Biometrika (2014), **xx**, x, *pp*. 1–28

Printed in Great Britain

An Extended Hazard Model with Longitudinal Covariates

BY Y. K. TSENG,

Graduate Institute of Statistics, National Central University, No. 300, Jhongda Rd., Jhong-Li, Taoyuan County 32049, Taiwan

tsengyk@ncu.edu.tw

Y. R. SU,

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, U.S.A.

ysu@fhcrc.org

M. MAO AND J. L. WANG

Department of Statistics, University of California, Davis, California 95616, U.S.A.

mmao@ucdavis.edu janelwang@ucdavis.edu

SUMMARY

In clinical trials and other medical studies, it has become increasingly common to observe an event time of interest and longitudinal covariates simultaneously. In the literature, joint modelling approaches have been employed to analyze both survival and longitudinal processes and to investigate their association. Early attention has mostly been placed on developing adaptive and flexible longitudinal processes based on a prespecified survival model, most commonly the Cox proportional model. In this paper, we propose a general class of semi-parametric hazard regres-

sion models, termed extended hazard model, for the survival component. This class includes two popular survival models, the Cox proportional hazards model and the accelerated failure time model, as special cases. The proposed model is flexible for modelling event data, and its nested structure facilitates model selection for the survival component through likelihood ratio tests. A pseudo joint likelihood approach is proposed to estimate the unknown parameters and components through a Monte Carlo EM algorithm. Asymptotic theory for the estimators is developed together with theory for the semiparametric likelihood ratio tests. The performance of the procedure is demonstrated through simulation studies. A case study featuring data from a Taiwan HIV/AIDS cohort study further illustrates the usefulness of the extended hazard model.

Some key words: Hazard smoothing; Joint modelling; Maximum likelihood estimation; Monte Carlo EM algorithm; Semiparametric likelihood ratio test.

1. INTRODUCTION

In many medical studies, longitudinal biomarkers and the event time of interest are collected simultaneously in order to explore their association. One main interest of these studies is to detect any impact of biomarkers and treatments on the event time. The patterns of the longitudinal biomarkers are of additional interest. A well-known example is an AIDS clinical trial and cohort study where the disease marker CD4 counts and event time of patients are both recorded. The primary goal of such a study is to explore the association between longitudinal CD4 counts and the event time, and a secondary goal is to model the patterns of the CD4 trajectories for a better understanding of the longitudinal time courses; see for instance Pawitan & Self (1993), Tsiatis et al. (1995), Wulfsohn & Tsiatis (1997), Bycott & Taylor (1998) and Wang & Taylor (2001). Partial likelihood estimation (Cox, 1975) for the Cox model encounters difficulties, because it requires knowledge of the entire history of longitudinal biomarkers and does not allow the longi-

tudinal covariates to contain measurement errors. Both requirements might fail in such medical studies, thereby inducing biases. Moreover the longitudinal processes are only accessible before the occurrence of the event, such as death, which results in informative missing or dropout. Several solutions for these difficulties have been proposed in the literature; see Tsiatis & Davidian (2004), Verbeke & Davidian (2008) and Fitzmaurice et al. (2008). The most efficient solution is the maximum likelihood approach to model jointly the longitudinal and event time processes. This involves selecting a longitudinal model for the biomarkers and a disease risk model for the event time data.

In this paper, the longitudinal component is assumed to be a linear mixed effects model with measurement errors,

$$W(t) = X(t) + e(t), \tag{1}$$

$$X(t) = b^{\mathrm{T}}\rho(t), \tag{2}$$

where the covariate process $W(t)$ is observed intermittently with measurement errors or random fluctuations $e(t)$, so $X(t)$ is the actual underlying covariate process that links the disease risks. The covariate process $X(t)$ is modelled through known basis functions $\{\rho_1(t), \ldots \rho_p(t)\}^{\mathrm{T}} = \rho(t)$. The measurement error $e(t)$ is independent of $b^{\mathrm{T}} = (b_1, \ldots, b_p)$, which follows a p-dimensional multivariate normal distribution $N_p(\mu, \Sigma)$. To accommodate various mechanisms that generate such data, a number of mixed effects models (1) have been proposed in the literature. Wulfsohn & Tsiatis (1997) considered one with $\{\rho_1(t), \rho_2(t)\} = (1, t)$, while Henderson et al. (2000) and Wang & Taylor (2001) added an extra Gaussian process to explain additional variation that cannot be fully described by the random effects and the measurement errors. Tsiatis & Davidian (2001) and Song et al. (2002) also applied simple linear mixed effects models for the longitudinal process but relaxed the normality assumption, by invoking a conditional

score approach and a flexible class of parametric density functions, respectively. Another flexible model by Brown et al. (2005) sets $\rho(t)$ to be a vector of B-spline basis functions with the number of knots determined by a model selection procedure. Ding & Wang (2008) simplified this by proposing a multiplicative random effects model with a single one-dimensional random effect.

For the survival component, the Cox proportional hazards model is usually employed to describe the risk at time t, stipulating

$$\lambda\{t \mid \bar{X}(t)\} = \lambda_0(t) \exp\{\beta X(t)\}, \tag{3}$$

where $\bar{X}(t) = \{X(s) : 0 \leq s < t\}$ is the covariate history up to time t, β is the regression parameter, and $\lambda_0(t)$ is the unspecified baseline hazard rate function. To address the challenge that the proportional hazard assumption may fail, Tseng et al. (2005) proposed a joint modelling approach based on the accelerated failure time model:

$$\lambda\{t \mid \bar{X}(t)\} = \lambda_0 \left[\int_0^t \exp\{\beta X(s)\} ds\right] \exp\{\beta X(t)\}. \tag{4}$$

A question that often arises in survival analysis is whether the Cox or the accelerated failure time model fits the data better. Unfortunately, there is no good tool to evaluate their adequacy, and neither may be suitable for a given data set, in which case an intermediate model may be needed. These problems can be addressed by enlarging the two models to a more general class, which then provides the needed flexibility in modelling the survival component. Developing such a flexible procedure complements the literature on joint modelling, which has largely focused on developing adaptive longitudinal models, while a flexible survival model is equally necessary for success. Moreover, Tseng et al. (2005) proposed only the methodology and did not study the properties of their estimators. Zeng & Lin (2007) provided efficient estimation for the accelerated failure time model with time-dependent covariates and a thorough investigation on large sample

properties of estimators. However, their model setting requires the complete history of the time-dependent covariates and no measurement error. In addition, the incorporation of random effects in the joint model setting leads to a complicated joint likelihood function, which is not considered in Zeng & Lin (2007). Consequently, adaptation of their approach to the joint model of (1) and (4) is not straightforward.

Our focus is extension of the Cox and accelerated failure time models in (3) and (4), which can be made more flexible by replacing the single parameter β in (4) by potentially different parameters β_1 and β_2. This leads to the extended hazard model:

$$\lambda\{t \mid \bar{X}(t)\} = \lambda_0 \left[\int_0^t \exp\{\beta_1 X(s)\} ds\right] \exp\{\beta_2 X(t)\}. \tag{5}$$

The baseline hazard function λ_0 is left unspecified, so (5) is a semiparametric model. When $\beta_1 = 0$, (5) becomes a Cox model (3), and when $\beta_1 = \beta_2$, (5) becomes an accelerated failure time model (4). In the Cox model, only the covariate at the most recent time has an impact on the subject's specific risk, while the accelerated failure time model (4) allows the entire covariate history to influence the risk. In extended hazard models, the covariates have both instantaneous and cumulative impact on the risk through the risk factor and the baseline hazard, respectively.

Remark. Another special case of (5) is called the accelerated hazard model when $\beta_2 = 0$. The accelerated hazard model with time-independent covariates was investigated by Chen & Wang (2000). To focus on the Cox model and the accelerated failure time model, we will not further study this model in this paper.

A special case of the extended hazard model (5) with time-independent covariates has been studied by Etezadi-Amoli & Ciampi (1987), Chen & Jewell (2001), and Tseng & Shu (2011). These approaches are inefficient and cannot handle longitudinal covariates. The model takes the

form:

$$\lambda\{t \mid \bar{X}(t)\} = \lambda_0\{t \exp(\beta_1 X)\} \exp(\beta_2 X). \quad (6)$$

Since Weibull regression model falls within both Cox and accelerated failure time families, there is an identifiability issue in model (6). However, this identifiability issue is alleviated in our setting, as long as the covariate $X(t)$ is not constant over time. This is a salient feature of the extended hazard model with longitudinal covariates. Recently, Tseng et al. (2014) investigated the extended hazard model (5) through weighted score equations. However, their approach also required complete covariate history and thus is not appropriate for the joint modelling setting.

In the rest of the paper, we consider a joint model formed by combining the extended hazard model (5) and the longitudinal model (1). The model and the pseudo likelihood function (Gong & Samaniego, 1981) are described in §2. Strong consistency and asymptotic distributions of these point estimators are derived in §3, along with semiparametric likelihood ratio tests of whether the Cox or accelerated failure time model is appropriate. The estimators derived from the proposed pseudo likelihood approach are shown to be semiparametric efficient. Simulation results are reported in §4, evaluating the performance of the procedure and demonstrating the flexibility of the extended hazard model. A case study of Taiwan HIV/AIDS cohort data is presented in §5 to illustrate the proposed joint modelling and model selection approaches.

2. JOINT MODEL AND ESTIMATION

2·1. *Joint model*

Without loss of generality, we assume a single time-dependent covariate $X_i(t)$, and a single time-independent covariate Z_i for subject i, where $i = 1, ..., n$. Let T_i and C_i be the event and censoring times for the ith individual. Conditional on $X_i(\cdot)$ and Z_i, we assume that T_i and

C_i are independent. The observed event-time is $V_i = \min(T_i, C_i)$ and $\Delta_i = 1(T_i \leq C_i)$ is the censoring indicator.

The longitudinal covariate process is scheduled to be measured at times t_{ij}, but no measurements are available after the event time. Therefore, the measurement schedule for subject i is $t_i = (t_{ij},\, t_{ij} \leq V_i)$ and the number of observable repeated measurements m_i depends on V_i. Due to measurement errors or random fluctuations, the observed longitudinal data are

$$W_{ij} = X_i(t_{ij}) + e_i(t_{ij}), \tag{7}$$

where

$$X_i(t) = b_i^{\mathrm{T}} \rho(t) \tag{8}$$

and $e_i(t)$ are realizations of $X(t)$ and $e(t)$ in (1) and (2), and b_i is independent of all $e_i(t_{ij})$. As in (1) and (2), we assume normal distributions for the random effects b_i and measurement error $e_i = e_i(t_i) = (e_{i1}, \ldots, e_{im_i})$. The normality assumption may be unrealistic in some applications and is difficult to verify as with any random effect model. However, robustness against the distribution of random effects has been noted in the literature; see Song et al. (2002) or Tsiatis & Davidian (2004). The property holds when there are a moderate number of repeated measurements for each subject. An explanation for this property is provided in Hsieh et al. (2006) and a theoretical justification can be found in Rizopoulos et al. (2008), who pointed out that if the longitudinal component is modelled by linear mixed effects, even small values of m_i ($\leq$4 in their studies) can suffice to retain robustness.

The survival component follows an extended hazard model similar to (5), but a different notation $\gamma = (\gamma_1, \gamma_2)$ is used for the regression parameters of the baseline covariate Z_i:

$$\lambda\{t \mid \bar{X}_i(t), Z_i\} = \lambda(t \mid \beta, \gamma, Z_i, b_i) = \lambda_0\{\psi(t; \beta_1, \gamma_1, Z_i, b_i)\}\psi'(t; \beta_2, \gamma_2, Z_i, b_i), \tag{9}$$

where $\lambda_0(\cdot)$ is the unspecified baseline hazard function, $\beta = (\beta_1, \beta_2)$ is the regression parameter for the longitudinal covariate $X_i(\cdot)$, and $\psi_i(t) = \psi(t; \beta_1, \gamma_1, Z_i, b_i) = \int_0^t \exp\{\beta_1 b_i^{\mathrm{T}} \rho(s) + \gamma_1 Z_i\} ds$, corresponds to a time scale transformation in the baseline hazard function. The derivative of $\psi(\cdot)$ is

$$\psi'(t; \beta_2, \gamma_2, Z_i, b_i) = \exp\{\beta_2 b_i^{\mathrm{T}} \rho(t) + \gamma_2 Z_i\}. \tag{10}$$

We assume a noninformative censoring and measurement schedule t_{ij}, so that the joint likelihood for model (7) and (9), based on the observed data, can be derived without difficulty to be

$$\begin{aligned} L(\theta) = \; & L(\beta, \gamma, \mu, \Sigma, \sigma_e^2, \lambda_0) \\ & = \Pi_{i=1}^n \left[\int \{\Pi_{j=1}^{m_i} f_l(W_{ij} \mid b_i, t_i, \sigma_e^2)\} f_s(V_i, \Delta_i \mid b_i, t_i, \lambda_0, \beta, \gamma, Z_i) f_r(b_i \mid \Sigma, \mu) db_i \right], \end{aligned} \tag{11}$$

where $f_l(W_{ij} \mid b_i, t_i, \sigma_e^2)$ is the density of $N\{b_i^{\mathrm{T}} \rho(t), \sigma_e^2\}$, $f_r(b_i \mid \Sigma, \mu)$ is the density of $N(\mu, \Sigma)$, and

$$f_s(V_i, \Delta_i \mid b_i, t_i, \lambda_0, \beta, \gamma, Z_i) = \left[\lambda\{V_i \mid \bar{X}_i(t), Z_i\}\right]^{\Delta_i} \exp\left[- \int_0^{V_i} \lambda\{t \mid \bar{X}_i(t), Z_i\} dt \right].$$

Because of the nonparametric model for λ_0, (11) is unbounded, so the maximum likelihood estimator does not exist. Furthermore, the nonparametric maximum likelihood approach also fails, as explained by Zeng & Lin (2007) for the accelerated failure time model, so we use a nonparametric pseudo-likelihood approach.

2·2. *Estimation*

The EM-algorithm is employed to maximize (11), with the unobserved random effects b_i treated as missing data. Given the complete data for subject i, $(V_i, \Delta_i, W_i, t_i, b_i, Z_i)$, the com-

plete data likelihood is

$$L^c(\theta) = \Pi_{i=1}^n \left[\{\Pi_{j=1}^{m_i} f_l(W_{ij} \mid b_i, t_i, \sigma_e^2)\} f_s(V_i, \Delta_i \mid b_i, t_i, \lambda_0, \beta, \gamma, Z_i) f_r(b_i \mid \Sigma, \mu) \right].$$

For details of the EM-algorithm, we refer to the joint accelerated failure time framework in Tseng et al. (2005). However, to adjust their EM procedure to our extended hazard model, we need to modify the estimation of λ_0 in the M-step, as elaborated below.

Let $T_1, \ldots, T_D$ denote the D observed failure times among the n subjects, accepting that there may be ties in the data. Define a time scale transformation $\psi(t) = \psi(t; \beta_1, \gamma_1, Z, b) = \int_0^t \exp\{\beta_1\, b^{\mathrm{T}} \rho(s) + \gamma_1 Z\} ds$. First, we consider an unspecified baseline hazard that is a step function taking a constant value C_j on $[\hat{u}_{(j-1)}, \hat{u}_{(j)})$, where $u = \psi(t)$ and $\hat{u}_{(k)}$ $(k = 1, ..., D)$ denote the ordered current estimates of $\psi(T_k)$. The baseline hazard function takes the form $\lambda_0(u) = \sum_{j=1}^D C_j 1_{\{\hat{u}_{(j-1)} \le u < \hat{u}_{(j)}\}}$.

For subject i, we rewrite the cumulative hazard function, Λ, as

$$\begin{aligned} &\int_0^{V_i} \lambda_0\{\psi_i(s)\} \exp\{\beta_2 X_i(s) + \gamma_2 Z_i\} ds \\ &= \sum_{j=1}^D C_j \exp(\gamma_2 Z_i) \int_{\psi_i^{-1}\{\hat{u}_{(j-1)}\}}^{\min[\psi_i^{-1}\{\hat{u}_{(j)}\}, V_i]} \exp\{\beta_2 X_i(s)\} ds 1_{\{\hat{u}_{(j-1)} \le u_i\}}, \end{aligned}$$

where $u_i = \psi(V_i)$.

For simplicity, we use $E_i\{g(b_i)\}$ to denote $E\{g(b_i) \mid V_i, \Delta_i, W_i, t_i, \hat{\theta}\}$, the conditional expectation of some function g, given the current estimate $\hat{\theta} = (\hat{\mu}, \hat{\Sigma}, \hat{\sigma}_e, \hat{\lambda}_0, \hat{\beta}, \hat{\gamma})$. The conditional score equation with respect to C_k is

$$\frac{\partial}{\partial C_k} \sum_{i=1}^n E_i[\Delta_i \log \lambda_0(u_i) - \Lambda\{\psi(V_i; \beta_1, \gamma_1, Z_i, b_i), \beta_2, \gamma_2, b_i\} \mid Z_i] = 0. \tag{12}$$

By solving the (12), the nonparametric pseudo maximum likelihood estimate for C_k is

$$\hat{C}_k = \frac{\sum_{i=1}^n E_i[\Delta_i 1_{\{\hat{u}_{(k-1)} \le u_i < \hat{u}_{(k)}\}}]}{\sum_{i=1}^n \exp(\gamma_2 Z_i) E_i \left(\int_{\psi_i^{-1}\{\hat{u}_{(k-1)}\}}^{\min[\psi_i^{-1}\{\hat{u}_{(k)}\}, V_i]} \exp[\beta_2 X(s) ds 1_{\{\hat{u}_{(k-1)} \le u_i\}}] \right)}. \tag{13}$$

There is no closed-form solution for the conditional score equation $(\partial\beta/\partial)E_i\{\log L^c(\theta)\}$. Furthermore, gradient methods cannot applied to the maximization of this likelihood function given the discontinuous baseline hazard function. In such a situation, direct maximization methods can be implemented but their convergence can be extremely slow. To overcome this numerical difficulty and to facilitate the theory, we turn to a smooth estimator of the baseline hazard function. A natural candidate is the kernel density estimator

$$\hat{\lambda}_0(u) = (1/h)\sum_{j=1}^{D} K[\{u - \hat{u}_{(j)}\}/h]d\hat{\Lambda}_0\{\hat{u}_{(j)}\},$$

where $d\hat{\Lambda}_0\{u_{(j)}\} = \hat{C}_j\{\hat{u}_{(j)} - \hat{u}_{(j-1)}\}$ with $\hat{C}_j$ defined in (13).

With this smooth estimate of λ_0, $\hat{\beta}$ and $\hat{\gamma}$ can be obtained using a Newton–Raphson method at the M-step. In §4 and §5, we choose a standard Gaussian kernel in the simulation and data analysis with the optimal bandwidths $(8\sqrt{2}/3)^{1/5}\sigma n^{-1/5}$, where σ is the sample standard deviation of the uncensored $\hat{u}_k$ (Jones, 1990; Jones & Sheather, 1991).

In the E-step, the Monte Carlo integration with antithetic variables in Henderson et al. (2000) is adapted to compute the conditional expectations. To ensure that the Monte Carlo EM is accurate to the fourth decimal place, we suggest a minimal Monte-Carlo size of 10000 to generate the multivariate normal vectors when the EM algorithm is close to convergence. The Supplementary Material has more information about the size of Monte Carlo samples and the corresponding accuracy. Properties of Monte-Carlo EM procedures have been extensively discussed; see Booth & Hobert (1999), Caffo et al. (2005) and Wei & Tanner (1990). To speed up the convergence, we suggest a two-stage procedure to obtain reasonable initial values. The first stage is to fit the longitudinal data with a mixed effects model and then to impute the entire covariate history. With the complete history, the second stage is then to apply the score equation procedure in Tseng et al. (2014) to derive initial estimates of the regression coefficients.

3. ASYMPTOTIC THEORY AND STANDARD ERRORS

3·1. *Asymptotic theory*

In this section, we show strong consistency and provide the asymptotic distribution of the proposed estimator with a general distribution assumption $f_{(b,e)(\cdot|\alpha)}$ with parameter α. For simplicity, the results do not include time-independent covariates Z but similar results hold for their estimates. We will use τ and $\tilde{\tau}$ to denote the time at the end of study and the largest transformed time at the end of study, respectively. We start by listing some conditions needed in the proofs.

Assumption 1. The parameter space for the finite-dimensional parameters is closed and bounded in the Euclidean space. Moreover, the true value is an interior point of the parameter space.

Assumption 2. The conditional expectations $E_i\left[\exp\{\beta_j b_i^{\mathrm{T}}\rho(t)\}\right]$ and $E_i\left[\exp\{-\beta_j b_i^{\mathrm{T}}\rho(t)\}\right]$ are uniformly bounded above and below from 0 with probability 1 in the parameter space of β_j ($j=1,2$). Furthermore, $E_\theta(\exp[b^{\mathrm{T}}\{\beta_2\rho(t_2)-\beta_1\rho(t_1)\}]I\{\psi(V)=\tilde{\tau}\})$ is bounded away from 0, for arbitrary t_1, t_2 in $[0,\tau]$ and β_1, and β_2 in the parameter space.

Assumption 3. The bandwidth h converges to 0 at rate n^{-r}, where $r<1$. The kernel function $K(\cdot)$ is a three times differentiable continuous density function. Its corresponding derivatives are of bounded variation on $\mathbb{R}$.

Assumption 4. The function λ_0 is positive, uniformly bounded, and differentiable on $[0,\tilde{\tau}]$. Moreover, λ_0 and λ_0' are Lipschitz continuous, and of bounded variation on $[0,\tilde{\tau}]$.

Assumption 5. The function ρ is of bounded variation. Both $E_\theta\left\{\beta_j b^{\mathrm{T}}\rho(t)\right\}$ and $E_\theta\left\{\psi^{(k)}(t,\beta_j,b)\right\}$ are bounded almost everywhere on the support of T, for $j=1,2$ and $k=1,2$.

Assumption 6. The distribution $f_{(b,e)(\cdot|\alpha)}$ has continuous second derivative with respect to α and its Fisher information matrix is positive definite.

THEOREM 1. *Define* $\hat{\Lambda}^s(u) = \int_0^u \hat{\lambda}_0(w)dw$. *Under Assumptions 1-4,* $(\hat{\beta}_1, \hat{\beta}_2, \hat{\alpha}, \hat{\Lambda}^s)$ *converges uniformly to the true values* $(\beta_1^0, \beta_2^0, \alpha^0, \Lambda_0)$ *under the joint Euclidean norm and the surpremum norm on* $[0, \tilde{\tau}]$ *as* $n \to \infty$.

THEOREM 2. *Let* A_q *be a direction space* $\{(a_{11}, a_{12}, a_2, a_3) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}^d \times \mathcal{L}_{BV} : |a_{11}| + |a_{12}| + |a_2| + \| a_3 \|_\infty < q\}$, *for some* q. *The notation* $\mathcal{L}_{BV}$ *stands for a class of bounded-variation functions on* $[0, \tilde{\tau}]$. *Under Assumptions 1-6,*

(i) $n^{1/2}\left\{(\hat{\beta}_1, \hat{\beta}_2, \hat{\alpha}, \hat{\Lambda}^s) - (\beta_1^0, \beta_2^0, \alpha^0, \Lambda_0)\right\}$ *converges weakly to Gaussian element with mean zero in the functional space* $l_\infty(A_q)$.

(ii) The estimators for the parametric components are semiparametrically efficient.

Theorem 2 facilitates the construction of semiparametric likelihood ratio tests, which can be used to test the adequacy of the Cox or accelerated failure time model as follows. For convenience, we rewrite the hazard function in (9) as $\lambda_0\{\psi(t; \beta_1, Z_i, b_i)\}\psi'(t; \beta_1 + \Delta\beta, Z_i, b_i)$ with $\Delta\beta = \beta_2 - \beta_1$. The Cox model corresponds to $\beta_1 = 0$; the accelerated failure time model corresponds to $\Delta\beta = 0$. Especially, $\Delta\beta$ is not involved in λ_0. Let $\hat{\phi}(\Delta\beta^0)$ and $\hat{\phi}(\beta_1^0)$ denote the maximum likelihood estimatiors obtained for $\theta \setminus \Delta\beta$ and $\theta \setminus \beta_1$ when setting $\Delta\beta = \Delta\beta^0$ and $\beta_1 = \beta_1^0$, respectively.

THEOREM 3. *Under Assumptions 1-6 and 7 in the Appendix, the likelihood ratio test statistic*

$$\xi = 2\log\frac{L(\hat{\theta})}{L\{\Delta\beta^0, \hat{\phi}(\Delta\beta^0)\}}$$

converges in distribution to a χ^2 distribution with degrees of freedom equal to the dimension of $\Delta\beta$.

THEOREM 4. *Under Assumptions 1-6 and 8 in the Appendix, the likelihood ratio test statistics*

$$\xi = 2\log\frac{L(\hat{\theta})}{L\{\beta_1^0, \hat{\phi}(\beta_1^0)\}}$$

converges in distribution to a χ^2 distribution with degrees of freedom equal to the dimension of β_1.

3·2. *Standard errors*

For the accelerated failure time model without longitudinal covariates, Zeng & Lin (2007) suggested a profile likelihood approach to estimate the inverse of the variance matrix, $\Sigma^{-1}_{(\beta,\gamma)}$, which then leads to standard errors. Unfortunately, this does not work well for the extended hazard model in the joint modelling setting. In our simulations, the profile variance matrices were often not invertible, could lead to negative standard error estimates, or had inaccurate inverses. An alternative and popular approach that involves the EM-algorithm is to employ the formula in Louis (1982). This, too, did not work well in the extended hazard joint modelling setting for similar reasons as for the profile method. We thus rely on the bootstrap to estimate standard errors. Details of the procedure and its numerical performance were first established in Tseng et al. (2005) for the accelerated failure time survival model with longitudinal covariates. A similar bootstrap method applies to the extended hazard model. We will not elaborate this further.

4. SIMULATION STUDIES

In this section, we investigate the performance of the proposed estimation procedure through simulations. Three scenarios are considered. The first serves to evaluate the accuracy of the estimators for both the longitudinal (7) and survival models (9). The second aims at evaluating

the impact on the regression coefficients when the survival model is misspecified. The third attempts to investigate properties of the likelihood ratio test statistics of Theorems 3 and 4.

In all scenarios, the preliminary scheduled measurement times for each subject are ten time points equally divided from zero to the time at the end of study, τ, which varies with the scenarios. In addition for scenario one and three, the lifetime is subject to independent censoring according to the exponential distribution. The values of τ and the means of exponential censoring are selected to result in a target censoring rate of about 30%. No measurement is available after death or censoring time.

In scenario one, the simulated sample sizes are n=100, 200, and 400. The baseline hazard function is $\lambda_0(t) = 10^{-5}\exp(2t)$. The baseline covariate Z is symbolic binary. The regression coefficients for the survival model are $(\beta_1, \beta_2, \gamma_1, \gamma_2) = (0 \cdot 5, 1, 0 \cdot 5, 1)$. For the longitudinal model, we used the benchmark provided by the linear growth curve model with $\rho_1(t) = 1$ and $\rho_2(t) = t$. The random effects are normally distributed with mean $\mu = (1, 0 \cdot 5)^{\mathrm{T}}$ and the variance components of Σ are $(\sigma_{11}, \sigma_{12}, \sigma_{22})^{\mathrm{T}} = (0 \cdot 04, -0 \cdot 01, 0 \cdot 01)^{\mathrm{T}}$. The measurement error is also normally distributed with mean zero and variance $\sigma_e^2 = 0 \cdot 04$. The mean of exponential random censor is set to be six and $\tau = 3$. Moreover, to evaluate the coverage probability of the bootstrap confidence intervals , 100 bootstrap resamples are taken for each of 500 Monte Carlo samples.

In scenario two, the sample size is n=100, and the settings of the baseline hazard function and the longitudinal model are the same as scenario one. However, no time-independent covariate is considered in the simulation and no random censoring is involved. Three sets of regression coefficients are considered for the single time-dependent covariate:

(a) $(\beta_1, \beta_2) = (-0 \cdot 5, 0)$, the true survival time follows the Cox model and $\tau = 9$;

(b) $(\beta_1, \beta_2) = (0 \cdot 5, 0 \cdot 5)$, the true survival time follows the accelerated failure time model and

673 $\tau = 3$;

674 (c) $(\beta_1, \beta_2) = (-0 \cdot 2, 0 \cdot 5)$, the true survival time follows the extended hazards model and $\tau =$

675 9.

676 The settings of scenario three are the same as in scenario one except for the regression coeffi-

677 cients and end of study times:

678 (d) $(\beta_1, \beta_2, \gamma_1, \gamma_2) = (0, 0 \cdot 5, 0, 0 \cdot 5)$, the true survival time follows the Cox model and $\tau = 6$;

679 (e) $(\beta_1, \beta_2, \gamma_1, \gamma_2) = (0 \cdot 5, 0 \cdot 5, 0 \cdot 5, 0 \cdot 5)$, the true survival time follows the accelerated fail-

680 ure time model and $\tau = 3$.

681 The regression coefficients in setting (d) correspond to testing the Cox model against the ex-

682 tended hazard model with $H_0 : \beta_1 = \gamma_1 = 0$, and setting (e) corresponds to testing accelerated

683 failure time model against the extended hazard model with $H_0 : \beta_1 = \beta_2, \gamma_1 = \gamma_2$. Error rejec-

684 tion rates at the 5% level of both tests are calculated based on 500 replications with sample sizes

685 $n =$100, 200, 400. The independent censoring times are generated from exponential distributions

686 with means fifteen and six for set (d) and (e), respectively.

687 Table 1 shows that the pseudo likelihood approach yields approximately unbiased estimators

688 for all the parameters. As expected, their efficiency increases with sample size. In addition, the

689 coverage probability of bootstrap confidence intervals approaching 95% nominal level. This in-

690 dicates that the bootstrap method works well in the estimation. Figure 1 (a)–(c) demonstrates

691 estimation of the baseline hazard function $\lambda_0(u)$ for $u \in [0, 6]$, which covers about 95% of the

692 transformed times $\hat{u}_k$ in the simulated data. The estimated hazard curves approach the target

693 hazard function as n becomes large and the proposed estimating procedures obtain satisfactory

694 estimators for both the longitudinal and survival components. Since the theorems in §3 apply only

695 to the smoothed baseline hazard estimators and we know that the piecewise constant baseline es-

696 timate is not consistent, a natural question is whether this consistency assumption is necessary

697

698

699

700

701

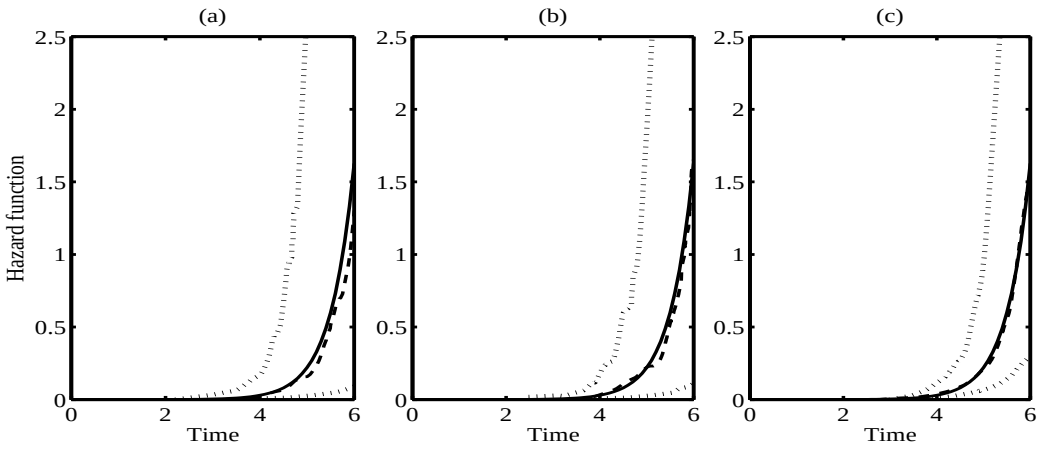

Fig. 1. Estimated baseline hazard function (dashed) and true baseline hazard function (solid) with 95% Monte Carlo confidence band (dotted) for (a) n=100, (b) n=200 and (c) n=400.

as well. Zeng & Lin (2007) made a similar note for the traditional accelerated failure time model with baseline covariates but did not pursue it. To check this empirically, we include in the top panel of Table 1 simulation results for the piecewise constant baseline hazard estimate for the case $n = 100$. The results of Table 1 reveal that even though the baseline estimator may not be consistent, its parametric estimates are comparable to those from the kernel smoothed ones, so the consistency assumption seems may not be necessary. Why then bother to smooth the piecewise constant baseline hazard estimates iteratively during the EM algorithm? The reason is its numerical advantage as explained in the M-step of the EM algorithm.

Table 2 shows the performance of three joint modelling approaches under different targeted survival models. The joint extended hazard approach produces approximately unbiased estimators in all three cases, but the other two approaches led to different conclusions and often insignificant estimates when the survival model is misspecified. The joint Cox and accelerated failure time approaches may give different results when the model is misspecified. It is reasonable that the standard errors of the estimates under the joint extended hazards approach are in some cases

Table 1. *simulation results of Scenario one*

	β_1	β_2	γ_1	γ_2	μ_1	μ_2	σ_{11}	σ_{12}	σ_{22}	σ_e^2
Target	0·5	1	0·5	1	1	0·5	0·05	$-0\cdot 01$	$0\cdot 01$	0·04
	$n = 100$ (piecewise constant baseline hazard without kernel smooth)									
Mean	0·464	0·973	0·512	0·969	1·002	0·505	0·041	−0·009	0·009	0·035
SD	0·064	0·096	0·041	0·131	0·022	0·013	0·007	0·006	0·006	0·005
Cover	93·6%	93·4%	93·2%	93·0%	93·0%	92·4%	93·2%	90·8%	92·4%	93·2%
	$n = 100$									
Mean	0.474	1·009	0·493	1·016	1·003	0·498	0·041	−0·009	0·010	0·036
SD	0.060	0·097	0·038	0·134	0·021	0·013	0·006	0·005	0·004	0·005
Cover	93·2%	94·0%	93·4%	93·0%	92·8%	93·6%	94·0%	91·2%	93·4%	93·8%
	$n = 200$									
Mean	0·488	0·991	0·498	1·007	1·001	0·498	0·041	−0·010	0·010	0·038
SD	0·052	0·078	0·031	0·111	0·016	0·009	0·005	0·002	0·002	0·004
Cover	94·0%	95·6%	93·8%	95·8%	94·2%	94·0%	94·4%	92·6%	94·0%	94·6%
	$n = 400$									
Mean	0·496	0·997	0·501	1·001	1·001	0·499	0·040	−0·010	0·010	0·040
SD	0·043	0·059	0·022	0·085	0·010	0·007	0·004	0·002	0·001	0·004
Cover	94·4%	95·0%	94·2%	95·4%	95·4%	94·8%	95·2%	93·6%	95·4%	95·2%

Mean, average of 500 Monte Carlo estimates; SD, standard deviation of 500 Monte Carlo estimates; Cover, coverage probability of bootstrap confidence intervals.

slightly larger than those obtained from the other two approaches due to the larger number of parameters in the joint extended hazard model.

The results of senario 3 are summarize in §3·3 of supplementary material which suggest that as the sample size increases, the error rejection rate approaches the nominal level $\alpha = 0 \cdot 05$.

In addition to the three scenarios, power analysis of likelihood ratio test for the Cox and the accelerated failure time models when the extended hazard model holds for data is simulated

Table 2. *Simulation results of scenario two-performance of the three joint models when the target survival time is misspecified*

	Cox		AFT		EH	
	Estimate	SD	Estimate	SD	Estimate	SD
Target survival - Cox						
$\beta_1 = 0$	-	-	$\beta_1 = \beta_2$	-	$-0{\cdot}028$	$0{\cdot}072$
$\beta_2 = -0 \cdot 5$	$-0{\cdot}503$	$0{\cdot}065$	$-0{\cdot}097$	$0{\cdot}077$	$-0{\cdot}488$	$0{\cdot}082$
Target survival - AFT						
$\beta_1 = 0 \cdot 5$	-	-	$\beta_1 = \beta_2$	-	$0{\cdot}455$	$0{\cdot}068$
$\beta_2 = 0 \cdot 5$	$0{\cdot}142$	$0{\cdot}135$	$0{\cdot}494$	$0{\cdot}106$	$0{\cdot}505$	$0{\cdot}127$
Target survival - EH						
$\beta_1 = -0 \cdot 2$	-	-	$\beta_1 = \beta_2$	-	$-0{\cdot}201$	$0{\cdot}083$
$\beta_2 = 0 \cdot 5$	$0{\cdot}130$	$0{\cdot}232$	$0{\cdot}427$	$0{\cdot}091$	$0{\cdot}502$	$0{\cdot}114$

AFT, accelerated failure time model; EH, extended hazard model.

with the same setting as scenario 2 but the regression coefficients are set to be four different values of regression coefficients, $(\beta_1, \beta_2) = (0 \cdot 2, 1), (0 \cdot 4, 1), (0 \cdot 6, 1), (0 \cdot 8, 1)$. Table 2 of supplementary material §3·2 shows the results of additional simulations. As expected, the larger the alternative models depart from the null model, the higher power of likelihood ratio tests dectect the departure. Additional simulations to investigate sensitivity of the estimates to larger measurement errors, non-normal random effects cases can be found in supplementary material.

5. APPLICATION TO TAIWANESE HIV/AIDS COHORT DATA

Data on 1164 HIV/AIDS infected patients in Taiwan were collected from 1990 to January, 2003. All patients were advised to return every three or four months for a follow-up check. We focus on in the association between the survival time, AIDS onset to death, and the biomarker, CD4 count, as well as evaluating the efficacy of HAART, highly active antiretroviral therapy ,

which consists of at least three anti-HIV drugs. We apply the proposed joint model procedure to the 137 patients who had developed AIDS in the cohort data, of which 37 patients received HAART right after AIDS onset. The other 100 patients received non-HAART all the time.

In the proposed joint modelling procedure, we need to select a suitable basis $\rho(t)$ for the longitudinal model (8). In practice, a parametric basis is often used for interpretability and computational efficiency. We found that a cubic random coefficient model, $\rho(t) = (1, t, t^2, t^3)$ seems appropriate for the pattern of logarithmic CD4 counts for both groups. Therefore, log(CD4) for the ith subject is modelled by

$$W_i(t) = b_i^{\mathrm{T}}\rho(t) + e_i(t) = b_{i0} + b_{i1}t + b_{i2}t^2 + b_{i3}t^3 + e_i(t),$$

where $e_{ij} = e_i(t_{ij}) \sim N(0, \sigma_e^2)$ and $b_i^{\mathrm{T}} \sim N(\mu_{4\times1}, \Sigma_{4\times4})$, for $i = 1, ..., 137$, $j = 1, ...m_i$ and $1 \leq m_i \leq 13$.

The treatment indicator Z was coded one when a patient received HAART treatment and zero otherwise. The results after fitting the joint extended hazard model are summarized in Table 3. The estimates of the regression coefficients γ_1 and γ_2 in the survival model are both significantly negative while the estimates for β_1 with p-value=0·08 and β_2 with p-value=0·06 are marginally negative, indicating the effectiveness of the HAART treatment and that lower CD4 counts are associated with higher risk of death due to AIDS. The fitted cubic log(CD4) curves for nine randomly selected subjects of each treatment group are plotted in supplementary material. The overall fit for these randomly selected 18 subjects seems reasonable.

Although the normality assumption on the random effects is difficult to verify, one might rely on the robustness property against normality assumption violations, as illustrated in Hsieh et al. (2006) and Rizopoulos et al. (2008), because this data set has a relatively large number of thirteen repeated measurements. The square root of estimated measurement error, 0·255, is less than 11% of the average of the log CD4 counts, 2·360. Moreover, the assumptions of noninformative

censoring and measurement schedule are reasonable for the HIV/AIDS cohort data due to the patients' high adherence to their assigned schedules.

Next, we conduct a semiparametric likelihood ratio test to check for the fit of the survival component in the joint model. Since the Cox model and accelerated failure time model are nested within the extended hazard model, we conducted a semiparametric likelihood ratio test with the extended hazard model as the full model and the other two models as reduced models. The two null hypotheses are H_0: $\beta_1 = \gamma_1 = 0$ and H_0: $\beta_1 = \beta_2, \gamma_1 = \gamma_2$ for testing the adequacy of the Cox or the accelerated failure time model, respectively. Table 4 shows that the Cox model is unsuitable, as it is rejected with p-value=0·001, and the accelerated failure time model could be selected for the data. This is an interesting finding, as all previous analyses of AIDS data with CD4 counts have assumed a Cox model for time to death, while our analysis for the Taiwanese AIDS data clearly indicates that the Cox model is unsuitable and an accelerated failure time model is better.

Upon refitting the Taiwan AIDS data with the accelerated failure time survival model, the regression parameter $\beta_1 = \beta_2$ in Table 3 now becomes significant, while under the extended hazard model, β_1 and β_2 were not significant at the 5% level.

ACKNOWLEDGEMENT

The research of Jane-Ling Wang is supported in part by National Institutes of Health and National Science Foundation grants. The research of Yi-Kuan Tseng is supported by a grant of National Science Council of Taiwan.

Table 3. *Parameter estimation of the HIV/AIDS data under the extended hazard model and the accelerated failure time model*

	Joint EH model			Joint AFT model		
	Estimate	Mean	SD	Estimate	Mean	SD
Regression coefficient						
β_1	− 0·814	−0·941	0·462	−1·017	−0·970	0·408
β_2	−0·772	−0·639	0·404	−1·017	−0·970	0·408
γ_1	−1·584	−1·504	0·528	−1·335	−1·593	0·416
γ_2	−1·594	−1·631	0·531	−1·335	−1·593	0·416
CD4 cell mean pattern						
μ_1	2·338	2·335	0·078	2·536	2·339	0·043
μ_2	−0·146	−0·112	0·202	−0·182	−0·150	0·200
μ_3	−0·013	−0·006	0·133	0·024	0·003	0·144
μ_4	0·005	0·007	0·022	0·004	0·008	0·024
Covariance matrix						
σ_{11}	0·410	0·407	0·061	0·411	0·398	0·050
σ_{12}	−0·184	−0·201	0·142	−0·174	−0·189	0·145
σ_{13}	0·011	0·027	0·087	0·008	0·033	0·095
σ_{14}	0·001	−0·001	0·015	0·001	−0·003	0·016
σ_{22}	0·882	1·029	0·175	0·898	1·035	0·157
σ_{23}	−0·273	−0·388	0·226	−0·279	−0·406	0·170
σ_{24}	0·025	0·044	0·041	0·226	0·047	0·042
σ_{33}	0·113	0·186	0·157	0·115	0·104	0·069
σ_{34}	−0·013	−0·025	0·025	−0·013	−0·027	0·027
σ_{44}	0·002	0·004	0·004	0·002	0·004	0·004
Measurement error						
σ_e^2	0·065	0·069	0·016	0·065	0·072	0·007

AFT, accelerated failure time model; EH, extended hazard model; Mean, mean of 100 bootstrap estimates; SD, standard deviation of 100 bootstrap estimates.

Table 4. *Summary of the estimates of regression parameters when fitting the HIV/AIDS cohort data by joint Cox, accelerated failure time and extended hazard models.*

	β_1	γ_1	β_2	γ_2	-2log likelihood	p-value
Cox	0	0	$-0{\cdot}610$	$-1{\cdot}845$	316·2	0·001
Accelerated failure time	$-1{\cdot}017$	$-1{\cdot}335$	$-1{\cdot}017$	$-1{\cdot}335$	304·2	0·301
Extended hazard	$-0{\cdot}814$	$-1{\cdot}584$	$-0{\cdot}772$	$-1{\cdot}594$	301·8	

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes more detail on the EM algorithm used under the proposed approach, part of the proofs, additional simulations, and two figures for the HIV/AIDS case study.

APPENDIX: PROOFS OF THE THEOREMS

Proof of Theorem 1

We first show the boundedness of $\hat{\Lambda}_n^s$ at $\tilde{\tau}$. By integrating $\hat{\lambda}_0$ from 0 to $\tilde{\tau}$, we have

$$\begin{aligned}\hat{\Lambda}_n^s(\tilde{\tau}) &= \int_0^{\tilde{\tau}} \frac{1}{h}\sum_{j=1}^{D} K\left\{\frac{w-\hat{u}_{(j)}}{h}\right\}\hat{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}dw \\ &= \frac{1}{h}\sum_{j=1}^{D}\left[\int_0^{\tilde{\tau}} K\left\{\frac{w-\hat{u}_{(j)}}{h}\right\}dw\hat{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}\right].\end{aligned}$$

Since the integral of a kernel function over its domain equals to 1, we have the following inequality:

$$\begin{aligned}\hat{\Lambda}_n^s(\tilde{\tau}) &\le \sum_{j=1}^{D}\hat{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\} \\ &= \sum_{j=1}^{D}\frac{\sum_{i=1}^{n}E_i\left[\Delta_i I\left\{\hat{u}_{(j-1)}\le\hat{\psi}_i(V_i)<\hat{u}_{(j)}\right\}\right]\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}}{\sum_{i=1}^{n}E_i\Big(\int_{\hat{\psi}_i^{-1}\{\hat{u}_{(j-1)}\}}^{\min[\hat{\psi}_i^{-1}\{\hat{u}_{(j)}\},V_i]}\exp\{\hat{\beta}_2 b_i^{\mathrm{T}}\rho(t)\}dtI\left\{\hat{u}_{(j-1)}\le\hat{\psi}_i(V_i)\right\}\Big)},\end{aligned}$$

where $\hat{\psi}$ denotes the estimated ψ with the involving β replaced by the corresponding $\hat{\beta}$. By the mean value theorem and the differentiability of $\hat{\psi}_i$ and $\hat{\psi}^{-1}$,

$$\hat{\Lambda}_n^s(\tilde{\tau}) \leq \frac{\frac{1}{n}\sum_{i=1}^{n}\Delta_i}{\frac{1}{n}\sum_{i=1}^{n}E_i\left[\exp\{\hat{\beta}_2 b_i^{\mathrm{T}}\rho(t^m)\}\exp\{-\hat{\beta}_1 b_i^{\mathrm{T}}\rho(t^M)\}I\left\{\hat{\psi}_i(V_i)=\tilde{\tau}\right\}\right]}, \tag{A1}$$

where t^m and t^M are the minimizer and maximizer, respectively, of $\rho(t)$ over $[0,\tau]$. Based on the strong law of large numbers and regularity conditions, the limit of the right-hand-side in (A1), as $n\to\infty$, is finite. This implies that $\hat{\Lambda}_n^s(\tilde{\tau})$ is bounded above with probability 1 as $n\to\infty$. Thus Helly's selection theorem and the Bolzano–Weierstrass theorem ensure the existence of a subsequence $\hat{\theta}_{\eta(n)}$ of $\hat{\theta}_n$ converging to a certain $\theta_0^* = (\beta_1^*, \beta_2^*, \alpha^*, \Lambda^*)$.

To show that the limit θ_0^* is exactly equal to θ_0, we need a new term $\bar{\Lambda}_n$ defined below to serve as the bridge between $\hat{\Lambda}_n^s$ and Λ_0. Let

$$\bar{\Lambda}_n(u) = \frac{1}{h}\sum_{j=1}^{D}\left[\int_0^u K\left\{\frac{w-\hat{u}_{(j)}}{h}\right\}dw\bar{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}\right],$$

where

$$\bar{C}_j = \frac{\sum_{i=1}^{n}E_{i0}\left[\Delta_i I\left\{\hat{u}_{(j-1)}\leq\psi_{i0}(V_i)<\hat{u}_{(j)}\right\}\right]}{\sum_{i=1}^{n}E_{i0}\left(\int_{\psi_{i0}^{-1}\{\hat{u}_{(j-1)}\}}^{\min[\psi_{i0}^{-1}\{\hat{u}_{(j)}\},V_i]}\exp\{\beta_{20}b_i^{\mathrm{T}}\rho(t)\}dtI\left\{\hat{u}_{(j-1)}\leq\psi_{i0}(V_i)\right\}\right)}.$$

Here E_{i0}, ψ_{i0}, and ψ_{i0}^{-1} stand for E_i, ψ_i, and ψ_i^{-1} with β_1 replaced by β_{10}. It can be shown that, as $n\to\infty$, $\bar{\Lambda}_n(u)$ converges uniformly to $\Lambda_0(u)$ on $[0,\tilde{\tau}]$; see §2 of the supplementary material for details. The proof of consistency will be completed by demonstrating that $\theta_0^* = \theta_0$.

It can be shown that $\hat{\Lambda}_{\eta(n)}^s$ is absolutely continuous with respect to $\bar{\Lambda}_{\eta(n)}$ by the following equation:

$$\hat{\Lambda}_{\eta(n)}^s(u) = \int_0^u \frac{\hat{\lambda}_0(t)}{\bar{\lambda}_{\eta(n)}(t)}d\bar{\Lambda}_{\eta(n)}(t) = \int_0^u \frac{\frac{1}{h}\sum_{j=1}^{D}K\left\{\frac{t-\hat{u}_{(j)}}{h}\right\}\hat{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}}{\frac{1}{h}\sum_{j=1}^{D}K\left\{\frac{t-\hat{u}_{(j)}}{h}\right\}\bar{C}_j\left\{\hat{u}_{(j)}-\hat{u}_{(j-1)}\right\}}d\bar{\Lambda}_{\eta(n)}(t). \tag{A2}$$

As a result of the Glivenko–Cantelli theorem, the integrand in (A2) converges uniformly to

$$R(t) = \frac{E\left(\frac{\Delta}{\exp\{(\hat{\beta}_2-\hat{\beta}_1)b^{\mathrm{T}}\rho(V)\}E\left[I\{t\leq\hat{\psi}(V)\}\right]}\right)}{E\left(\frac{\Delta}{\exp\{(\beta_{20}-\beta_{10})b^{\mathrm{T}}\rho(V)\}E[I\{t\leq\psi_0(V)\}]}\right)}.$$

By taking limits on both sides of (A2), we obtain that $\Lambda^*(u) = \int_0^u R(t)d\Lambda_0(t)$, which illustrates the absolute continuity of Λ^* with respect to Λ_0 which is continuous; hence $\hat{\Lambda}^s_{\eta(n)}$ converges to Λ^* uniformly on $[0, \tilde{\tau}]$.

On the other hand, $\hat{\theta}_n$ maximizes the observed likelihood hence

$$\frac{1}{\eta(n)}\sum_{i=1}^{\eta(n)}\left[l_i\{\hat{\theta}_{\eta(n)}\} - l_i\{\beta_{10}, \beta_{20}, \alpha_0, \bar{\Lambda}_{\eta(n)}\}\right] \geq 0.$$

Since for any $\eta(n)$, as $m \to \infty$,

$$\frac{1}{m}\sum_{i=1}^{m}\left[l_i\{\hat{\theta}_{\eta(n)}\} - l_i\{\beta_{10}, \beta_{20}, \alpha_0, \bar{\Lambda}_{\eta(n)}\}\right] \to E\left[l\{\hat{\theta}_{\eta(n)}\} - l\{\beta_{10}, \beta_{20}, \alpha_0, \bar{\Lambda}_{\eta(n)}\}\right],$$

where $l(\cdot)$ is the generic observed log-likelihood on single subject and $l_i(\cdot)$ is the observed log-likelihood on subject i. It implies that

$$E\left[l\{\hat{\theta}_{\eta(n)}\} - l\{\beta_{10}, \beta_{20}, \alpha_0, \bar{\Lambda}_{\eta(n)}\}\right] \geq -o(1)$$

with probability 1. The dominated convergence theorem thus ensures that

$$E\left[l\{\hat{\theta}_{\eta(n)}\} - l\{\beta_{10}, \beta_{20}, \alpha_0, \bar{\Lambda}_{\eta(n)}\}\right] \to E\left\{l(\theta_0^*) - l(\theta_0)\right\},$$

as $\eta(n) \to \infty$, where the right-hand-side is non-positive. Hence $E\left\{l(\theta_0^*) - l(\theta_0)\right\} = 0$, and the equivalence of θ_0^* and θ_0 is obtained by the identifiability of the model. Since every convergent subsequence converges to the same limit, the consistency of $\hat{\theta}_n$ to θ_0 is concluded.

Proof of Theorem 2

The proof of the asymptotic normality of the proposed estimator can be shown based on Theorem 3.3.1 in Van der Vaart and Wellner (1996) by verifying the four sufficient conditions there. Details can be found in the supplementary material. Below we focus on the proof of semiparametric efficiency.

The Fréchet differentiability of the scores and the consistency of $\hat{\theta}_n$ lead to the following equation:

$$n^{1/2}\nabla_{\hat{\theta}_n - \theta_0} S_{\theta_0}(\theta_0)(a) = n^{1/2}\left\{S_{n,\theta_0}(\theta_0)(a) - S_{\theta_0}(\theta_0)(a)\right\} + o_p(1),$$

where$\nabla_{\hat{\theta}_n-\theta_0} S_{\theta_0}(\theta_0)(a)$ is the Fréchet derivative of the scores as defined in the supplementary material, and $o_p(1)$ converges in probability to the zero element in $l_\infty(A_q)$. On the other hand, the invertibility of the information operator shown in the supplementary material $\sigma = (\sigma_{11}, \sigma_{12}, \sigma_2, \sigma_3) : A_q \to A_q$ implies that, for each a, there exists $\tilde{a} = (\tilde{a}_{11}, \tilde{a}_{12}, \tilde{a}_2, \tilde{a}_3) = \sigma^{-1}(a)$. Thus we have

$$\begin{aligned} & n^{1/2}\left(S_{n,\theta_0}(\theta_0)(\tilde{a}) - S_{\theta_0}(\theta_0)(\tilde{a})\right) + o_p(1) \\ & \quad = n^{1/2}\nabla_{\hat{\theta}_n-\theta_0}(\theta_0)(\tilde{a}) = n^{1/2}\nabla_{\hat{\theta}_n-\theta_0}(\theta_0)\{\sigma^{-1}(a)\}, \\ & \quad = n^{1/2}\left\{-(\hat{\beta}_{1,n}-\beta_{10})a_{11} - (\hat{\beta}_{2,n}-\beta_{20})a_{12} - (\hat{\alpha}-\alpha_0)a_2 - \int_0^\tau a_3(u)d(\hat{\Lambda}-\Lambda_0)(u)\right\}. \quad \text{(A3)} \end{aligned}$$

By taking $a_3 = 0$ in (A3) and solving for the corresponding $\tilde{a}$ such that $\sigma(\tilde{a}) = (a_{11}, a_{12}, a_2, 0)$, a simple calculation shows that the influence function of $(\hat{\beta}_1 a_{11}, \hat{\beta}_2 a_{12}, \hat{\alpha} a_2)$ lies within the linear span of the tangent (score) space. The semiparametric efficiency of $(\hat{\beta}_1, \hat{\beta}_2, \hat{\alpha})$ is then implied.

Proof of Theorem 3 and 4

Denote the expectations taken with respect to the empirical and true distribution by P_n and P_0 respectively. Two technical assumptions are made below to validate the likelihood ratio tests proposed in Theorem 3 and Theorem 4. These are

Assumption 7. $n^{1/2}P_n\left[\frac{\partial l\{\Delta\beta, \hat{\phi}(\Delta\beta^0)\}}{\partial \Delta\beta}\,|_{\Delta\beta^0}\right] = o_p(1)$.

Assumption 8. $n^{1/2}P_n\left[\frac{\partial l\{\beta_1, \hat{\phi}(\beta_1^0)\}}{\partial \beta_1}\,|_{\beta_1^0}\right] = o_p(1)$.

Following a similar scheme as the proof for Theorem 3.1 in Murphy & van der Vaart (1997), the likelihood ratio statistics can be written as

$$\xi = 2nP_n\left[l(\hat{\theta}) - l\{\Delta\beta^0, \hat{\phi}(\Delta\beta^0)\}\right]. \quad \text{(A4)}$$

By the definition of $\hat{\phi}(\Delta\beta^0)$, $l\{\Delta\beta^0, \hat{\phi}(\Delta\beta^0)\} \geq l\{\Delta\beta^0, (\hat{\beta}_1, \hat{\alpha}, \hat{\lambda}_0)\}$, so (A4) implies

$$\begin{aligned}\xi &\leq 2nP_n \left[l(\hat{\theta}) - l\{\Delta\beta^0, (\hat{\beta}_1, \hat{\alpha}, \hat{\lambda}_0)\} \right], \\ &= 2nP_n \left[- \left. \frac{\partial l(\theta)}{\partial \Delta\beta} \right|_{\hat{\theta}} (\Delta\beta^0 - \Delta\hat{\beta}) \right. , \\ &\left. -\frac{1}{2} \left. \frac{\partial^2 l\{\Delta\beta, (\hat{\beta}_1, \hat{\alpha}, \hat{\lambda}_0)\}}{\partial \Delta\beta \partial \Delta\beta^{\mathrm{T}}} \right|_{\Delta\tilde{\beta}} (\Delta\beta^0 - \Delta\hat{\beta})^2 \right], \end{aligned} \tag{A5}$$

for some $\Delta\tilde{\beta}$ between $\Delta\hat{\beta}$ and $\Delta\beta^0$. Since $\hat{\theta}$ is the maximum likelihood estimator, the linear term of the right-hand side of (A5) is equal to zero. It is not difficult to show that the second term converges in distribution, under the null hypothesis, to a χ^2 distribution with degrees of freedom equal to the dimension of $\Delta\beta$.

For the other direction, we have $l(\hat{\theta}) \geq l\{\Delta\hat{\beta}, \hat{\phi}(\Delta\beta^0)\}$, so (A4) implies

$$\begin{aligned}\xi &\geq 2nP_n \left[l\{\Delta\hat{\beta}, \hat{\phi}(\Delta\beta^0)\} - l\{\Delta\beta^0, \hat{\phi}(\Delta\beta^0)\} \right], \\ &= 2n^{1/2}P_n \left. \frac{\partial l\{\Delta\beta, \hat{\phi}(\Delta\beta^0)\}}{\partial \Delta\beta} \right|_{\Delta\beta^0} n^{1/2}(\Delta\hat{\beta} - \Delta\beta^0) \\ &\left. -\left[-nP_n \left. \frac{\partial^2 l\{\Delta\beta, \hat{\phi}(\Delta\beta^0)\}}{\partial \Delta\beta \partial \Delta\beta^{\mathrm{T}}} \right|_{\Delta\tilde{\beta}} (\Delta\hat{\beta} - \Delta\beta^0)^2 \right]. \end{aligned} \tag{A6}$$

By the consistency of the proposed estimators and Assumption 7, under the null, the first term converges in probability to 0, while the asymptotic normality leads to the convergence in distribution of the second term to a χ^2 distribution with degrees of freedom as the dimension of $\Delta\beta$. The combination of (A5) and (A6) completes the proof of Theorem 3. The proof of Theorem 4 can be obtained analogously as Theorem 3.

REFERENCES

BOOTH, J. G. & HOBERT, J. P. (1999). Maximizing Generalized Linear Mixed Model Likelihoods with an Automated Monte Carlo EM Algorithm. *J. R. Statist. Soc.* B **61**, 265–85.

BROWN, E. R., IBRAHIM, J. G. & DEGRUTTOLA, V. (2005). A Flexible B-spline Model for Multiple Longitudinal Biomarkers and Survival. *Biometrics* **61**, 64–73.

BYCOTT, P. W. & TAYLOR, J. M. G. (1998). An Evaluation of a Measure of the Proportion of the Treatment Effect Explained by a Surrogate Marker. *Control Clin. Trials* **19**, 555–68.

CAFFO, B., JANK, W. & JONES, G. (2005). Ascent-based Monte Carlo EM. *J. R. Statist. Soc.* B **67**, 235–52.

CHEN, Y. Q. & JEWELL, N. P. (2001). On a General Class of Semiparametric Hazards Regression Models. *Biometrika* **88**, 687–702.

CHEN, Y. Q. & WANG, M. C. (2000). Analysis of accelerated hazards models. *J. Am. Statist. Ass.* **95**, 608–18.

COX, D. R. (1975). Partial Likelihood. *Biometrika* **62**, 269–76.

DING, J. & WANG, J. L. (2008). Modeling Longitudinal Data with Nonparametric Multiplicative Random Effects Jointly with Survival Data. *Biometrics* **64**, 546–56.

ETEZADI-AMOLI, J. & CAMPI, A. (1987). Extended Hazard Regression for Censored Survival Data with Covariates: A Spline Approximation for the Baseline Hazard Function. *Biometrics, B* **43**, 181–92.

FITZMAURICE, G., DAVIDIAN, M., VERBEKE, G. & Molenberghs, G. (2008). *Longitudinal Data Analysis: Handbooks of Modern Statistical Methods*. Chapman and Hall/CRC.

GONG, G. & SAMANIEGO, F. J. (1981). Pseudo Maximum Likelihood Estimation: Theory and Applications. *Ann. Statist.* **9**, 861–69.

GRAMBSCH, P. M. & THERNEAU, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, **81**, 515–26.

HENDERSON, R., DIGGLE, P. J. & DOBSON, A. (2000). Joint Modelling of Longitudinal Measurements and Event Time Data. *Biostatistics* **4**, 465–80.

HSIEH, F., TSENG, Y. K. & WANG, J. L. (2006). Joint Modeling of Survival Time and Longitudinal Data: Likelihood Approach Revisit. *Biometrics* **62**, 1037–43.

JONES, M. C. (1990). The Performance of Kernel Density Functions in Kernel Distribution Function Estimation. *Statist. Prob. Lett.* **9**, 129–32.

JONES, M. C. & SHEATHER, S. J. (1991). Using Non-stochastic Terms to Advantage in Kernel-based Estimation of Integrated Squared Density Derivatives. *Statist. Prob. Lett.* **11**, 511–14.

LOUIS, T. A. (1982). Finding the Observed Information Matrix when Using the EM Algorithm. *J. R. Statist. Soc.* B **44**, 226–33.

MURPHY, S. A. & VAN DER VAART, A. W. (1997). Semiparametric Likelihood Ratio Inference. *Ann. Statist.* **25**, 1471–509.

PAWITAN, Y. & SELF, S. (1993). Modeling Disease Marker Processes in AIDS. *J. Am. Statist. Ass.* **88**, 719–26.

RIZOPOULOS, D., VERBEKE, G. & MOLENBERGHS, G. (2008). Shared Parameter Models under Random Effects Misspecification. *Biometrika* **95**, 63–74.

SONG, X., DAVIDIAN, M. & TSIATIS, A. A. (2002). A Semiparametric Likelihood Approach to Joint Modelling of Longitudinal and Time-to-event Data. *Biometrics* **58**, 742–53.

THERNEAU, T. M. & GRAMBSCH, P. M. (2000). *Modeling Survival Data*. New York: Springer.

TSENG, Y. K., HSIEH, F., & WANG, J. L. (2005). Joint Modeling of Accelerated Failure Time and Longitudinal Data. *Biometrika* **92**, 587–603.

TSENG, Y. K. & SHU, K. N. (2011). Efficient Estimation for a Semiparametric Extended Hazards Model. *Commun. Statist.- Simul. Comput.* **40**, pp.270–85.

TSENG, Y. K., HSU, K. N. & YANG, Y. F. (2014). A Semiparametric Extended Hazard Regression Model with Time-dependent Covariates. *J. Nonpara. Statist.* **26**, 115–28.

TSIATIS, A. A. & DAVIDIAN, M. (2001). A Semiparametric Estimator for the Proportional Hazards Model with Longitudinal Covariates Measured with Error. *Biometrika* **88**, 447–58.

TSIATIS, A. A. & DAVIDIAN, M. (2004). Joint Modelling of Longitudinal and Time-to-event Data: An Overview. *Statist. Sin.* **14**, 809–34.

TSIATIS, A. A., DEGRUTTOLA, V. & WULFSOHN, M. S. (1995). Modelling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *J. Am. Statist. Ass.* **90**, 27–37.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

VERBEKE, G. & DAVIDIAN, M. (2008). Joint Models for Longitudinal Data: Introduction and Overview. *Longitudinal Data Analysis: HandBooks of Modern Statistical Methods* Ed. Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G., 319–26. Chapman and Hall/CRC.

WANG, Y. & TAYLOR, J. M. G. (2001). Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *J. Am. Statist. Ass.* **96**, 895–905.

WEI, G. C. G. & TANNER, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and Poor Man's Data Augmentation Algorithm. *J. Am. Statist. Ass.* **85**, 699–704.

WULFSOHN, M. S. & TSIATIS, A. A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error. *Biometrics*, **53**, 330–39.

ZENG, D. & LIN, D. Y. (2007). Efficient Estimation in the Accelerated Failure Time Model. *J. Am. Statist. Ass.* **102**, 1387–96.