given in the appendix.

For any $\omega \geq 0$, $\hat{\beta}_C(\omega)$ is a consistent estimator of β_0, and its asymptotic variance is a function of ω. By taking the derivative of Γ_β with respect to ω, we see that Γ_β achieves its minimum at $\omega_{\text{opt}} = \bar{\alpha}\Gamma/(\alpha\Gamma_r + \bar{\alpha}\Gamma_v)$. Thus, $\hat{\beta}_C(\omega_{\text{opt}})$ has the smallest asymptotic variance in the class of consistent estimators $\hat{\beta}_C(\omega)$ $(\omega \geq 0)$. In particular, at least for large n, $\hat{\beta}_C(\omega_{\text{opt}})$ will be more efficient than $\hat{\beta}_C(0)$, which is the estimator of the complete-case analysis. We can estimate ω_{opt} by $\hat{\omega}_{\text{opt}} = \hat{\bar{\alpha}}\hat{\Gamma}/(\hat{\alpha}\hat{\Gamma}_r + \hat{\bar{\alpha}}\hat{\Gamma}_v)$, where $\hat{\alpha} = n^{-1}\sum_{i=1}^n \xi_i$, $\hat{\bar{\alpha}} = 1 - \hat{\alpha}$, and $\hat{\Gamma}$, $\hat{\Gamma}_r$ and $\hat{\Gamma}_v$ are given in the appendix. Because $\hat{\omega}_{\text{opt}}$ depends on $\hat{\beta}_C$, these two estimators can be obtained simultaneously by a simple iterative procedure, or a preliminary consistent estimator of β_0 can be calculated with $\omega = 0$ or 0.5 and plugged into the expression for $\hat{\omega}_{\text{opt}}$, which is in turn used to obtain the final $\hat{\beta}_C(\hat{\omega}_{\text{opt}})$. Our simulation results show that the one-step computation is satisfactory.

The next theorem establishes the weak convergence and uniform consistency of $\hat{\Lambda}_C(\cdot)$, and the proof is again given in the appendix.

Theorem 3
Assume that conditions 1–5 hold. Then $n^{1/2}\{\hat{\Lambda}_C(t) - \Lambda_0(t)\}$ *converges weakly to a zero-mean Gaussian process with covariance function* $\phi(t,s)$ *at time points* (t,s), *where* ϕ *is defined by (13) in the appendix. Furthermore,* $\sup_{t\in[0,\tau]}|\hat{\Lambda}_C(t) - \Lambda_0(t)| \to_p 0$.

3. Classical error model with multiple covariates

In this section, we extend the results of section 2 to accommodate multiple covariates. Suppose that, for the ith subject, a p-vector of time-independent covariates $\mathbf{Z}_i$ is measured with error and a q-vector of possibly time-dependent covariates $\mathbf{X}_i(t)$ is measured precisely. The surrogate for $\mathbf{Z}_i$ is $\mathbf{W}_i = \mathbf{Z}_i + \boldsymbol{\epsilon}_i$, where the vector-valued error terms $\boldsymbol{\epsilon}_i$ $(i = 1,\ldots,n)$ are i.i.d. with mean $\mathbf{0}_p$ and are independent of all other random variables. Here, $\mathbf{0}_p$ is a p-vector of 0s. Let $\mathbf{H}_i(t) = (\mathbf{Z}_i^{\mathrm{T}}, \mathbf{X}_i(t)^{\mathrm{T}})^{\mathrm{T}}$ and $\hat{\mathbf{H}}_i(t) = (\mathbf{W}_i^{\mathrm{T}}, \mathbf{X}_i(t)^{\mathrm{T}})^{\mathrm{T}}$. The Cox model specifies that $d\Lambda(t|\mathbf{H}_i) = \exp(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{H}_i(t))d\Lambda_0(t)$, where $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}})^{\mathrm{T}}$ with $\boldsymbol{\beta}_0$ and γ_0 pertaining to $\mathbf{Z}_i$ and $\mathbf{X}_i(t)$, respectively.

For vectors $\mathbf{a} = (a_1,\ldots,a_p)^{\mathrm{T}}$ and $\mathbf{b} = (b_1,\ldots,b_q)^{\mathrm{T}}$, define $\mathbf{a}\otimes\mathbf{b}$ to be the $p\times q$ matrix with a_ib_j as the (i,j)th element, and let $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\otimes\mathbf{a}$. Write $\mathbf{S}^{(k)}(\boldsymbol{\theta},t) = n^{-1}\sum_{i=1}^n Y_i(t)\exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t))\mathbf{H}_i^{\otimes k}(t)$ $(k = 0,1,2)$, and $\mathbf{s}^{(k)}(\boldsymbol{\theta},t) = E\{\mathbf{S}^{(k)}(\boldsymbol{\theta},t)\}$. We further define $\mathbf{e}(\boldsymbol{\theta},t) = \mathbf{s}^{(1)}(\boldsymbol{\theta},t)/s^{(0)}(\boldsymbol{\theta},t)$, $\mathbf{v}(\boldsymbol{\theta},t) = \mathbf{s}^{(2)}(\boldsymbol{\theta},t)/s^{(0)}(\boldsymbol{\theta},t) - \mathbf{e}^{\otimes 2}(\boldsymbol{\theta},t)$, and $\boldsymbol{\Gamma} = \int_0^\tau \mathbf{v}(\boldsymbol{\theta}_0,t)s^{(0)}(\boldsymbol{\theta}_0,t)d\Lambda_0(t)$.

If $\mathbf{Z}$ were measured on all study subjects, the partial likelihood score function for $\boldsymbol{\theta}_0$ would be $\mathbf{U}(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \{\mathbf{H}_i(t) - \mathbf{S}^{(1)}(\boldsymbol{\theta},t)/S^{(0)}(\boldsymbol{\theta},t)\}dN_i(t)$. When only surrogate $\mathbf{W}$ can be observed for subjects not in the validation set, we define $\boldsymbol{\eta}_k(\boldsymbol{\beta}) = E(\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\epsilon}_1)\boldsymbol{\epsilon}_1^{\otimes k})$ $(k = 0, 1, 2)$, and, for $i = 1, \ldots, n$, let

$$R_i^{(0)}(\boldsymbol{\theta}, t) = \xi_i \exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t)) + \bar{\xi}_i \eta_0^{-1}(\beta) \exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t),$$

$$\mathbf{R}_i^{(1)}(\boldsymbol{\theta}, t) = \xi_i \exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t))\mathbf{H}_i(t) + \bar{\xi}_i \eta_0^{-1}(\beta) \exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t)) \left\{ \hat{\mathbf{H}}_i(t) - \begin{bmatrix} \eta_0^{-1}(\boldsymbol{\beta})\boldsymbol{\eta}_1(\boldsymbol{\beta}) \\ \mathbf{0}_q \end{bmatrix} \right\}.$$

It is not difficult to see that $E\{n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{R}_i^{(k)}(\boldsymbol{\theta}, t)|\mathscr{F}_\tau\} = \mathbf{S}^{(k)}(\boldsymbol{\theta}, t)$ $(k = 0, 1)$, where $\mathscr{F}_\tau$ is the σ-algebra generated by $\{N_i(t), Y_i(t), \mathbf{H}_i(t) : t \in [0, \tau]; i = 1, \ldots, n\}$. For fixed $\boldsymbol{\beta}$, we can estimate $\boldsymbol{\eta}_k(\boldsymbol{\beta})$ consistently by $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}) = \sum_{i=1}^n \xi_i \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\epsilon}_i)\boldsymbol{\epsilon}_i^{\otimes k} / \sum_{i=1}^n \xi_i$ $(k = 0, 1, 2)$. Then, for $k = 0, 1$ and $i = 1, \ldots, n$, we define $\hat{\mathbf{R}}_i^{(k)}(\boldsymbol{\theta}, t)$ in the same way as $\mathbf{R}_i^{(k)}(\boldsymbol{\theta}, t)$, but with $\eta_0(\boldsymbol{\beta})$ and $\boldsymbol{\eta}_1(\boldsymbol{\beta})$ replaced by $\hat{\eta}_0(\boldsymbol{\beta})$ and $\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta})$, respectively. Also, as in section 2, in deriving the estimating function, we may downweight the contributions of the subjects in the non-validation set to improve efficiency. Here the weighting is achieved by a $(p+q) \times (p+q)$ matrix $\boldsymbol{\Omega}$. Let $\mathbf{A}_i = \xi_i \mathbf{I}_{p+q} + \bar{\xi}_i \boldsymbol{\Omega}$, where $\mathbf{I}_{p+q}$ is the $(p+q) \times (p+q)$ identity matrix. Let $\mathbf{S}_C^{(k)}(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) = n^{-1}\sum_{j=1}^n Y_j(t)\mathbf{A}_j\hat{\mathbf{R}}_j^{(k)}(\boldsymbol{\theta}, t)$ $(k = 0, 1)$, and $\mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) = \{\mathbf{S}_C^{(0)}(\boldsymbol{\theta}, t, \boldsymbol{\Omega})\}^{-1}\mathbf{S}_C^{(1)}(\boldsymbol{\theta}, t, \boldsymbol{\Omega})$. Then we can define the estimating function for $\boldsymbol{\theta}_0$:

$$\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}) = \sum_{i=1}^n \mathbf{A}_i \int_0^\tau \left\{ \xi_i \mathbf{H}_i(t) + \bar{\xi}_i \hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) \right\} dN_i(t).$$

Let $\hat{\boldsymbol{\theta}}_C \equiv \hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega})$ be the solution to $\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}) = \mathbf{0}$. As a generalization of (2), the corrected version of the Breslow estimator for $\Lambda_0(t)$ is

$$\hat{\Lambda}_C(t) \equiv \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)\hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s)}.$$

Similar to section 2, let $\boldsymbol{\Gamma}_v = E[\int_0^\tau \{\hat{\mathbf{H}}_1(t) - \mathbf{e}(\boldsymbol{\theta}_0, t)\}[dN_1(t) - \eta_0^{-1}(\boldsymbol{\beta}_0)\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_1(t)) \times Y_1(t)d\Lambda_0(t)] + \eta_0^{-2}(\boldsymbol{\beta}_0)\int_0^\tau \exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_1(t))Y_1(t)d\Lambda_0(t)\{\boldsymbol{\eta}_1(\boldsymbol{\beta}_0)^{\mathrm{T}}, \mathbf{0}_q^{\mathrm{T}}\}^{\mathrm{T}}]^{\otimes 2}$, and

$$\boldsymbol{\Gamma}_r = (\bar{\alpha}/\alpha)^2 \varphi^2 \eta_0^{-4}(\boldsymbol{\beta}_0) \begin{bmatrix} \boldsymbol{\Gamma}_r^\dagger & \mathbf{0}_{p \times q} \\ \mathbf{0}_{q \times p} & \mathbf{0}_{q \times q} \end{bmatrix},$$

where $\boldsymbol{\Gamma}_r^\dagger = \eta_0^2(\boldsymbol{\beta}_0)\boldsymbol{\eta}_2(2\boldsymbol{\beta}_0) - \eta_0(\boldsymbol{\beta}_0)\boldsymbol{\eta}_1(2\boldsymbol{\beta}_0) \otimes \boldsymbol{\eta}_1(\boldsymbol{\beta}_0) - \eta_0(\boldsymbol{\beta}_0)\boldsymbol{\eta}_1(\boldsymbol{\beta}_0) \otimes \boldsymbol{\eta}_1(2\boldsymbol{\beta}_0) + \eta_0(2\boldsymbol{\beta}_0)\,\boldsymbol{\eta}_1^{\otimes 2}(\boldsymbol{\beta}_0)$, and $\varphi = \int_0^\tau s^{(0)}(\boldsymbol{\theta}_0, t)d\Lambda_0(t)$. The asymptotic normality of $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0, \boldsymbol{\Omega})$ is established in the following theorem.

Theorem 4
Under conditions 1, 2′, 3′, 4, 5 and 6 given in the appendix, $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0, \boldsymbol{\Omega}) \to_d N(\mathbf{0}, \boldsymbol{\Gamma}_C)$, *where* $\boldsymbol{\Gamma}_C = \alpha\boldsymbol{\Gamma} + \boldsymbol{\Omega}(\bar{\alpha}\boldsymbol{\Gamma}_v + \alpha\boldsymbol{\Gamma}_r)\boldsymbol{\Omega}^{\mathrm{T}}$.

A brief proof can be found in the appendix. The next theorem follows from theorem 4, together with the arguments given in the proof of theorem 2.

Theorem 5
Under conditions 1, 2′, 3′, 4, 5 and 6, $\hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega})$ *exists and is unique in a neighbourhood of* $\boldsymbol{\theta}_0$ *with probability converging to one as* $n \to \infty$, *and* $\hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega}) \to_p \boldsymbol{\theta}_0$. *Furthermore,* $n^{1/2}\{\hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega}) - \boldsymbol{\theta}_0\} \to_d N(0, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{\Omega}))$, *where* $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{\Omega}) = \boldsymbol{\Gamma}^{-1}(\alpha\mathbf{I}_{p+q} + \bar{\alpha}\boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}_C(\alpha\mathbf{I}_{p+q} + \bar{\alpha}\boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}^{-1}$.

It can be shown through simple matrix algebra that $\boldsymbol{\Omega}_{\mathrm{opt}} \equiv \bar{\alpha}\boldsymbol{\Gamma}(\bar{\alpha}\boldsymbol{\Gamma}_v + \alpha\boldsymbol{\Gamma}_r)^{-1}$ is the optimal weight matrix in that $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{\Omega}) - \boldsymbol{\Gamma}_{\boldsymbol{\theta}}(\boldsymbol{\Omega}_{\mathrm{opt}})$ is positive semidefinite for all $\boldsymbol{\Omega}$. One can estimate

$\boldsymbol{\Omega}_{\text{opt}}$ in the same manner as in the case of ω in section 2. Analogous to theorem 3, the corrected Breslow estimator is again asymptotically normal and uniformly consistent.

4. Generalized error model with multiple covariates

It may not be always appropriate to use the classical error model, which assumes that the measured surrogate is unbiased for the underlying true covariate. In this section, we extend the results of sections 2–3 to allow a more general measurement error model. All the notation is defined in the same way as in the previous section unless otherwise indicated. Under the generalized error model, the lth $(l = 1, \ldots, p)$ surrogate is a linear function of the corresponding true covariate plus a random noise: $W_{il} = a_{0l} + b_{0l} Z_{il} + \tilde{\epsilon}_{il}$, where a_{0l} and $b_{0l} \neq 0$ are constants, and $\tilde{\epsilon}_{il}$ $(i = 1, \ldots, n)$ are i.i.d. zero-mean random variables, which are independent of the true covariate, other covariates as well as the failure and censoring times. Note that $(W_{il} - a_{0l})/b_{0l} = Z_{il} + \epsilon_{il}$, where $\epsilon_{il} = \tilde{\epsilon}_{il}/b_{0l}$. Let $\mathbf{a}_0 = (a_{01}, \ldots, a_{0p})$, $\mathbf{b}_0 = (b_{01}, \ldots, b_{0p})$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^{\mathrm{T}}$. Again, we do not assume any specific distributions for $\mathbf{Z}_i$ and $\boldsymbol{\epsilon}_i$. For any vector $\mathbf{a}$, let diag($\mathbf{a}$) be the square matrix with $\mathbf{a}$ as its diagonal and 0s elsewhere. For any p-vectors $\mathbf{a}$ and $\mathbf{b}$, define $\tilde{\mathbf{W}}_i(\mathbf{a}, \mathbf{b}) = \text{diag}(\mathbf{b})^{-1}(\mathbf{W}_i - \mathbf{a})$ and $\hat{\mathbf{H}}_i(t, \mathbf{a}, \mathbf{b}) = (\tilde{\mathbf{W}}_i(\mathbf{a}, \mathbf{b})^{\mathrm{T}}, \mathbf{X}_i(t)^{\mathrm{T}})^{\mathrm{T}}$ $(i = 1, \ldots, n)$. Let $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}) = \sum_{i=1}^n \xi_i \exp[\boldsymbol{\beta}^{\mathrm{T}}\{\tilde{\mathbf{W}}_i(\mathbf{a}, \mathbf{b}) - \mathbf{Z}_i\}]\{\tilde{\mathbf{W}}_i(\mathbf{a}, \mathbf{b}) - \mathbf{Z}_i\}^{\otimes k} / \sum_{i=1}^n \xi_i$ $(k = 0, 1, 2)$. Also let

$$\hat{R}_i^{(0)}(\boldsymbol{\theta}, t, \mathbf{a}, \mathbf{b}) = \xi_i \exp(\boldsymbol{\theta}^{\mathrm{T}} \mathbf{H}_i(t)) + \bar{\xi}_i \hat{\eta}_0^{-1}(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}) \exp(\boldsymbol{\theta}^{\mathrm{T}} \hat{\mathbf{H}}_i(t, \mathbf{a}, \mathbf{b})),$$

$$\hat{\mathbf{R}}_i^{(1)}(\boldsymbol{\theta}, t, \mathbf{a}, \mathbf{b}) = \xi_i \exp(\boldsymbol{\theta}^{\mathrm{T}} \mathbf{H}_i(t)) \mathbf{H}_i(t) + \bar{\xi}_i \hat{\eta}_0^{-1}(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}) \exp(\boldsymbol{\theta}^{\mathrm{T}} \hat{\mathbf{H}}_i(t, \mathbf{a}, \mathbf{b})) \times \left\{ \hat{\mathbf{H}}_i(t, \mathbf{a}, \mathbf{b}) - \begin{bmatrix} \hat{\eta}_0^{-1}(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}) \hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b}) \\ \mathbf{0}_q \end{bmatrix} \right\}.$$

We then define $\mathbf{S}_C^{(k)}(\boldsymbol{\theta}, t, \Omega, \mathbf{a}, \mathbf{b}) = n^{-1} \sum_{j=1}^n Y_j(t) \mathbf{A}_j \hat{\mathbf{R}}_j^{(k)}(\boldsymbol{\theta}, t, \mathbf{a}, \mathbf{b})$ $(k = 0, 1)$, where $\mathbf{A}_i = \xi_i \mathbf{I}_{p+q} + \bar{\xi}_i \boldsymbol{\Omega}$. Finally, let $\mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b}) = \{\mathbf{S}_C^{(0)}(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b})\}^{-1} \mathbf{S}_C^{(1)}(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b})$, and

$$\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b}) = \sum_{i=1}^n \mathbf{A}_i \int_0^\tau \{\xi_i \mathbf{H}_i(t) + \bar{\xi}_i \hat{\mathbf{H}}_i(t, \mathbf{a}, \mathbf{b}) - \mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \mathbf{a}, \mathbf{b})\} dN_i(t).$$

If $\mathbf{a}_0$ and $\mathbf{b}_0$ are known, we are back to the classical error model discussed in section 3, since $\tilde{\mathbf{W}}_i(\mathbf{a}_0, \mathbf{b}_0) = \mathbf{Z}_i + \boldsymbol{\epsilon}_i$. Using $\tilde{\mathbf{W}}_i(\mathbf{a}_0, \mathbf{b}_0)$ as the surrogate and applying the method previously developed, we end up with an estimating function which is exactly $\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}, \mathbf{a}_0, \mathbf{b}_0)$. This can be easily seen from the definitions of $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}, \mathbf{a}, \mathbf{b})$ and $\hat{\mathbf{R}}_i^{(k)}(\boldsymbol{\theta}, t, \mathbf{a}, \mathbf{b})$ $(k = 0, 1)$.

Although in general $\mathbf{a}_0$ and $\mathbf{b}_0$ are unknown, we can estimate individual components a_{0l} and b_{0l} $(l = 1, \ldots, p)$ consistently by the least-squares estimators $\hat{a}_l$ and $\hat{b}_l$ based on the validation set. Let $\hat{\mathbf{a}} = (\hat{a}_1, \ldots, \hat{a}_p)^{\mathrm{T}}$ and $\hat{\mathbf{b}} = (\hat{b}_1, \ldots, \hat{b}_p)^{\mathrm{T}}$. The proposed estimator $\hat{\boldsymbol{\theta}}_C \equiv \hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega})$ is the solution to $\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = \mathbf{0}$. We then estimate $\Lambda_0(t)$ by

$$\hat{\Lambda}_C(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s) \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \hat{\mathbf{a}}, \hat{\mathbf{b}})}.$$

Let $\boldsymbol{\Gamma}_p$ be the first p columns of $\boldsymbol{\Gamma}$. Also, let $\boldsymbol{\mu}_Z = \text{diag}\,(\text{E}\,\mathbf{Z}_1)$, $\boldsymbol{\sigma}_Z^2 = (\text{var}(Z_{11}), \ldots, \text{var}(Z_{1p}))^{\mathrm{T}}$, $\boldsymbol{\Gamma}_{Zd} = \text{diag}\,(\boldsymbol{\sigma}_Z^2)$, $\boldsymbol{\Gamma}_\epsilon = \text{var}(\boldsymbol{\epsilon}_1)$, $\mathbf{B}_0 = \text{diag}(\mathbf{b}_0)$, and $\mathbf{B}_1 = \text{diag}\,(\boldsymbol{\beta}_0)$. Define $\boldsymbol{\Gamma}_{Z\epsilon}$ to be the $p \times p$ matrix with the (i,j)th element $\text{cov}(Z_{1i}, Z_{1j})\text{cov}(\epsilon_{1i}, \epsilon_{1j})$. Write $\mathbf{W}_i^* = \tilde{\mathbf{W}}_i(\mathbf{a}_0, \mathbf{b}_0)$ and $\hat{\mathbf{H}}_i^*(t) = (\mathbf{W}_i^{*\mathrm{T}}, \mathbf{X}_i^{\mathrm{T}}(t))^{\mathrm{T}}$. We then define $\boldsymbol{\Gamma}_v^*$ like $\boldsymbol{\Gamma}_v$ in the previous section, but with $\hat{\mathbf{H}}_1(t)$ replaced by $\hat{\mathbf{H}}_1^*(t)$. Also, let $\boldsymbol{\Gamma}_q = \bar{\alpha}^2 \alpha^{-2} (\varphi^2 \mathbf{J} \mathbf{B}_0^{-1} \boldsymbol{\Gamma}_\epsilon \mathbf{B}_0^{-1} \mathbf{J}^{\mathrm{T}} + \boldsymbol{\Gamma}_p \mathbf{B}_0^{-1} \mathbf{B}_1 \boldsymbol{\Gamma}_{Zd}^{-1} \boldsymbol{\Gamma}_{Z\epsilon} \boldsymbol{\Gamma}_{Zd}^{-1} \mathbf{B}_1 \mathbf{B}_0^{-1} \boldsymbol{\Gamma}_p^{\mathrm{T}})$,

$\boldsymbol{\Gamma}_{qr} = \bar{\alpha}^2\alpha^{-2}\varphi^2\eta_0^{-2}(\boldsymbol{\beta}_0)\mathbf{J}\mathbf{B}_0^{-1}\{\boldsymbol{\eta}_1^{\otimes 2}(\boldsymbol{\beta}_0) - \eta_0(\boldsymbol{\beta}_0)\boldsymbol{\eta}_2(\boldsymbol{\beta}_0)\}\mathbf{J}^{\mathrm{T}}$, and $\boldsymbol{\Gamma}_{rq} = \boldsymbol{\Gamma}_{qr}^{\mathrm{T}}$, where $\mathbf{J} = (\mathbf{I}_{p\times p}, \mathbf{0}_{p\times q})^{\mathrm{T}}$. The asymptotic distribution of $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0, \boldsymbol{\Omega}, \mathbf{a}_0, \mathbf{b}_0)$ can be obtained from theorem 4. Then a Taylor series expansion of $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$ around $(\mathbf{a}_0, \mathbf{b}_0)$ yields the following theorem.

Theorem 6

Under conditions 1, 2′, 3′, 4, 5, 6 and 7 given in the appendix, $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) \to_d N(\mathbf{0}, \boldsymbol{\Gamma}_C^*)$, *where* $\boldsymbol{\Gamma}_C^* = \alpha\boldsymbol{\Gamma} + \boldsymbol{\Omega}(\bar{\alpha}\boldsymbol{\Gamma}_v^* + \alpha\boldsymbol{\Gamma}_r + \alpha\boldsymbol{\Gamma}_{rq} + \alpha\boldsymbol{\Gamma}_{qr} + \alpha\boldsymbol{\Gamma}_q)\boldsymbol{\Omega}^{\mathrm{T}}$. *Furthermore,* $n^{1/2}\{\hat{\boldsymbol{\theta}}_C(\boldsymbol{\Omega}) - \boldsymbol{\theta}_0\} \to_d N(0, \boldsymbol{\Gamma}_{\boldsymbol{\theta}}^*)$, *where* $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^* = \boldsymbol{\Gamma}^{-1}(\alpha\mathbf{I}_{p+q} + \bar{\alpha}\boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}_C^*(\alpha\mathbf{I}_{p+q} + \bar{\alpha}\boldsymbol{\Omega})^{-1}\boldsymbol{\Gamma}^{-1}$.

The proof of this theorem and a consistent estimator for $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}^*$ are given in the appendix. The optimal weighting matrix $\boldsymbol{\Omega}_{\mathrm{opt}}^* \equiv \bar{\alpha}\boldsymbol{\Gamma}(\bar{\alpha}\boldsymbol{\Gamma}_v^* + \alpha\boldsymbol{\Gamma}_r + \alpha\boldsymbol{\Gamma}_{rq} + \alpha\boldsymbol{\Gamma}_{qr} + \alpha\boldsymbol{\Gamma}_q)^{-1}$ can also be estimated consistently.

The following result for the corrected Breslow estimator is analogous to theorem 3 in section 2:

Theorem 7

Under the same conditions of theorem 6, $n^{1/2}\{\hat{\Lambda}_C(t) - \Lambda_0(t)\}$ *converges weakly to a zero-mean Gaussian process. Furthermore,* $\sup_{t\in[0,\tau]}|\hat{\Lambda}_C(t) - \Lambda_0(t)| \to_p 0$.

The proof is similar to that of theorem 3 and is outlined in the appendix.

5. Simulation studies

Extensive simulation studies were carried out to investigate the performance of the proposed estimators in practical situations. The baseline hazard was set to be 1, and two covariates Z and X were generated from a bivariate normal distribution with $\mathrm{var}(Z) = \mathrm{var}(X) = 1$ and $\mathrm{cov}(Z, X) = 0.5$. The surrogate W of Z was generated from the classical error model $W = Z + \epsilon$, where ϵ is mean-zero normal with standard deviation $\sigma_\epsilon = 0.2, 0.5$ or 1; X was supposed to be measured precisely. The regression coefficient β_0 of Z varied among 0.2, 0.5 and 1, while the coefficient γ_0 of X was fixed at 0.5. Censoring times were generated from the uniform distribution on $[0, \tau]$, where τ was chosen to yield a censoring rate of approximately 30%, 50% or 70%. For a total sample size of 500, three sizes of the validation set were considered: $\alpha = 0.1, 0.2, 0.5$. For each combination of the simulation parameters, 10 000 data sets were generated, and for each data set five estimators of $\boldsymbol{\theta}_0 = (\beta_0, \gamma_0)^{\mathrm{T}}$ were calculated: the full-data estimator, which uses the true value of Z for all subjects; the complete-case estimator, which uses only the validation set; the first naïve estimator, which uses W instead of Z for all subjects; the second naïve estimator, which uses W for subjects in the non-validation set and Z for subjects in the validation set; and the proposed adaptive estimator described in section 4. For the proposed estimator, we assumed the generalized error model. To calculate the proposed estimator, the optimal weight is obtained by the one-step computation, and the initial $\boldsymbol{\Omega}$ was set to be the diagonal matrix with 0.5 on its diagonal.

Table 1 summarizes the results for β_0 when $\beta_0 = 0.5$ and the censoring rate is approximately 30% or 70%. Clearly, the two naïve estimators are biased towards 0, and the bias increases with the standard deviation of the measurement error. The proposed estimator corrects this bias very well. The standard error of the proposed estimator is always smaller than that of the complete-case estimator, and is not much larger than that of the full-data estimator when the error is small ($\sigma_\epsilon = 0.2$). The Newton–Raphson algorithm broke down in 40 to 50 out of

Table 1. *Simulation results under the model* $d\Lambda(t) = \exp(\beta_0 Z + \gamma_0 X)dt$ *with* $\beta_0 = \gamma_0 = 0.5$*, and censoring rates* 30% or 70%

α	σ_ϵ	Estimator	30% censored Mean	SE	SEE	CP	70% censored Mean	SE	SEE	CP
*	*	Full data	0.502	0.066	0.066	0.951	0.503	0.098	0.098	0.952
0.1	*	CC	0.526	0.244	0.229	0.946	0.544	0.386	0.354	0.949
	0.2	Naïve I	0.475	0.064	0.064	0.925	0.477	0.095	0.095	0.944
		Naïve II	0.477	0.064	0.064	0.931	0.479	0.095	0.095	0.946
		Proposed	0.503	0.070	0.069	0.949	0.504	0.102	0.101	0.950
	0.5	Naïve I	0.368	0.057	0.056	0.342	0.373	0.084	0.084	0.658
		Naïve II	0.378	0.057	0.057	0.423	0.383	0.085	0.085	0.705
		Proposed	0.502	0.094	0.093	0.947	0.504	0.127	0.125	0.951
	1.0	Naïve I	0.204	0.042	0.042	0.000	0.210	0.063	0.063	0.006
		Naïve II	0.217	0.044	0.043	0.000	0.223	0.065	0.065	0.015
		Proposed	0.476	0.160	0.147	0.900	0.484	0.213	0.193	0.915
0.2	*	CC	0.512	0.158	0.154	0.944	0.519	0.241	0.232	0.946
	0.2	Naïve I	0.475	0.064	0.064	0.925	0.477	0.095	0.095	0.944
		Naïve II	0.480	0.064	0.064	0.935	0.482	0.095	0.096	0.950
		Proposed	0.504	0.069	0.068	0.946	0.505	0.101	0.100	0.950
	0.5	Naïve I	0.368	0.057	0.056	0.342	0.373	0.084	0.084	0.658
		Naïve II	0.389	0.058	0.058	0.502	0.393	0.086	0.086	0.753
		Proposed	0.505	0.084	0.083	0.949	0.507	0.119	0.117	0.950
	1.0	Naïve I	0.204	0.042	0.042	0.000	0.210	0.063	0.063	0.006
		Naïve II	0.231	0.045	0.044	0.000	0.237	0.067	0.067	0.032
		Proposed	0.490	0.120	0.116	0.929	0.495	0.165	0.159	0.939
0.5	*	CC	0.505	0.095	0.094	0.949	0.508	0.141	0.140	0.949
	0.2	Naïve I	0.475	0.064	0.064	0.925	0.477	0.095	0.095	0.944
		Naïve II	0.488	0.065	0.065	0.944	0.490	0.096	0.096	0.951
		Proposed	0.504	0.068	0.067	0.949	0.506	0.100	0.099	0.951
	0.5	Naïve I	0.368	0.057	0.056	0.342	0.373	0.084	0.084	0.658
		Naïve II	0.424	0.061	0.060	0.748	0.428	0.090	0.090	0.869
		Proposed	0.505	0.074	0.073	0.948	0.507	0.108	0.107	0.949
	1.0	Naïve I	0.204	0.042	0.042	0.000	0.210	0.063	0.063	0.006
		Naïve II	0.289	0.052	0.049	0.020	0.294	0.075	0.074	0.215
		Proposed	0.502	0.085	0.084	0.948	0.504	0.123	0.122	0.948

Note: Mean and SE are the mean and standard error of the estimator, SEE is the mean of the standard error estimator, and CP is the coverage probability of the 95% confidence interval.
When $\sigma_\epsilon = 1$ and censoring rate is 30% the proposed method broke down 44 times in the 10 000 replicates for $\alpha = 0.1$ and 12 times for $\alpha = 0.2$. When $\sigma_\epsilon = 1$ and 70% censored there were 46 breakdowns for $\alpha = 0.1$, 11 for $\alpha = 0.2$, and 1 for $\alpha = 0.5$. There was 1 when $\sigma_\epsilon = 0.5$, 30% censored and $\alpha = 0.1$. The breakdown cases were excluded from the calculations of the summary statistics.
* Estimator does not depend on this parameter.

the 10 000 replicates under the extreme condition where the error is large ($\sigma_\epsilon = 1$) and the size of validation set is small ($\alpha = 0.1$). There were much less or no breakdowns in all the other settings. The estimated standard error of the proposed estimator is on average quite close to the true standard error, and the coverage probability of the 95% confidence interval is also very satisfactory, except for the extreme combination of $\sigma_\epsilon = 1$ and $\alpha = 0.1$. In addition, the simulation studies showed that the naïve estimators for γ_0 are also biased, though the magnitude of this bias is in general smaller than that of the naïve estimators for β_0. The proposed method corrects this bias as well.

When $\beta_0 = 0.2$, the results are similar, though the proposed estimator is less likely to break down, and the coverage probabilities are excellent even in extreme settings. When $\beta = 1.0$, the proposed estimator performs a little bit worse. Additional simulation studies revealed similar pictures for $n = 200$ and $\alpha = 0.2$ or 0.5, and for uniform rather than normal Z. Simulations with $\beta_0 = -0.5$ also showed that the naïve approach biases the estimator of β_0 towards 0 and the bias is well corrected by the proposed method.

6. Classical error model with replicate covariate measurements

In many applications, only replicate measurements are available for the covariate subject to measurement error. Here, we show how to extend the proposed methods to this important and realistic setting. It is infeasible to estimate the bias in the surrogate based on replicates alone. Thus, the classical error model is required in this section. Suppose that, for the ith subject, a p-vector of time-independent covariates $\mathbf{Z}_i$ is measured with error and a q-vector of possibly time-dependent covariates $\mathbf{X}_i(t)$ is measured precisely. For conciseness, we assume that $\mathbf{Z}_i$ is only measured twice by the surrogates $\mathbf{W}_{i1} \equiv \mathbf{Z}_i + \boldsymbol{\epsilon}_{i1}$ and $\mathbf{W}_{i2} \equiv \mathbf{Z}_i + \boldsymbol{\epsilon}_{i2}$, where the vector-valued error terms $\boldsymbol{\epsilon}_{i,m}$ $(i = 1, \dots, n; m = 1, 2)$ are i.i.d. with mean $\mathbf{0}_p$ and are independent of all other random variables. We make an additional assumption that the error distribution is symmetric, i.e., $\boldsymbol{\epsilon}_{11}$ and $-\boldsymbol{\epsilon}_{11}$ have the same distribution.

Let $\mathbf{H}_i(t) = (\mathbf{Z}_i^{\mathrm{T}}, \mathbf{X}_i(t)^{\mathrm{T}})^{\mathrm{T}}$ and $\hat{\mathbf{H}}_i(t) = (\hat{\mathbf{W}}_i^{\mathrm{T}}, \mathbf{X}_i(t)^{\mathrm{T}})^{\mathrm{T}}$, where $\hat{\mathbf{W}}_i = (\mathbf{W}_{i1} + \mathbf{W}_{i2})/2$. We assume the same Cox model as in section 3. Write $\boldsymbol{\eta}_k(\boldsymbol{\beta}) = E(\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\epsilon}_{11})\boldsymbol{\epsilon}_{11}^{\otimes k})$ $(k = 0, 1, 2)$. Note that

$$E(\exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t))|F_\tau) = \eta_0^2(\boldsymbol{\beta}/2)\exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t)),$$

$$E(\exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t))\hat{\mathbf{H}}_i(t)|F_\tau) = \eta_0^2(\boldsymbol{\beta}/2)\exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t))\mathbf{H}_i(t) + \exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t))\begin{bmatrix}\eta_0(\boldsymbol{\beta}/2)\boldsymbol{\eta}_1(\boldsymbol{\beta}/2)\\ \mathbf{0}_q\end{bmatrix}.$$

Define

$$R_i^{(0)}(\boldsymbol{\theta}, t) = \eta_0^{-2}(\boldsymbol{\beta}/2)\exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t)),$$

$$\mathbf{R}_i^{(1)}(\boldsymbol{\theta}, t) = \eta_0^{-2}(\boldsymbol{\beta}/2)\exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t))\hat{\mathbf{H}}_i(t) - \exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t))\begin{bmatrix}\eta_0^{-3}(\boldsymbol{\beta}/2)\boldsymbol{\eta}_1(\boldsymbol{\beta}/2)\\ \mathbf{0}_q\end{bmatrix}.$$

It is easy to see that $E\{n^{-1}\sum_{i=1}^n Y_i(t)\mathbf{R}_i^{(k)}(\boldsymbol{\theta}, t)|\mathscr{F}_\tau\} = \mathbf{S}^{(k)}(\boldsymbol{\theta}, t)$ $(k = 0, 1)$, where $S^{(k)}(\boldsymbol{\theta}, t)$ are defined in section 3. By the symmetry of the error distribution, $E\{\exp(\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2}))\} = \eta_0^2(\boldsymbol{\beta})$ and $E\{(\mathbf{W}_{i1} - \mathbf{W}_{i2})\exp(\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2}))\} = 2\eta_0(\boldsymbol{\beta})\boldsymbol{\eta}_1(\boldsymbol{\beta})$. Thus, $\eta_0(\boldsymbol{\beta})$ and $\boldsymbol{\eta}_1(\boldsymbol{\beta})$ can be estimated by $\hat{\eta}_0(\boldsymbol{\beta}) = \{n^{-1}\sum_{i=1}^n \exp(\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2}))\}^{1/2}$ and $\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta}) = \{2n\,\hat{\eta}_0(\boldsymbol{\beta})\}^{-1}\sum_{i=1}^n(\mathbf{W}_{i1} - \mathbf{W}_{i2})\exp(\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2}))$. For $k = 0, 1$, we define $\hat{\mathbf{R}}_i^{(k)}(\boldsymbol{\theta}, t)$ in the same way as $\mathbf{R}_i^{(k)}(\boldsymbol{\theta}, t)$, but with $\boldsymbol{\eta}_l(\boldsymbol{\beta}/2)$ $(l = 0, 1)$ replaced by $\hat{\boldsymbol{\eta}}_l(\boldsymbol{\beta}/2)$. Let $\mathbf{S}_C^{(k)}(\boldsymbol{\theta}, t) = n^{-1}\sum_{j=1}^n Y_j(t)\hat{\mathbf{R}}_j^{(k)}(\boldsymbol{\theta}, t)$ $(k = 0, 1)$, and $\mathbf{E}_C(\boldsymbol{\theta}, t) = \mathbf{S}_C^{(1)}(\boldsymbol{\theta}, t)/S_C^{(0)}(\boldsymbol{\theta}, t)$. We then propose the estimating function

$$\mathbf{U}_C(\boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\boldsymbol{\theta}, t)\}dN_i(t).$$

Denote the resultant estimator by $\hat{\boldsymbol{\theta}}_C$. The corresponding estimator of $\Lambda_0(t)$ is

$$\hat{\Lambda}_C(t) \equiv \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{\sum_{j=1}^n Y_j(s)\hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s)}.$$

Let

$$\mathbf{u}_i = \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{e}(\boldsymbol{\theta}_0, t)\}\left\{dN_i(t) - R_i^{(0)}(\boldsymbol{\theta}_0, t)Y_i(t)d\Lambda_0(t)\right\} + \int_0^\tau \exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)\,d\Lambda_0(t)\;\eta_0^{-3}(\boldsymbol{\beta}_0/2)\begin{bmatrix}\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix},$$

$$\mathbf{r}_i = \eta_0^{-2}(\boldsymbol{\beta}_0/2)\varphi\exp(\boldsymbol{\beta}_0^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2})/2)\left\{2^{-1}\begin{bmatrix}\mathbf{W}_{i1} - \mathbf{W}_{i2}\\ \mathbf{0}_q\end{bmatrix} - \eta_0^{-1}(\boldsymbol{\beta}_0/2)\begin{bmatrix}\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix}\right\},$$

where $\mathbf{e}(\boldsymbol{\theta}, t)$ and φ are defined in section 3. Also, let $\boldsymbol{\Gamma}_u = E(\mathbf{u}_1^{\otimes 2})$, $\boldsymbol{\Gamma}_{ur} = E(\mathbf{u}_1 \otimes \mathbf{r}_1)$, $\boldsymbol{\Gamma}_{ru} = E(\mathbf{r}_1 \otimes \mathbf{u}_1)$ and $\boldsymbol{\Gamma}_r = E(\mathbf{r}_1^{\otimes 2})$. We then have the following result for $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0)$:

Theorem 8
Under conditions 1, 2″, 3″, 4, 5, 6 and 8 given in the appendix, $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0) \to_d N(\mathbf{0}_{p+q}, \boldsymbol{\Gamma}_C)$, *where* $\boldsymbol{\Gamma}_C = \boldsymbol{\Gamma}_u + \boldsymbol{\Gamma}_{ur} + \boldsymbol{\Gamma}_{ru} + \boldsymbol{\Gamma}_r$.

The asymptotic properties of $\hat{\boldsymbol{\theta}}_C$ are established in the following theorem:

Theorem 9
Under conditions 1, 2″, 3″, 4, 5, 6 and 8, $\hat{\boldsymbol{\theta}}_C$ *exists and is unique in a neighbourhood of* $\boldsymbol{\theta}_0$ *with probability converging to one as* $n \to \infty$, *and* $\hat{\boldsymbol{\theta}}_C \to_p \boldsymbol{\theta}_0$. *Furthermore,* $n^{1/2}\{\hat{\boldsymbol{\theta}}_C - \boldsymbol{\theta}_0\} \to_d N(0, \boldsymbol{\Gamma}_{\boldsymbol{\theta}})$, *where* $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Gamma}_C\boldsymbol{\Gamma}^{-1}$ *and* $\boldsymbol{\Gamma}$ *is defined in section 3.*

Remark 1. A consistent estimator of $\boldsymbol{\Gamma}_C$ is $\hat{\boldsymbol{\Gamma}}_C \equiv n^{-1}\sum_{i=1}^n(\hat{\mathbf{u}}_i + \hat{\mathbf{r}}_i)^{\otimes 2}$, where $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{r}}_i$ are obtained from $\mathbf{u}_i$ and $\mathbf{r}_i$ by replacing $\boldsymbol{\theta}_0$, $\mathbf{e}(\boldsymbol{\theta}_0, t)$, $R_i^{(0)}(\boldsymbol{\theta}_0, t)$, $\Lambda_0(t)$ and $\boldsymbol{\eta}_k(\boldsymbol{\beta}_0/2)$ $(k = 0, 1)$ with $\hat{\boldsymbol{\theta}}_C$, $\mathbf{E}_C(\hat{\boldsymbol{\theta}}_C, t)$, $\hat{R}_i^{(0)}(\hat{\boldsymbol{\theta}}_C, t)$, $\hat{\Lambda}_C(t)$ and $\hat{\boldsymbol{\eta}}_k(\hat{\boldsymbol{\beta}}_C/2)$. In addition, a consistent estimator of $\boldsymbol{\Gamma}$ is $\hat{\boldsymbol{\Gamma}} \equiv n^{-1}\sum_{i=1}^n\int_0^\tau\{\mathbf{S}_C^{(2)}(\hat{\boldsymbol{\theta}}_C, t)/\mathbf{S}_C^{(0)}(\hat{\boldsymbol{\theta}}_C, t) - \mathbf{E}_C^{\otimes 2}(\hat{\boldsymbol{\theta}}_C, t)\}\,dN_i(t)$, where $\mathbf{S}_C^{(2)}(\boldsymbol{\theta}, t) = \partial\mathbf{S}_C^{(1)}(\boldsymbol{\theta}, t)/\partial\boldsymbol{\theta}$. Thus, $\hat{\boldsymbol{\Gamma}}_{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\Gamma}}^{-1}\hat{\boldsymbol{\Gamma}}_C\hat{\boldsymbol{\Gamma}}^{-1}$ consistently estimates $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$.

Remark 2. As in section 2, $\hat{\Lambda}_C(\cdot)$ can be shown to be asymptotically normal and uniformly consistent with a covariance function which can be consistently estimated.

7. Discussion

The approximately corrected score estimator by Nakamura (1992) was developed under the restrictive assumptions of the classical error model and a normally distributed error with known covariance matrix. Recently, Kong & Gu (1999) proved the consistency and asymptotic normality of this estimator and provided a modification for non-normal error. When the error $\boldsymbol{\epsilon}$ is normally distributed with a known covariance matrix, $\boldsymbol{\eta}_k(\boldsymbol{\beta})$ $(k = 0, 1)$ defined in section 3 can be written out in terms of $\boldsymbol{\beta}$ and the covariance matrix of $\boldsymbol{\epsilon}$. Then replacing $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta})$ by $\boldsymbol{\eta}_k(\boldsymbol{\beta})$ $(k = 0, 1)$ and disregarding the validation set and weight, the estimators proposed in our section 3 would reduce to Nakamura's estimator. In addition, assuming a known moment generating function for $\boldsymbol{\epsilon}$, replacing $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta})$ by $\boldsymbol{\eta}_k(\boldsymbol{\beta})$ $(k = 0, 1)$ and disregarding the validation set and weight, our estimators would reduce to the modification suggested by Kong & Gu (1999).

Nakamura (1992) and Kong & Gu (1999) assume that the covariance matrix or the moment generating function of the error is known. We have removed these assumptions by making use of a validation set or replicate measurements, which are two general schemes to study measurement error problems (Carroll *et al.*, 1995, p. 12). We have developed formal inference procedures for these two practical settings. Kong & Gu (1999) only mentioned in passing that the error parameters might be estimated from a validation set or replicates, and did not account for the extra variability due to such estimation in their variance estimators.

Nakamura (1992) and Kong & Gu (1999) confined their attention to the classical additive measurement error model requiring the surrogate be unbiased for the true covariate. The methods developed in section 4 pertain to a more general error model which allows biased surrogates. This generalization is essential in many real applications. Another new feature of this work is the clear distinction between covariates that are measured with error and those without error.

Under the classical error model, Buzas (1998) provided a method to correct the partial likelihood score using the moment generating function of the error distribution, which is assumed to be known. In correcting the partial likelihood score function, Buzas proposed to replace $S^{(0)}(\boldsymbol{\theta}, t)$ with $n^{-1}\sum_{i=1}^{n} Y_i(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\hat{\mathbf{Z}}_i + \boldsymbol{\gamma}^{\mathrm{T}}\mathbf{X}_i)$, where $\hat{\mathbf{Z}}_i$ is the predicted value of $\mathbf{Z}_i$ based on the regression of $\mathbf{W}_i$ on $\mathbf{X}_i$, and if $\mathbf{X}_i$ does not exist, i.e. no covariate is measured precisely, Buzas proposed to use just $n^{-1}\sum_{j=1}^{n} Y_j(t)$ in place of $S^{(0)}(\boldsymbol{\theta}, t)$. The efficiency of this estimator depends on the existence of $\mathbf{X}_i$, and on the relation of $\mathbf{Z}_i$ and $\mathbf{X}_i$ if $\mathbf{X}_i$ does exist. The techniques developed in this paper can potentially be used to study the asymptotic properties of the estimator suggested by Buzas.

Acknowledgements

This research was supported by the National Institutes of Health grants CA82659, GM47845 and CA70227. The authors are grateful to the reviewers for their helpful comments.

References

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100–1120.

Breslow, N. E. (1972). Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.

Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plann. Inference* **67**, 247–257.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman & Hall, London.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187–220.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *J. Amer. Statist. Assoc.* **72**, 147–148.

Fuller, W. A. (1987). *Measurement error models*. Wiley, New York.

Hu, P., Tsiatis, A.A. & Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **54**, 1407–1419.

Kong, F. H. & Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statist. Sinica* **9**, 953–969.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711–730.

Nakamura, T. (1990). Corrected score functions for errors-in-variables models: methodology and application to generalized linear models. *Biometrika* **77**, 127–137.

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* **48**, 829–838.

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **69**, 331–342.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Statist. Ser. A* **18**, 4335–4358.
van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer-Verlag, New York.
Wang, C. Y., Hsu, L., Feng, Z. D. & Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics* **53**, 131–145.
Zhou, H. & Pepe, M. S. (1995). Auxiliary covariate data in failure time regression. *Biometrika* **82**, 139–149.

Received January 2000, in final form August 2001

D. Y. Lin, University of North Carolina, Biostatistics Department, 3101E McGavran-Greenberg Hall, CB #7420, Chapel Hill, NC 27599-7420, USA.
E-mail: lin@bios.unc.edu

Appendix

Regularity conditions. For a vector $\mathbf{a} = (a_1, \ldots, a_p)^{\mathrm{T}}$, let $|\mathbf{a}| = (\sum_{i=1}^{p} a_i^2)^{1/2}$. In the case of a single covariate, the following conditions are needed.

1. $\Lambda_0(t)$ is continuous and $\Lambda_0(\tau) < \infty$.
2. $E(|Z_1|^2) < \infty$, $E(|\epsilon_1|^2) < \infty$.
3. There exists a compact neighbourhood $\mathscr{B}$ of β_0 such that:
   (a) $E\{\sup_{\beta \in \mathscr{B}} |Z_1|^2 \exp(\beta Z_1)\} < \infty$; (b) $E\{\sup_{\beta \in \mathscr{B}} |\epsilon_1|^2 \exp(\beta \epsilon_1)\} < \infty$;
   (c) $E\{\sup_{\beta \in \mathscr{B}} |Z_1|^2 \exp(2\beta Z_1)\} < \infty$; (d) $E\{\sup_{\beta \in \mathscr{B}} |\epsilon_1|^2 \exp(2\beta \epsilon_1)\} < \infty$.
4. $P\{Y_1(\tau) = 1\} > 0$.
5. Γ is positive definite.

For the case of multiple covariates in section 3, we need to restate conditions 2 and 3:

2′. $E\{\sup_{t \in [0,\tau]} |\mathbf{H}_1(t)|^2\} < \infty, E(|\boldsymbol{\epsilon}_1|^2) < \infty$.
3′. There exists a compact neighbourhood $\mathscr{B}$ of $\boldsymbol{\beta}_0$ and a compact neighbourhood $\mathscr{C}$ of γ_0 such that:
   (a) $E\{\sup_{t \in [0,\tau], \boldsymbol{\theta} \in \mathscr{B} \times \mathscr{C}} |\mathbf{H}_1(t)|^2 \exp(\boldsymbol{\theta}^{\mathrm{T}} \mathbf{H}_1(t))\} < \infty$;
   (b) $E\{\sup_{\boldsymbol{\beta} \in \mathscr{B}} |\boldsymbol{\epsilon}_1|^2 \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\epsilon}_1)\} < \infty$;
   (c) $E\{\sup_{t \in [0,\tau], \boldsymbol{\theta} \in \mathscr{B} \times \mathscr{C}} |\mathbf{H}_1(t)|^2 \exp(2\boldsymbol{\theta}^{\mathrm{T}} \mathbf{H}_1(t))\} < \infty$;
   (d) $E\{\sup_{\boldsymbol{\beta} \in \mathscr{B}} |\boldsymbol{\epsilon}_1|^2 \exp(2\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\epsilon}_1)\} < \infty$.

We also need a condition for time-dependent covariates:

6. For all sample paths of $\mathbf{X}_i(\cdot)$, $|X_{ji}(0)| + \int_0^\tau |dX_{ji}(t)| < \infty$, where $X_{ji}(\cdot)$ is the jth component of $\mathbf{X}_i(\cdot)$.

Besides the above conditions, we need one more for the generalized error model:

7. There exists a compact neighbourhood $\mathscr{B}_0$ of $\mathbf{0}$ such that: $E\Big\{\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathscr{B}_0} |\mathbf{Z}_1|^2 \exp(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_1)\} < \infty$.

For the situation of replicate covariate measurements studied in section 7, we replace $\boldsymbol{\epsilon}_1$ in conditions 2′ and 3′ by $\boldsymbol{\epsilon}_{11}$, and also modify part (d) of condition 3′ to be $E\{\sup_{\boldsymbol{\beta} \in \mathscr{B}} |\boldsymbol{\epsilon}_{11}|^2 \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\epsilon}_{11}/2)\} < \infty$. We label them as conditions 2″ and 3″. An additional condition is required:

8. The distribution of $\boldsymbol{\epsilon}_{11}$ is symmetric with respect to $\mathbf{0}$, i.e. $\boldsymbol{\epsilon}_{11}$ and $-\boldsymbol{\epsilon}_{11}$ are identically distributed.

Proof of theorem 1. We decompose $U_C(\beta_0, \omega)$ into two terms:

$$U_C(\beta_0,\omega) = \sum_{i=1}^{n}(\xi_i + \omega\bar{\xi}_i)\int_0^\tau \{\xi_i Z_i + \bar{\xi}_i W_i - E_C(\beta_0,t,\omega)\}\Big\{dN_i(t) - \hat{R}_i^{(0)}(\beta_0)Y_i(t)d\Lambda_0(t)\Big\}$$
$$+ \sum_{i=1}^{n}(\xi_i + \omega\bar{\xi}_i)\int_0^\tau \{\xi_i Z_i + \bar{\xi}_i W_i - E_C(\beta_0,t,\omega)\}\hat{R}_i^{(0)}(\beta_0)Y_i(t)d\Lambda_0(t). \quad (8)$$

We further split the first term into $B_1 + B_2 + B_3 + B_4$, where

$$B_1 = \sum_{i=1}^{n}(\xi_i + \omega\bar{\xi}_i)\int_0^\tau \{\xi_i Z_i + \bar{\xi}_i W_i - e(\beta_0,t)\}\Big\{dN_i(t) - R_i^{(0)}(\beta_0)Y_i(t)d\Lambda_0(t)\Big\},$$

$$B_2 = \sum_{i=1}^{n}(\xi_i + \omega\bar{\xi}_i)\int_0^\tau \{e(\beta_0,t) - E_C(\beta_0,t,\omega)\}\{dN_i(t) - \exp(\beta_0 Z_i)Y_i(t)d\Lambda_0(t)\},$$

$$B_3 = \omega\sum_{i=1}^{n}\bar{\xi}_i\int_0^\tau \{e(\beta_0,t) - E_C(\beta_0,t,\omega)\}\{\exp(\beta_0 Z_i) - \eta_0^{-1}(\beta_0)\exp(\beta_0 W_i)\}Y_i(t)d\Lambda_0(t),$$

$$B_4 = \omega\{\eta_0^{-1}(\beta_0) - \hat{\eta}_0^{-1}(\beta_0)\}\sum_{i=1}^{n}\bar{\xi}_i\int_0^\tau \{W_i - E_C(\beta_0,t,\omega)\}\exp(\beta_0 W_i)Y_i(t)\,d\Lambda_0(t).$$

It follows from the arguments of Andersen & Gill (1982) that $\sup_{0\le t\le\tau,\beta\in\mathscr{B}} |E_C(\beta,t,\omega) - e(\beta,t)| \to_p 0$. By expanding $\mathscr{F}_t$ to include the selection of the validation set and the surrogate covariates, we see that $n^{-1/2}B_2$ is a martingale integral with variance converging to 0. Thus $n^{-1/2}B_2 \to_p 0$.

Let $J_n(t) = n^{-1/2}\sum_{i=1}^n \bar{\xi}_i\{\exp(\beta_0 Z_i) - \eta_0^{-1}(\beta_0)\exp(\beta_0 W_i)\}\int_0^t Y_i(s)\,d\Lambda_0(s)$. Then $n^{-1/2}B_3 = \omega\int_0^\tau\{e(\beta_0,t) - E_C(\beta_0,t,\omega)\}dJ_n(t)$. In light of conditions 3(c) and 3(d) and the fact that $J_n(t)$ is the difference of two non-decreasing processes, $J_n(t)$ converges weakly to a process $J(t)$ with continuous sample paths by ex. 2.11.16 of van der Vaart & Wellner (1996, p. 215). By the strong embedding theorem (Shorack & Wellner, 1986, pp. 47–48), there exists a new probability space in which $\sup_{t\in[0,\tau]}|S_C^{(k)}(\beta_0,t,\omega) - (\alpha+\omega\bar{\alpha})s^{(k)}(\beta_0,t)| \to_{\text{a.s.}} 0$ $(k = 0, 1)$ and $J_n(t) \to_{\text{a.s}} J(t)$. Since $S_C^{(0)}(\beta_0,t,\omega)$ is monotone in t and left-continuous, lem. 1 of Lin *et al.* (2000) implies that $\int_0^t S_C^{(0)}(\beta_0,t,\omega)^{-1}dJ_n(s) \to_{\text{a.s.}} \int_0^t\{(\alpha+\omega\bar{\alpha})s^{(0)}(\beta_0,t)\}^{-1}dJ(s)$. Another use of this lemma yields $\int_0^\tau E_C(\beta_0,t,\omega)dJ_n(s) \to_{\text{a.s.}} \int_0^\tau e(\beta_0,t)dJ(s)$. Likewise, $\int_0^\tau e(\beta_0,t)dJ_n(s) \to_{\text{a.s.}} \int_0^\tau e(\beta_0,t)dJ(s)$. Hence $n^{-1/2}B_3 \to_{\text{a.s.}} 0$. Returning to the original probability space, we have $n^{-1/2}B_3 \to_p 0$.

Note that

$$n^{-1/2}B_4 = \omega n^{1/2}\eta_0^{-2}(\beta_0)\{\hat{\eta}_0(\beta_0) - \eta_0(\beta_0)\}\bar{\alpha}$$
$$\times \underline{E}\left\{\int_0^\tau\{W_1 - e(\beta_0,t)\}\exp(\beta_0 W_1)Y_1(t)d\Lambda_0(t)\right\} + o_p(1).$$

By (3) and (4), we have $n^{-1/2}B_4 = \omega(\bar{\alpha}/\alpha)\varphi\quad \eta_0^{-2}(\beta_0)\eta_1(\beta_0)n^{-1/2}\sum_{i=1}^n \xi_i\{\exp(\beta_0\epsilon_i) - \eta_0(\beta_0)\} + o_p(1)$.

By simple algebraic calculations, the second term of (8) is equal to $\omega\hat{\eta}_0^{-2}(\beta_0)\hat{\eta}_1(\beta_0)\sum_{i=1}^n \bar{\xi}_i \int_0^\tau \exp(\beta_0 W_i)Y_i(t)d\Lambda_0(t)$. We can split this into $B_5 + B_6$, where

$$B_5 = \omega\eta_0^{-2}(\beta_0)\eta_1(\beta_0)\sum_{i=1}^{n}\bar{\xi}_i\int_0^\tau \exp(\beta_0 W_i)Y_i(t)d\Lambda_0(t),$$

$$B_6 = \omega\{\hat{\eta}_0^{-2}(\beta_0)\hat{\eta}_1(\beta_0) - \eta_0^{-2}(\beta_0)\eta_1(\beta_0)\}\sum_{i=1}^{n}\bar{\xi}_i\int_0^\tau \exp(\beta_0 W_i)Y_i(t)\,d\Lambda_0(t).$$

It is easy to see that $B_1 + B_5 = \sum_{i=1}^n(\xi_i u_i + \omega\bar{\xi}_i v_i)$, where

$$u_i = \int_0^\tau \{Z_i - e(\beta_0, t)\} dM_i(t),$$

$$v_i = \int_0^\tau \{W_i - e(\beta_0, t)\}\{dN_i(t) - \eta_0^{-1}(\beta_0)\exp(\beta_0 W_i) Y_i(t)\, d\Lambda_0(t)\}$$

$$+ \eta_0^{-2}(\beta_0)\eta_1(\beta_0)\int_0^\tau \exp(\beta_0 W_i) Y_i(t)\, d\Lambda_0(t),$$

and $M_i(t) = N_i(t) - \int_0^t Y_i(s)\exp(\beta_0 Z_i)\, d\Lambda_0(s)$. Because u_i is a martingale integral for each i, $E(u_i) = 0$. It follows from (3) and (4) that $E(v_i|\mathscr{F}_\tau) = u_i$, which implies that $E(v_i) = 0$.

For B_6, we note that $n^{-1}\sum_{i=1}^n \bar{\xi}_i \int_0^\tau \exp(\beta W_i) Y_i(t)\, d\Lambda_0(t) \rightarrow_{\text{a.s.}} \bar{\alpha}\varphi\eta_0(\beta_0)$. Hence

$$n^{-1/2} B_6 = \omega(\bar{\alpha}/\alpha)\varphi\eta_0^{-2}(\beta_0) n^{-1/2} \sum_{i=1}^n \xi_i [\eta_0(\beta_0)\{\exp(\beta_0\epsilon_i)\epsilon_i - \eta_1(\beta_0)\}$$

$$- 2\eta_1(\beta_0)\{\exp(\beta_0\epsilon_i) - \eta_0(\beta_0)\}] + o_p(1).$$

It then follows that $n^{-1/2}(B_4 + B_6) = \omega n^{-1/2}\sum_{i=1}^n \xi_i r_i + o_p(1)$, where

$$r_i = (\bar{\alpha}/\alpha)\varphi\eta_0^{-2}(\beta_0)\{\eta_0(\beta_0)\exp(\beta_0\epsilon_i)\epsilon_i - \eta_1(\beta_0)\exp(\beta_0\epsilon_i)\}.$$

Summarizing the above results, we have

$$n^{-1/2} U_C(\beta_0, \omega) = n^{-1/2}\sum_{i=1}^n \left(\xi_i u_i + \omega\xi_i r_i + \omega\bar{\xi}_i v_i\right) + o_p(1), \tag{9}$$

which is essentially a normalized sum of zero-mean i.i.d. random variables. By assumption, ϵ_i is independent of $\{Y_i(t), N_i(t), Z_i\}$, which implies that u_i and r_i are independent. Clearly, $\text{cov}(\xi_i u_i, \bar{\xi}_i v_i) = 0$ and $\text{cov}(\xi_i r_i, \bar{\xi}_i v_i) = 0$. It is easy to see that $\text{var}(u_1) = \Gamma$, $\text{var}(v_1) = \Gamma_v$ and $\text{var}(r_1) = \Gamma_r$, so that $\text{var}(\xi_1 u_1 + \omega\xi_1 r_1 + \omega\bar{\xi}_1 v_1) = \Gamma_C$. It then follows from the central limit theorem that $n^{-1/2} U_C(\beta_0, \omega) \rightarrow_d N(0, \Gamma_C)$.

Proof of theorem 2. Let $D(\beta, \omega) = -n^{-1}\partial U_C(\beta, \omega)/\partial\beta$. Then

$$D(\beta, \omega) = n^{-1}\sum_{i=1}^n (\xi_i + \omega\bar{\xi}_i)\int_0^\tau \partial E_C(\beta, t, \omega)/\partial\beta dN_i(t),$$

where $\partial E_C(\beta, t, \omega)/\partial\beta = S_C^{(2)}(\beta, t, \omega)/S_C^{(0)}(\beta, t, \omega) - S_C^{(1)}(\beta, t, \omega)^2/S_C^{(0)}(\beta, t, \omega)^2$ and $S_C^{(2)}(\beta, t, \omega) = \partial S_C^{(1)}(\beta, t, \omega)/\partial\beta$. It follows from condition 4 and th. III.1 of Andersen & Gill (1982) that

$$\sup_{t\in[0,\tau], \beta\in\mathscr{B}} |\partial E_C(\beta, t, \omega)/\partial\beta - v(\beta, t)| \rightarrow_p 0. \tag{10}$$

Hence

$$D(\beta, \omega) \rightarrow_p (\alpha + \omega\bar{\alpha})\int_0^\tau v(\beta, t) s^{(0)}(\beta, t) d\Lambda_0(t) = (\alpha + \omega\bar{\alpha})\Gamma. \tag{11}$$

By th. III.1 of Andersen & Gill (1982), condition 1 and (10), we see that the convergence in (11) is uniform for $\beta \in \mathscr{B}$, i.e.

$$\sup_{\beta\in\mathscr{B}} |D(\beta, \omega) - (\alpha + \omega\bar{\alpha})\Gamma| \rightarrow_p 0. \tag{12}$$

For any positive ω, the limit is non-negative everywhere and positive at β_0 by condition 5. Theorem 1 implies that $n^{-1} U_C(\beta_0, \omega) \rightarrow_p 0$ for any ω. It then follows from the proof of theorem 2 of Foutz (1977) that $\hat{\beta}_C$ exists and is unique in $\mathscr{B}$ with probability converging to one as $n \rightarrow \infty$, and $\hat{\beta}_C \rightarrow_p \beta_0$.

The Taylor series expansion yields $n^{1/2}(\hat\beta_C-\beta_0)=D^{-1}(\beta^*,\omega)n^{-1/2}U_C(\beta_0,\omega)$, where β^* is on the line segment between $\hat\beta_C$ and β_0. By (12) and the consistency of $\hat\beta_C$, we have $D(\beta^*,\omega)\to_p(\alpha+\omega\bar\alpha)\Gamma$. Then the theorem follows from the asymptotic normality of $n^{-1/2}U_C(\beta_0,\omega)$.

Estimation of variance Γ_β. Based on (12), $(\alpha+\omega\bar\alpha)\Gamma$ can be estimated consistently by $-n^{-1}\partial U_C(\hat\beta_C,\omega)/\partial\beta$. We can estimate Γ by

$$\hat\Gamma=(\hat\alpha+\omega\hat{\bar\alpha})^{-1}n^{-1}\sum_{i=1}^n(\xi_i+\omega\bar\xi_i)\int_0^\tau\left\{S_C^{(2)}(\hat\beta_C,t,\omega)/S_C^{(0)}(\hat\beta_C,t,\omega)-E_C(\hat\beta_C,t,\omega)^2\right\}dN_i(t).$$

It is not difficult to see that φ can be consistently estimated by $\hat\varphi=n^{-1}\sum_{i=1}^n\delta_i$. Thus, we can estimate Γ_r by $\hat\Gamma_r\equiv(\hat{\bar\alpha}/\hat\alpha)^2\hat\varphi^2\hat\eta_0^{-4}(\hat\beta_C)\{\hat\eta_0^2(\hat\beta_C)\hat\eta_2(2\hat\beta_C)-2\hat\eta_0(\hat\beta_C)\hat\eta_1(\hat\beta_C)\hat\eta_1(2\hat\beta_C)+\hat\eta_1^2(\hat\beta_C)\times\hat\eta_0(2\hat\beta_C)\}$, which is also consistent. In addition, we can estimate Γ_v by $\hat\Gamma_v\equiv(n\hat{\bar\alpha})^{-1}\sum_{i=1}^n\bar\xi_i\hat v_i^2$, where $\hat v_i=\int_0^\tau\{W_i-E_c(\hat\beta_C,t,\omega)\}\{dN_i(t)-\hat\eta_0^{-1}(\hat\beta_C)\exp(\hat\beta_CW_i)\quad Y_i(t)d\hat\Lambda_C(t)\}+\hat\eta_0^{-2}(\hat\beta_C)\hat\eta_1(\hat\beta_C)\int_0^\tau\exp(\hat\beta_CW_i)Y_i(t)d\hat\Lambda_C(t)$ $(i=1,\dots,n)$. The consistency of $\hat\Gamma_v$ follows from the law of large numbers, together with the consistency of $\hat\beta_C$ and the uniform consistency of$\hat\Lambda_C(\cdot)$ given in theorem 3. Thus, Γ_β is consistently estimated by $\hat\Gamma_\beta\equiv\hat\Gamma_C/\{(\hat\alpha+\omega\hat{\bar\alpha})\hat\Gamma\}^2$, where $\hat\Gamma_C=\hat\alpha\hat\Gamma+\omega^2(\hat{\bar\alpha}\hat\Gamma_v+\hat\alpha\hat\Gamma_r)$.

Proof of theorem 3. We can write $n^{1/2}\{\hat\Lambda_C(t)-\Lambda_0(t)\}=D_1(t)+D_2(t)+D_3(t)+o_p(1)$, where

$$D_1(t)=n^{1/2}\sum_{i=1}^n\int_0^t\left\{\frac{1}{\sum_{j=1}^nY_j(s)\hat R_j^{(0)}(\hat\beta_C)}-\frac{1}{\sum_{j=1}^nY_j(s)\hat R_j^{(0)}(\beta_0)}\right\}dN_i(s),$$

$$D_2(t)=n^{1/2}\sum_{i=1}^n\int_0^t\left\{\frac{1}{\sum_{j=1}^nY_j(s)\hat R_j^{(0)}(\beta_0)}-\frac{1}{\sum_{j=1}^nY_j(s)R_j^{(0)}(\beta_0)}\right\}dN_i(s),$$

$$D_3(t)=n^{1/2}\left\{\sum_{i=1}^n\int_0^t\frac{dN_i(s)}{\sum_{j=1}^nY_j(s)R_j^{(0)}(\beta_0)}-\Lambda_0^*(t)\right\},$$

and $\Lambda_0^*(t)=\int_0^tI\{\sum_{i=1}^nY_i(s)>0\}d\Lambda_0(s)$. By Taylor expansions,

$$D(t)=h(t)n^{-1/2}\sum_{i=1}^n\{\xi_iu_i+\omega(\bar\xi_iv_i+\xi_ir_i)\}+o_p(1),$$

where $h(t)=-\int_0^te(\beta_0,s)d\Lambda_0(s)/\{(\alpha+\omega\bar\alpha)\Gamma\}$. Clearly,

$$D_2(t)=(\bar\alpha/\alpha)\eta_0^{-1}(\beta_0)\Lambda_0(t)n^{-1/2}\sum_{i=1}^n\xi_i\{\exp(\beta_0\epsilon_i)-\eta_0(\beta_0)\}+o_p(1),$$

and

$$D_3(t)=n^{-1/2}\sum_{i=1}^n\left\{\xi_i\int_0^t\frac{dM_i(s)}{s^{(0)}(\beta_0,s)}+\bar\xi_i\int_0^t\frac{dN_i(s)-Y_i(s)\eta_0^{-1}(\beta_0)\exp(\beta_0W_i)d\Lambda_0(s)}{s^{(0)}(\beta_0,s)}\right\}+o_p(1).$$

Summarizing the above results for $D_k(t)$ $(k=1,2,3)$, we get

$$n^{1/2}\{\hat{\Lambda}_C(t)-\Lambda_0(t)\}=n^{-1/2}\sum_{i=1}^{n}\{\xi_i\tilde{u}_i(t)+\bar{\xi}_i\tilde{v}_i(t)+\xi_i\tilde{r}_i(t)\}+o_p(1),$$

where

$$\tilde{u}_i(t)=h(t)u_i+\int_0^t\frac{dM_i(s)}{s^{(0)}(\beta_0,s)},$$
$$\tilde{v}_i(t)=\omega h(t)v_i+\int_0^t\frac{dN_i(s)-Y_i(s)\eta_0^{-1}(\beta_0)\exp(\beta_0W_i)d\Lambda_0(s)}{s^{(0)}(\beta_0,s)},$$
$$\tilde{r}_i(t)=\omega h(t)r_i+(\bar{\alpha}/\alpha)\eta_0^{-1}(\beta_0)\Lambda_0(t)\{\exp(\beta_0\epsilon_i)-\eta_0(\beta_0)\}.$$

It is easy to see that the means of $\tilde{u}_i$, $\tilde{v}_i$ and $\tilde{r}_i$ are all 0, and $\tilde{u}_i$ and $\tilde{r}_i$ are independent for every i. Furthermore, $\mathrm{cov}(\tilde{u}_i,\tilde{v}_i)=0$ and $\mathrm{cov}(\tilde{r}_i,\tilde{v}_i)=0$. Let $\phi_u(t,s)=E\{\tilde{u}_1(t)\tilde{u}_1(s)\}$, $\phi_v(t,s)=E\{\tilde{v}_1(t)\tilde{v}_1(s)\}$ and $\phi_r(t,s)=E\{\tilde{r}_1(t)\tilde{r}_1(s)\}$. Write

$$\phi(t,s)=\alpha\phi_u(t,s)+\bar{\alpha}\phi_v(t,s)+\alpha\phi_r(t,s). \tag{13}$$

It can be shown that $\tilde{u}_i(t)$, $\tilde{v}_i(t)$ and $\tilde{r}_i(t)$ are sums of monotone functions in t. It then follows from ex. 2.11.16 of van der Vaart & Wellner (1996) that $n^{1/2}\{\hat{\Lambda}_C(t)-\Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\phi(t,s)$.

Using the above decomposition for $\hat{\Lambda}_C(t)-\Lambda_0(t)$, we can show the uniform consistency of $\hat{\Lambda}_C(\cdot)$.

Proof of theorem 4. As in the proof of theorem 1, we can show that

$$n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega})=n^{-1/2}\sum_{i=1}^{n}\left(\xi_i\mathbf{u}_i+\xi_i\boldsymbol{\Omega}\mathbf{r}_i+\bar{\xi}_i\boldsymbol{\Omega}\mathbf{v}_i\right)+o_p(1), \tag{14}$$

where

$$\mathbf{u}_i=\int_0^\tau\{\mathbf{H}_i(t)-\mathbf{e}(\boldsymbol{\theta}_0,t)\}dM_i(t),$$
$$\mathbf{v}_i=\int_0^\tau\{\hat{\mathbf{H}}_i(t)-\mathbf{e}(\boldsymbol{\theta}_0,t)\}\{dN_i(t)-\eta_0^{-1}(\boldsymbol{\beta}_0)\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)d\Lambda_0(t)\}$$
$$+\eta_0^{-2}(\boldsymbol{\beta}_0)\int_0^\tau\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)d\Lambda_0(t)\begin{bmatrix}\boldsymbol{\eta}_1(\boldsymbol{\beta}_0)\\ \mathbf{0}_q\end{bmatrix}, \tag{15}$$
$$\mathbf{r}_i=(\bar{\alpha}/\alpha)\varphi\eta_0^{-2}(\boldsymbol{\beta}_0)\exp(\boldsymbol{\beta}_0^{\mathrm{T}}\epsilon_i)\begin{bmatrix}\eta_0(\boldsymbol{\beta}_0)\boldsymbol{\epsilon}_i-\boldsymbol{\eta}_1(\boldsymbol{\beta}_0)\\ \mathbf{0}_q\end{bmatrix}.$$

Here $M_i(t)=N_i(t)-\int_0^t Y_i(s)\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{H}_i(s))d\Lambda_0(s)$.

Proof of theorem 6. Since the asymptotic normality of $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}_0,\mathbf{b}_0)$ has already been proved in theorem 4, we expand $n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\hat{\mathbf{a}},\hat{\mathbf{b}})$ around $(\mathbf{a}_0,\mathbf{b}_0)$:

$$n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\hat{\mathbf{a}},\hat{\mathbf{b}})=n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}_0,\mathbf{b}_0)+n^{-1}\partial\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}^*,\mathbf{b}^*)/\partial\mathbf{a}\,n^{1/2}(\hat{\mathbf{a}}-\mathbf{a}_0)$$
$$+n^{-1}\partial\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}^*,\mathbf{b}^*)/\partial\mathbf{b}\,n^{1/2}(\hat{\mathbf{b}}-\mathbf{b}_0),$$

where $(\mathbf{a}^{*\mathrm{T}},\mathbf{b}^{*\mathrm{T}})^{\mathrm{T}}$ is on the line segment between $(\hat{\mathbf{a}}^{\mathrm{T}},\hat{\mathbf{b}}^{\mathrm{T}})^{\mathrm{T}}$ and $(\mathbf{a}_0^{\mathrm{T}},\mathbf{b}_0^{\mathrm{T}})^{\mathrm{T}}$. Since $\hat{\mathbf{a}}\rightarrow_p\mathbf{a}_0$ and $\hat{\mathbf{b}}\rightarrow_p\mathbf{b}_0$, we have $\mathbf{a}^*\rightarrow_p\mathbf{a}_0$ and $\mathbf{b}^*\rightarrow_p\mathbf{b}_0$. Through some tedious but straightforward calculations,

$$n^{-1}\partial\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}^*,\mathbf{b}^*)/\partial\mathbf{a}\rightarrow_p-\bar{\alpha}\varphi\boldsymbol{\Omega}\mathbf{J}\mathbf{B}_0^{-1},$$
$$n^{-1}\partial\mathbf{U}_C(\boldsymbol{\theta}_0,\boldsymbol{\Omega},\mathbf{a}^*,\mathbf{b}^*)/\partial\mathbf{b}\rightarrow_p\bar{\alpha}\boldsymbol{\Omega}(\boldsymbol{\Gamma}_p\mathbf{B}_0^{-1}\mathbf{B}_1-\varphi\mathbf{J}\mathbf{B}_0^{-1}\boldsymbol{\mu}_Z).$$

Condition 7 is needed to guarantee the consistency of $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$ for $\boldsymbol{\eta}_k(\boldsymbol{\beta})$ $(k = 0, 1, 2)$. Let $\mu_{Zl} = E(Z_{1l})$ and $\sigma^2_{Zl} = \text{var}(Z_{1l})(l = 1, \ldots, p)$. By the properties of the least-squares estimators,

$$n^{1/2}(\hat{a}_l - a_{0l}) = \frac{1}{\alpha\sigma^2_{Zl}} n^{-1/2} \sum_{i=1}^{n} \xi_i \{(\sigma^2_{Zl} + \mu^2_{Zl})\epsilon_{il} - \mu_{Zl} Z_{il}\epsilon_{il}\} + o_p(1), \tag{16}$$

$$n^{1/2}(\hat{b}_l - b_{0l}) = \frac{1}{\alpha\sigma^2_{Zl}} n^{-1/2} \sum_{i=1}^{n} \xi_i(-\mu_{Zl}\epsilon_{il} + Z_{il}\epsilon_{il}) + o_p(1). \tag{17}$$

It follows from theorem 4 that

$$n^{-1/2}\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) = n^{-1/2} \sum_{i=1}^{n} \left(\xi_i \mathbf{u}_i + \xi_i \boldsymbol{\Omega}\mathbf{r}_i + \xi_i \boldsymbol{\Omega}\mathbf{q}_i + \bar{\xi}_i \boldsymbol{\Omega}\mathbf{v}^*_i\right) + o_p(1), \tag{18}$$

where $\mathbf{q}_i = (\bar{\alpha}/\alpha)\left[-\varphi \mathbf{J}\mathbf{B}_0^{-1}\boldsymbol{\epsilon}_i + \boldsymbol{\Gamma}_p \mathbf{B}_0^{-1}\mathbf{B}_1 \boldsymbol{\Gamma}^{-1}_{Zd}\{\text{diag}(\mathbf{Z}_i) - \boldsymbol{\mu}_Z\}\boldsymbol{\epsilon}_i\right]$, and $\mathbf{u}_i$, $\mathbf{v}^*_i$ and $\mathbf{r}_i$ are defined as in (15), but with $\hat{\mathbf{H}}_i(t)$ replaced by $\hat{\mathbf{H}}^*_i(t)$ in $\mathbf{v}^*_i$. We see that $\boldsymbol{\Gamma}_{rq} = \text{cov}(\mathbf{r}_1, \mathbf{q}_1)$, $\boldsymbol{\Gamma}_{qr} = \text{cov}(\mathbf{q}_1, \mathbf{r}_1)$ and $\boldsymbol{\Gamma}q = \text{var}(\mathbf{q}_1)$, while $\mathbf{q}_1$ and $\mathbf{u}_1$ are uncorrelated. The theorem readily follows.

Estimation of $\boldsymbol{\Gamma}^*_{\boldsymbol{\theta}}$. For $k = 0, 1, 2$, let $\hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}) = \hat{\boldsymbol{\eta}}_k(\boldsymbol{\beta}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$, and for $k = 0, 1$, let $\hat{\mathbf{R}}^{(k)}_i(\boldsymbol{\theta}, t) = \hat{\mathbf{R}}^{(k)}_i(\boldsymbol{\theta}, t, \hat{\mathbf{a}}, \hat{\mathbf{b}})$ and $\mathbf{S}^{(k)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) = \mathbf{S}^{(k)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$. Also, let $\mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) = \mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}})$ and $\hat{\mathbf{H}}_i(t) = \hat{\mathbf{H}}(t, \hat{\mathbf{a}}, \hat{\mathbf{b}})$. We then define $\hat{\mathbf{R}}^{(2)}_i(\boldsymbol{\theta}, t) = \partial\hat{\mathbf{R}}^{(1)}_i(\boldsymbol{\theta}, t)/\partial\boldsymbol{\theta}$. It is easy to see that

$$\hat{\mathbf{R}}^{(2)}_i(\boldsymbol{\theta}, t) = \xi_i \exp(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{H}_i(t))\mathbf{H}^{\otimes 2}_i(t) + \bar{\xi}_i \hat{\eta}_0^{-1}(\boldsymbol{\beta}) \exp(\boldsymbol{\theta}^{\mathrm{T}}\hat{\mathbf{H}}_i(t)) \Bigg[\hat{\mathbf{H}}^{\otimes 2}_i(t) - \hat{\eta}_0^{-1}(\boldsymbol{\beta})\mathbf{J}\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta})\hat{\mathbf{H}}_i(t)^{\mathrm{T}}$$
$$- \hat{\eta}_0^{-1}(\boldsymbol{\beta})\hat{\mathbf{H}}_i(t)\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta})^{\mathrm{T}}\mathbf{J}^{\mathrm{T}} + 2\hat{\eta}_0^{-2}(\boldsymbol{\beta})\mathbf{J}\hat{\boldsymbol{\eta}}^{\otimes 2}_1(\boldsymbol{\beta})\mathbf{J}^{\mathrm{T}} - \hat{\eta}_0^{-1}(\boldsymbol{\beta})\mathbf{J}\hat{\boldsymbol{\eta}}_2(\boldsymbol{\beta})\mathbf{J}^{\mathrm{T}}\Bigg].$$

Let $\mathbf{S}^{(2)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) = n^{-1}\sum_{j=1}^{n} Y_j(t)\mathbf{A}_j\hat{\mathbf{R}}^{(2)}_j(\boldsymbol{\theta}, t)$. We can estimate $\boldsymbol{\Gamma}$ consistently by $\hat{\boldsymbol{\Gamma}} = (\hat{\alpha}\mathbf{I}_{p+q} + \hat{\bar{\alpha}}\boldsymbol{\Omega})^{-1} n^{-1}\partial\mathbf{U}_C(\hat{\boldsymbol{\theta}}_C, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}})/\partial\boldsymbol{\theta}$, where

$$\frac{\partial\mathbf{U}_C(\boldsymbol{\theta}, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}})}{\partial\boldsymbol{\theta}}$$
$$= \sum_{i=1}^{n} \mathbf{A}_i \int_0^{\tau} \left[\left\{\mathbf{S}^{(0)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega})\right\}^{-1}\left\{\mathbf{S}^{(2)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}) - \mathbf{E}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega}, \hat{\mathbf{a}}, \hat{\mathbf{b}}) \otimes \mathbf{S}^{(1)}_C(\boldsymbol{\theta}, t, \boldsymbol{\Omega})\right\}\right] dN_i(t).$$

Also, we can estimate $\boldsymbol{\Gamma}^*_v$ consistently by $\hat{\boldsymbol{\Gamma}}^*_v = n^{-1}\sum_{i=1}^{n} \hat{\mathbf{v}}^{*\otimes 2}_i$, where

$$\hat{\mathbf{v}}^*_i = \int_0^{\tau} \{\hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\hat{\boldsymbol{\theta}}_C, t, \boldsymbol{\Omega})\}[dN_i(t) - \hat{\eta}_0^{-1}(\hat{\boldsymbol{\beta}}_C)\exp(\hat{\boldsymbol{\theta}}^{\mathrm{T}}_C\hat{\mathbf{H}}_i(t))Y_i(t)d\hat{\Lambda}_C(t)]$$
$$+ \hat{\eta}_0^{-2}(\hat{\boldsymbol{\beta}}_C)\int_0^{\tau} \exp(\hat{\boldsymbol{\theta}}^{\mathrm{T}}_C\hat{\mathbf{H}}_i(t))Y_i(t)d\hat{\Lambda}_C(t)\mathbf{J}\hat{\boldsymbol{\eta}}_1(\hat{\boldsymbol{\beta}}_C).$$

A consistent estimator for $\boldsymbol{\Gamma}_r$ is $\hat{\boldsymbol{\Gamma}}_r \equiv (\hat{\bar{\alpha}}/\hat{\alpha})^2\varphi^2\hat{\eta}_0^{-4}(\hat{\boldsymbol{\beta}}_C)\mathbf{J}\{\hat{\eta}_0(2\boldsymbol{\beta}_C)\hat{\boldsymbol{\eta}}^{\otimes 2}_1(\hat{\boldsymbol{\beta}}_C) - \hat{\eta}_0(\hat{\boldsymbol{\beta}}_C)\hat{\boldsymbol{\eta}}_1(\hat{\boldsymbol{\beta}}_C) \otimes \hat{\boldsymbol{\eta}}_1(2\hat{\boldsymbol{\beta}}_C) - \hat{\eta}_0(\hat{\boldsymbol{\beta}}_C)\hat{\boldsymbol{\eta}}_1(2\hat{\boldsymbol{\beta}}_C) \otimes \hat{\boldsymbol{\eta}}_1(\hat{\boldsymbol{\beta}}_C) + \hat{\eta}^2_0(\hat{\boldsymbol{\beta}}_C)\hat{\boldsymbol{\eta}}_2(2\hat{\boldsymbol{\beta}}_C)\}\mathbf{J}^{\mathrm{T}}$. Let $\hat{\boldsymbol{\Gamma}}_\epsilon = \sum_{i=1}^{n} \xi_i\{\hat{\mathbf{B}}_0^{-1}(\mathbf{W}_i - \hat{\mathbf{a}}) - \mathbf{Z}_i\}^{\otimes 2}/(\sum_{i=1}^{n} \xi_i - 2)$ where $\hat{\mathbf{B}}_0 = \text{diag}(\hat{\mathbf{b}})$. Let $\hat{\sigma}^2_{Zl}$ be the sample variance of all the Z_{il}s of subjects in the validation set, $\hat{\boldsymbol{\sigma}}^2_Z = (\hat{\sigma}^2_{Z1}, \ldots, \hat{\sigma}^2_{Zp})^{\mathrm{T}}$, and $\hat{\boldsymbol{\Gamma}}_{Zd} = \text{diag}(\hat{\boldsymbol{\sigma}}^2_Z)$. We also define $\hat{\boldsymbol{\Gamma}}_{Z\epsilon}$ to be the $p \times p$ matrix with estimated $\text{cov}(Z_{1i}, Z_{1j})\text{cov}(\epsilon_{1i}, \epsilon_{1j})$ as the (i, j)th element, and let $\hat{\boldsymbol{\Gamma}}_p$ be the first p columns of $\hat{\boldsymbol{\Gamma}}$. By plugging all these estimators into the expressions defining $\boldsymbol{\Gamma}_q$, $\boldsymbol{\Gamma}_{rq}$ and $\boldsymbol{\Gamma}_{qr}$, we get consistent estimators $\hat{\boldsymbol{\Gamma}}_q$, $\hat{\boldsymbol{\Gamma}}_{rq}$ and $\hat{\boldsymbol{\Gamma}}_{qr}$, respectively. Then we estimate $\boldsymbol{\Gamma}^*_{\boldsymbol{\theta}}$ by $\hat{\boldsymbol{\Gamma}}^*_{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\Gamma}}^{-1}(\hat{\alpha}\mathbf{I}_{p+q} + \hat{\bar{\alpha}}\boldsymbol{\Omega})^{-1}\hat{\boldsymbol{\Gamma}}^*_C(\hat{\alpha}\mathbf{I}_{p+q} + \hat{\bar{\alpha}}\boldsymbol{\Omega}\ T)^{-1}\hat{\boldsymbol{\Gamma}}^{-1}$, where $\hat{\boldsymbol{\Gamma}}^*_C = \hat{\alpha}\hat{\boldsymbol{\Gamma}} + \boldsymbol{\Omega}(\hat{\bar{\alpha}}\hat{\boldsymbol{\Gamma}}^*_v + \hat{\alpha}\hat{\boldsymbol{\Gamma}}_r + \hat{\alpha}\hat{\boldsymbol{\Gamma}}_{rq} + \hat{\alpha}\hat{\boldsymbol{\Gamma}}_{qr} + \hat{\alpha}\hat{\boldsymbol{\Gamma}}_q)\boldsymbol{\Omega}^{\mathrm{T}}$.

Proof of theorem 7. Let

$$\hat{\Lambda}_C(t, \mathbf{a}_0, \mathbf{b}_0) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s) \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \mathbf{a}_0, \mathbf{b}_0)}.$$

The Taylor expansion of $\hat{\Lambda}_C(t)$ around $(\mathbf{a}_0, \mathbf{b}_0)$ yields

$$n^{1/2}\hat{\Lambda}_C(t) = n^{1/2}\hat{\Lambda}_C(t, \mathbf{a}_0, \mathbf{b}_0) + (\mathbf{D}_n^a)^{\mathrm{T}} n^{1/2}(\hat{\mathbf{a}} - \mathbf{a}_0) + (\mathbf{D}_n^b)^{\mathrm{T}} n^{1/2}(\hat{\mathbf{b}} - \mathbf{b}_0), \tag{19}$$

where

$$\mathbf{D}_n^a = -n^{-1} \sum_{i=1}^{n} \int_0^t \frac{n^{-1} \sum_{j=1}^{n} Y_j(s) \partial \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \mathbf{a}^*, \mathbf{b}^*)/\partial \mathbf{a} \, d\, N_i(s)}{\left\{ n^{-1} \sum_{j=1}^{n} Y_j(s) \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \mathbf{a}^*, \mathbf{b}^*) \right\}^2},$$

$$\mathbf{D}_n^b = -n^{-1} \sum_{i=1}^{n} \int_0^t \frac{n^{-1} \sum_{j=1}^{n} Y_j(s) \partial \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \mathbf{a}^*, \mathbf{b}^*)/\partial \mathbf{b} \, dN_i(s)}{\left\{ n^{-1} \sum_{j=1}^{n} Y_j(s) \hat{R}_j^{(0)}(\hat{\boldsymbol{\theta}}_C, s, \mathbf{a}^*, \mathbf{b}^*) \right\}^2}.$$

By applying to $\hat{\Lambda}_C(t, \mathbf{a}_0, \mathbf{b}_0)$ a decomposition similar to that used in the proof of theorem 3, calculating the limits of $\mathbf{D}_n^a$ and $\mathbf{D}_n^b$, and plugging them, along with (16) and (17), into (19), we can approximate $n^{1/2}\{\hat{\Lambda}_C(t) - \Lambda_0(t)\}$ by a sum of n i.i.d. zero-mean terms. Then the theorem can be established in a similar fashion as theorem 3.

Proof of theorem 8. This proof is parallel to that of theorem 1. First we write

$$\begin{aligned}\mathbf{U}_C(\boldsymbol{\theta}_0) = {} & \sum_{i=1}^{n} \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\boldsymbol{\theta}_0, t)\}\{dN_i(t) - \hat{R}_i^{(0)}(\boldsymbol{\theta}_0, t) Y_i(t) d\Lambda_0(t)\} \\ & + \sum_{i=1}^{n} \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\boldsymbol{\theta}_0, t)\} \hat{R}_i^{(0)}(\boldsymbol{\theta}_0, t) Y_i(t) d\Lambda_0(t).\end{aligned} \tag{20}$$

Then we split the first term into $\mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3 + \mathbf{B}_4$, where

$$\mathbf{B}_1 = \sum_{i=1}^{n} \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{e}(\boldsymbol{\theta}_0, t)\}\{dN_i(t) - R_i^{(0)}(\boldsymbol{\theta}_0, t) Y_i(t) d\Lambda_0(t)\},$$

$$\mathbf{B}_2 = \sum_{i=1}^{n} \int_0^\tau \{\mathbf{e}(\boldsymbol{\theta}_0, t) - \mathbf{E}_C(\boldsymbol{\theta}_0, t)\}\{dN_i(t) - \exp(\boldsymbol{\theta}_0^{\mathrm{T}} \mathbf{H}_i(t)) Y_i(t) d\Lambda_0(t)\},$$

$$\mathbf{B}_3 = \sum_{i=1}^{n} \int_0^\tau \{\mathbf{e}(\boldsymbol{\theta}_0, t) - \mathbf{E}_C(\boldsymbol{\theta}_0, t)\}\{\exp(\boldsymbol{\theta}_0^{\mathrm{T}} \mathbf{H}_i(t)) - \eta_0^{-2}(\boldsymbol{\beta}_0/2) \exp(\boldsymbol{\theta}_0^{\mathrm{T}} \hat{\mathbf{H}}_i(t))\} Y_i(t) d\Lambda_0(t),$$

$$\mathbf{B}_4 = \sum_{i=1}^{n} \int_0^\tau \{\hat{\mathbf{H}}_i(t) - \mathbf{E}_C(\boldsymbol{\theta}_0, t)\}\{R_i^{(0)}(\boldsymbol{\theta}_0, t) - \hat{R}_i^{(0)}(\boldsymbol{\theta}_0, t)\} Y_i(t) d\Lambda_0(t).$$

As in the proof of theorem 1, we can show that $n^{-1/2}\mathbf{B}_2 \to_p \mathbf{0}$ and $n^{-1/2}\mathbf{B}_3 \to_p \mathbf{0}$ via martingale theory and empirical process theory. We can also show that

$$\mathbf{B}_4 = \varphi \sum_{i=1}^{n} \left\{ \exp(\boldsymbol{\beta}_0^{\mathrm{T}}(\mathbf{W}_{i1} - \mathbf{W}_{i2})/2) - \eta_0^2(\boldsymbol{\beta}_0/2) \right\} \begin{bmatrix} \eta_0^{-3}(\boldsymbol{\beta}_0/2)\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2) \\ \mathbf{0}_q \end{bmatrix} + o_p(\mathbf{1}).$$

The second term on the right hand side of (20) is equal to

$$\sum_{i=1}^{n}\int_0^{\tau}\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)d\Lambda_0(t)\begin{bmatrix}\hat{\eta}_0^{-3}(\boldsymbol{\beta}_0/2)\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix},$$

which can be split further into $\mathbf{B}_5+\mathbf{B}_6$, where

$$\mathbf{B}_5=\sum_{i=1}^{n}\int_0^{\tau}\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)d\Lambda_0(t)\begin{bmatrix}\eta_0^{-3}(\boldsymbol{\beta}_0/2)\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix},$$

$$\mathbf{B}_6=\sum_{i=1}^{n}\int_0^{\tau}\exp(\boldsymbol{\theta}_0^{\mathrm{T}}\hat{\mathbf{H}}_i(t))Y_i(t)d\Lambda_0(t)\begin{bmatrix}\hat{\eta}_0^{-3}(\boldsymbol{\beta}_0/2)\hat{\boldsymbol{\eta}}_1(\boldsymbol{\beta}_0/2)-\eta_0^{-3}(\boldsymbol{\beta}_0/2)\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix}.$$

Note that $\mathbf{B}_1+\mathbf{B}_5=\sum_{i=1}^{n}\mathbf{u}_i$, and we can show that $E(\mathbf{u}_i)=\mathbf{0}$ $(i=1,\ldots,n)$. We also have

$$\mathbf{B}_6=\varphi\sum_{i=1}^{n}\left[2^{-1}\eta_0^{-2}(\boldsymbol{\beta}_0/2)\left\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}(\mathbf{W}_{i1}-\mathbf{W}_{i2})/2)\begin{bmatrix}\mathbf{W}_{i1}-\mathbf{W}_{i2}\\ \mathbf{0}_q\end{bmatrix}-2\eta_0(\boldsymbol{\beta}_0/2)\begin{bmatrix}\boldsymbol{\eta}_1(\boldsymbol{\beta}_0/2)\\ \mathbf{0}_q\end{bmatrix}\right\}\right.$$
$$\left.-2\eta_0^{-3}(\boldsymbol{\beta}_0/2)\left\{\exp(\boldsymbol{\beta}_0^{\mathrm{T}}(\mathbf{W}_{i1}-\mathbf{W}_{i2})/2)-\eta_0^2(\boldsymbol{\beta}_0/2)\right\}\begin{bmatrix}\boldsymbol{\eta}_1(\boldsymbol{\beta}_0)\\ \mathbf{0}_q\end{bmatrix}\right]+o_p(\mathbf{1}).$$

We see that $\mathbf{B}_4+\mathbf{B}_6=\sum_{i=1}^{n}\mathbf{r}_i+o_p(\mathbf{1})$. It is not hard to see that $E(\mathbf{r}_i)=\mathbf{0}$ $(i=1,\ldots,n)$. The asymptotic normality of $n^{-1/2}\mathbf{U}_C(\theta_0)$ follows easily.

Proof of theorem 9. Let $\mathbf{S}_C^{(2)}(\boldsymbol{\theta},t)=\partial\mathbf{S}_C^{(1)}(\boldsymbol{\theta},t)/\partial\boldsymbol{\theta}$. As in the proof of theorem 2, we can show that $\sup_{t\in[0,\tau],\boldsymbol{\theta}\in\mathscr{B}\times\mathscr{C}}|\mathbf{S}_C^{(2)}(\boldsymbol{\theta},t)-s^{(2)}(\boldsymbol{\theta},t)|\rightarrow_p\mathbf{0}$, $\sup_{t\in[0,\tau],\boldsymbol{\theta}\in\mathscr{B}\times\mathscr{C}}|\partial\mathbf{E}_C(\boldsymbol{\theta},t)/\partial\boldsymbol{\theta}-\mathbf{v}(\boldsymbol{\theta},t)|\rightarrow_p\mathbf{0}$, and $\sup_{t\in[0,\tau],\boldsymbol{\theta}\in\mathscr{B}\times\mathscr{C}}|-n^{-1}\partial\mathbf{U}_C(\boldsymbol{\theta})/\partial\boldsymbol{\theta}-\boldsymbol{\Gamma}|\rightarrow_p\mathbf{0}$, where $s^{(2)}(\boldsymbol{\theta},t)$, $\mathbf{v}(\boldsymbol{\theta},t)$ and $\boldsymbol{\Gamma}$ are defined in section 3. Then the existence, uniqueness and consistency of $\hat{\boldsymbol{\theta}}_C$ can be proved in the same manner as in theorem 2. Asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}}_C-\boldsymbol{\theta}_0)$ then follows from the Taylor expansion.