Acta Mathematica Sinica, English Series
Mar., 2015, Vol. 31, No. 3, pp. 511–525
Published online: February 15, 2015
DOI: 10.1007/s10114-015-3628-7
Http://www.ActaMath.com

Acta Mathematica Sinica, English Series

Additive Hazards Regression with Random Effects for Clustered Failure Times

Deng PAN
Department of Statistics, Chinese University of Hong Kong, Hong Kong, P. R. China
E-mail: *dpan@cuhk.edu.hk*

Yan Yan LIU **Yuan Shan WU**[1)]
School of Mathematics and Statistics, Wuhan University, Wuhan 430072*, P. R. China*
E-mail: *liuyy@whu.edu.cn* *shan@whu.edu.cn*

Abstract Additive hazards model with random effects is proposed for modelling the correlated failure time data when focus is on comparing the failure times within clusters and on estimating the correlation between failure times from the same cluster, as well as the marginal regression parameters. Our model features that, when marginalized over the random effect variable, it still enjoys the structure of the additive hazards model. We develop the estimating equations for inferring the regression parameters. The proposed estimators are shown to be consistent and asymptotically normal under appropriate regularity conditions. Furthermore, the estimator of the baseline hazards function is proposed and its asymptotic properties are also established. We propose a class of diagnostic methods to assess the overall fitting adequacy of the additive hazards model with random effects. We conduct simulation studies to evaluate the finite sample behaviors of the proposed estimators in various scenarios. Analysis of the Diabetic Retinopathy Study is provided as an illustration for the proposed method.

Keywords Additive hazards regression, clustered failure times, counting process, empirical process, frailty, model checking, random effects

MR(2010) Subject Classification 62N01, 62N02

1 Introduction

Multivariate failure time data are encountered frequently in scientific investigation, in which the study subjects are sampled in clusters and the failure times within the same cluster tend to be correlated. Much work in this context has focused on the marginal hazards models, including but not limited to, the marginal proportional hazards model (Spiekerman and Lin [17]) and the marginal additive hazards model (Yin and Cai [20]), which do not specify the dependence structure but adjusts for it in inference. When the intracluster dependence is of interest as well as the effects of covariates on the failure times, a useful approach is to incorporate a random effect or frailty into the hazards regression to describe the intracluster dependence. In the

Received November 5, 2013, revised May 28, 2014, accepted July 25, 2014
Supported by National Natural Science Foundation of China (Grant Nos. 11171263, 11201350 and 11371299), Doctoral Fund of Ministry of Education of China (Grant Nos. 20110141110004 and 20110141120004), and Fundamental Research Funds for the Central Universities
1) Corresponding author

proportional hazards regression, the hazard function for the k-th subject of the i-th cluster associated with covariates $Z_{ik}(\cdot)$ takes the form

$$\lambda(t|Z_{ik};\xi_i) = \xi_i\lambda_0(t)\mathrm{e}^{\beta^{\mathrm{T}}Z_{ik}(t)}, \tag{1.1}$$

where β is a vector of unknown regression parameters, $\lambda_0(\cdot)$ is an unspecified baseline hazard function, and ξ_i is an unobserved random effect for the i-th cluster, which induces the dependence between subjects within the cluster.

Model (1.1) with Gamma random effect has been investigated rigorously by Murphy [13, 14] for the case without covariates, and by Parner [15] for the case with covariates; Model (1.1) with positive stable random effect has been studied by Fine et al. [2] and Martinussen and Pipper [11]. Recently, Zeng and Lin [21] proposed a general class of semiparametric transformation models with random effects based on the model (1.1), which includes the Cox proportional hazards and proportional odds models with random effects as its special cases.

Although various hazards regression models with random effects have been extensively studied and applied in practice, we are not aware of the studies on the additive hazards model with random effects for survival analysis. Both the proportional and additive hazards models have sound biological and empirical bases, and construct two principal frameworks for studying the association between covariates and disease occurrence or death. In contrast to the proportional hazards model, the additive hazards model specifies that the hazards function associated with a set of possibly time-dependent covariates is the sum of, rather than the product of, the baseline hazard function and the regression function of covariates (see Lin and Ying [10]). The additive hazards model offers a valuable alternative to the proportional hazards model when the investigator is interested in the hazards difference instead of the hazards ratio or the proportional hazards assumption is violated in practice. Consequently, it is imperative to establish the estimate method for the additive hazards model with random effects. To this point, in this article we develop the statistical inference for fitting the correlated failure time data using the additive hazards model with random effects. Specifically, we propose the additive hazards model with random effects as follows:

$$\lambda(t|Z_{ik}, X_{ik};\xi_i) = \lambda_0(t) + \beta^{\mathrm{T}}Z_{ik}(t) + \xi_i^{\mathrm{T}}X_{ik}(t), \tag{1.2}$$

where β is a p-vector of unknown regression parameters, $\lambda_0(\cdot)$ is a completely unspecified function, ξ_i is an unobserved positive frailty variable for the i-th cluster, Z_{ik} and X_{ik} are the p-vector and q-vector covariate processes, respectively, associated with the fixed and random effects. To identify the proposed model (1.2), we require that $Z_{ik}(t)$ and $X_{ik}(t)$ do not share any components and that neither $Z_{ik}(t)$ nor $X_{ik}(t)$ contains intercept term, where the intercept term means the deterministic function of t. Our model (1.2) is different from the partly parametric version of the Aalen additive model proposed by McKeague and Sasieni [12], in which the effect of some covariates varies nonparametrically over time and that of the remaining is constant. Our model (1.2) is akin to the mixed effect model in the longitudinal data analysis. Thus, the parameters in our model possess similar interpretations as in the usual mixed effect models, but in terms of hazard risks.

The remainder of this article is organized as follows. In Section 2, we introduce some notation and derive the estimation procedures for the model parameters. Section 3 establishes the asymptotic properties for proposed estimators with proofs relegated to Section 8. Model checking procedures are provided in Section 4. We conduct simulation studies in Section 5 to evaluate the finite-sample behavior of the asymptotic approximation. The proposed methods are illustrated with the Diabetic Retinopathy Study in Section 6. Some concluding remarks are given in Section 7.

2 Estimation

Let $\widetilde{T}_{ik}$ and C_{ik} denote the failure time and censoring time, respectively, for the k-th individual in the i-th cluster, $i = 1, 2, \ldots, n, k = 1, 2, \ldots, K$. Correspondingly, let $T_{ik} = \min(\widetilde{T}_{ik}, C_{ik})$ be the observed time and denote the censoring indicator by $\Delta_{ik} = I(\widetilde{T}_{ik} \le C_{ik})$, where $I(\,\cdot\,)$ is the indicator function. Let τ denote the end time of study. Assume that $\{\widetilde{T}_i, C_i, \xi_i, Z_i(t), X_i(t)\colon t \in [0,\tau]\}$ are independent and identically distributed (i.i.d.) for $i = 1, 2, \ldots, n$, where $\widetilde{T}_i = (\widetilde{T}_{i1}, \ldots, \widetilde{T}_{iK})$, and $C_i, Z_i(t)$, and $X_i(t)$ are defined in the same manner. Also, we assume that, given the covariates in the i-th cluster, the censoring C_i is independent of ξ_i and $\widetilde{T}_i$ and that $\widetilde{T}_{ik}$ for $k = 1, 2, \ldots, K$ are independent given $\xi_i, Z_i(\,\cdot\,)$, and $X_i(\,\cdot\,)$. The counting process is denoted by $N_{ik}(t) = I(T_{ik} \le t)\Delta_{ik}$ and the at-risk process by $Y_{ik}(t) = I(T_{ik} \ge t)$.

Let the distribution of ξ be P_α indexed by an unknown parameter α, whose dimension is assumed to be equal to that of ξ to ensure the identifiability. Denote the Laplace transform by $G_\alpha(u) = E_\alpha\{\exp(-\xi^{\mathrm{T}} u)\}$, where $E_\alpha(\,\cdot\,)$ is the expectation with respect to P_α. An interesting feature of the model (1.2) is that, when marginalized over the random effect ξ_i, it still maintains the structure of additive hazards regression. Specifically, noting that the marginal survival function for the k-th subject in the i-th cluster is

$$\begin{aligned}&P(T_{ik} > t | Z_{ik}, X_{ik})\\&\quad = \exp\left\{-\int_0^t \lambda_0(u)du - \beta^{\mathrm{T}} \int_0^t Z_{ik}(u)du\right\} E_\alpha\left[\exp\left\{-\xi_i^{\mathrm{T}} \int_0^t X_{ik}(u)du\right\}\right],\end{aligned}$$

we can obtain the corresponding marginal hazards function as follows:

$$\lambda(t|Z_{ik}, X_{ik}) = \lambda_0(t) + \beta^{\mathrm{T}} Z_{ik}(t) + H_\alpha\{\bar{X}_{ik}(t)\}^{\mathrm{T}} X_{ik}(t), \tag{2.1}$$

where $H_\alpha(u) = -\partial\{\log G_\alpha(u)\}/\partial u$ and $\bar{X}_{ik}(t) = \int_0^t X_{ik}(u)du$. For example, if ξ is from the gamma distribution with mean μ and variance σ^2, then $H_\alpha(u) = \frac{\alpha^2}{\alpha+\sigma^2 u}$ if one takes $\alpha = \mu$.

We aim to make inference, based on the observed data $\{T_i, \Delta_i, Z_i(\,\cdot\,), X_i(\,\cdot\,)\}$ $(i = 1, 2, \ldots, n)$, about the regression parameter $\theta \equiv (\beta^{\mathrm{T}}, \alpha^{\mathrm{T}})^{\mathrm{T}}$ of interest and the cumulative hazards function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$ under the model (2.1). Let

$$M_{ik}(t,\theta) = N_{ik}(t) - \int_0^t Y_{ik}(u)\lambda(u|Z_{ik}, X_{ik})du$$

and θ_0 denote the true value of θ. Then $M_{ik}(t,\theta_0)$ is a counting process martingale (Fleming and Harrington [3]), which is denoted by $M_{ik}(t)$ in what follows for short.

Noting that $E\{dM_{ik}(t)|Z_{ik}(t), X_{ik}(t)\} = 0$ and as elucidated by Lin and Ying [10], we specify the estimating equations as follows:

$$\sum_{i=1}^n \sum_{k=1}^K \int_0^t M_{ik}(du) = 0, \tag{2.2}$$

and

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^{\tau} Q_{ik}(u,\theta)M_{ik}(du,\theta)=0, \tag{2.3}$$

for the cumulative baseline function $\Lambda_0(t)$ and the regression parameter θ, respectively. $Q_{ik}(t,\theta)$ is a smooth (with respect to θ) $(p+q)$-vector-valued function of $Z_{ik}(t)$, $X_{ik}(t)$ and θ, but not involving $\Lambda_0(t)$.

Solving (2.2) with given θ_0, the Aalen–Breslow type estimator for $\Lambda_0(t)$ is given by $\widehat{\Lambda}_n(t,\theta_0)$, where

$$\widehat{\Lambda}_n(t,\theta)=\int_0^t \frac{\sum_{i=1}^n\sum_{k=1}^K\left[dN_{ik}(u)-Y_{ik}(u)\left\{\beta^{\mathrm{T}}Z_{ik}(u)+H_\alpha(\bar{X}_{ik}(u))^{\mathrm{T}}X_{ik}(u)\right\}du\right]}{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(u)}.$$

Substituting $\widehat{\Lambda}_n(t,\theta)$ for $\Lambda_0(t)$ in (2.3), the resultant estimating equation for θ is then given by $\mathbf{U}(\theta,\tau)=0$, where

$$\begin{aligned}\mathbf{U}&(\theta,t)\\&=\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\{Q_{ik}(u,\theta)-\bar{Q}(u,\theta)\}[dN_{ik}(u)-Y_{ik}(u)\{\beta^{\mathrm{T}}Z_{ik}(u)+H_\alpha(\bar{X}_{ik}(u))^{\mathrm{T}}X_{ik}(u)\}du]\\&=\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\{Q_{ik}(u,\theta)-\bar{Q}(u,\theta)\}M_{ik}(du,\theta)\end{aligned}$$

and

$$\bar{Q}(t,\theta)=\frac{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(t)Q_{ik}(t,\theta)}{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(t)}.$$

The solution to $\mathbf{U}(\theta,\tau)=0$, denoted by $\widehat{\theta}$, is used to as an estimate for θ_0. A natural estimator for $\Lambda_0(t)$ is $\widehat{\Lambda}_n(t,\widehat{\theta})$. To ensure monotonicity, we make a minor modification, that is, $\widehat{\Lambda}_n^*(t)\equiv\max_{0\le u\le t}\widehat{\Lambda}_n(u,\widehat{\theta})$. Following similar arguments in Lin and Ying [10], we have that $\widehat{\Lambda}_n^*(t)-\widehat{\Lambda}_n(t,\widehat{\theta})=o_p(n^{-\frac{1}{2}})$, uniformly in $t\in[0,\tau]$. Usually, we may choose

$$Q_{ik}(t,\theta)=\begin{bmatrix} Z_{ik}(t)\\ h_\alpha(\bar{X}_{ik}(t))X_{ik}(t)\end{bmatrix} \tag{2.4}$$

for simplicity, where $h_\alpha(t)=\frac{\partial H_\alpha(t)^{\mathrm{T}}}{\partial\alpha}$.

3 Asymptotic Properties

In this section, we establish the asymptotic properties of the proposed estimators. We impose the following regularity conditions throughout our discussion.

(C1) $\{N_i(\cdot),Y_i(\cdot),Z_i(\cdot),X_i(\cdot),\xi_i,Q_i(\cdot,\theta)\}$ are i.i.d. for $i=1,2,\dots,n$, where N_i, Y_i and Q_i are defined in the similar manner as Z_i. The parameter space, denoted by Θ, is compact and contains the true value θ_0 as its interior point.

(C2) $P(Y_{ik}(t)=1$, for all $t\in[0,\tau])>0$ for $k=1,2,\dots,K$ and $i=1,2,\dots,n$.

(C3) For $k=1,2,\dots,K$ and $i=1,2,\dots,n$, both $Z_{ik}(\cdot)$ and $X_{ik}(\cdot)$ have bounded total variations. $Q_{ik}(\cdot,\theta)$ has bounded total variation uniformly in $\theta\in\Theta$.

(C4) The matrix A is nonsingular, where

$$A = E\Big[\sum_{k=1}^{K}\Big\{\int_0^\tau Y_{1k}(u)\{Q_{1k}(u,\theta_0)-\bar{q}(u,\theta_0)\}\begin{pmatrix} Z_{1k}(u)du \\ h_{\alpha_0}(\bar{X}_{1k}(u))X_{1k}(u)du\end{pmatrix}^{\mathrm{T}}\Big\}\Big]$$

with $\bar{q}(t,\theta) = \frac{E\{\sum_{k=1}^K Y_{1k}(t)Q_{1k}(t,\theta)\}}{E\{\sum_{k=1}^K Y_{1k}(t)\}}$.

(C5) The class

$$\Big\{\frac{\partial Q_{ik}(t,\cdot)}{\partial\theta}, H_\alpha(\bar{X}_{ik}(t)), h_\alpha(\bar{X}_{ik}(t))X_{ik}(t)\colon t\in[0,\tau], k=1,2,\ldots,K, i=1,2,\ldots,n\Big\}$$

are equicontinuous and bounded uniformly in parameter space Θ.

(C6) $H_\alpha(\bar{X}) = H_{\alpha_0}(\bar{X})$ for almost every $\bar{X}$ implies that $\alpha=\alpha_0$. If $h_{\alpha_0}(\bar{X})v_\alpha = 0$ almost everywhere for $\bar{X}$, then $v_\alpha = 0$.

Let $\mathbf{S}_k^{(0)}(t,\theta) = n^{-1}\sum_{i=1}^n Y_{ik}(t)$, $\mathbf{S}_k^{(1)}(t,\theta) = n^{-1}\sum_{i=1}^n Y_{ik}(t)[Z_{ik}^{\mathrm{T}}(t), X_{ik}(t)^{\mathrm{T}}h_\alpha(\bar{X}_{ik}(t))^{\mathrm{T}}]^{\mathrm{T}}$, and write $\mathbf{E}(t,\theta) = \sum_{k=1}^K \mathbf{S}_k^{(1)}(t,\theta)\{\sum_{k=1}^K \mathbf{S}_k^{(0)}(t,\theta)\}^{-1}$. Denote the limiting values of $\mathbf{S}_k^{(0)}(t,\theta)$, $\mathbf{S}_k^{(1)}(t,\theta)$, and $\mathbf{E}(t,\theta)$ by $\mathbf{s}_k^{(0)}(t,\theta), \mathbf{s}_k^{(1)}(t,\theta)$, and $\mathbf{e}(t,\theta)$, respectively. Conditions (C1), (C2), and (C5) imply that as $n\to\infty$ for $k=1,2,\ldots,K$ and $d=0,1$,

$$\sup_{(t,\theta)\in[0,\tau]\times\Theta}\|\mathbf{S}_k^{(d)}(t,\theta)-\mathbf{s}_k^{(d)}(t,\theta)\| \to_{\text{a.s.}} 0, \qquad \sup_{(t,\theta)\in[0,\tau]\times\Theta}\|\mathbf{E}(t,\theta)-\mathbf{e}(t,\theta)\| \to_{\text{a.s.}} 0,$$
$$\sup_{(t,\theta)\in[0,\tau]\times\Theta}\|\bar{Q}(t,\theta)-\bar{q}(t,\theta)\| \to_{\text{a.s.}} 0,$$

where $\|\mathbf{a}\|$ is defined as the maximum norm for a vector or matrix $\mathbf{a}$.

We summarize the asymptotic properties of $\widehat{\theta}$ and $\widehat{\Lambda}_n(t,\widehat{\theta})$ in the following theorems.

Theorem 3.1 *Under conditions* (C1)–(C6), $\widehat{\theta}$ *converges almost surely to* θ_0, *while* $\sqrt{n}(\widehat{\theta}-\theta_0)$ *converges weakly to a normal distribution with mean* 0 *and covariance* $A^{-1}\Sigma(A^{\mathrm{T}})^{-1}$, *where* $\Sigma = \sum_{j=1}^K\sum_{k=1}^K \Sigma_{jk}(\tau,\tau)$ *with*

$$\Sigma_{jk}(s,t) = E\left[\int_0^s\{Q_{1j}(u,\theta_0)-\bar{q}(u,\theta_0)\}dM_{1j}(u)\int_0^t\{Q_{1k}(v,\theta_0)-\bar{q}(v,\theta_0)\}^{\mathrm{T}}dM_{1k}(v)\right]$$

for s and t in $[0,\tau]$.

The asymptotic covariance can be estimated empirically by replacing the limiting values with their empirical counterparts, say $\widehat{A}^{-1}\widehat{\Sigma}(\widehat{A}^{\mathrm{T}})^{-1}$, where

$$\widehat{A} = n^{-1}\sum_{i=1}^n\sum_{k=1}^K\left[\int_0^\tau Y_{ik}(u)\{Q_{ik}(u,\widehat{\theta})-\bar{Q}(u,\widehat{\theta})\}\begin{Bmatrix} Z_{ik}(u)du \\ h_{\widehat{\alpha}}(\bar{X}_{ik}(u))X_{ik}(u)du\end{Bmatrix}^{\mathrm{T}}\right],$$

and $\widehat{\Sigma} = \sum_{j=1}^K\sum_{k=1}^K\widehat{\Sigma}_{jk}(\tau,\tau)$ with

$$\widehat{\Sigma}_{jk}(s,t) = n^{-1}\sum_{i=1}^n\left[\int_0^s\{Q_{ij}(u,\widehat{\theta})-\bar{Q}(u,\widehat{\theta})\}d\widehat{M}_{ij}(u)\int_0^t\{Q_{ik}(v,\widehat{\theta})-\bar{Q}(v,\widehat{\theta})\}^{\mathrm{T}}d\widehat{M}_{ik}(v)\right]$$

and

$$d\widehat{M}_{ik}(t) = dN_{ik}(t) - Y_{ik}(t)\big\{d\widehat{\Lambda}_n(t,\widehat{\theta}) + \widehat{\beta}^{\mathrm{T}}Z_{ik}(t)dt + H_{\widehat{\alpha}}(\bar{X}_{ik}(t))^{\mathrm{T}}X_{ik}(t)dt\big\}.$$

The proof of consistency in Theorem 3.1 involves several applications of the strong law of large numbers and the inverse function theorem. The proof of asymptotic normality follows from that $n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)$ asymptotically behaves as a scaled normalized sum of independent and identically distributed random vectors. The proof of Theorem 3.1 is provided in Section 8.

Theorem 3.2 *Under conditions* (C1)–(C6), $\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t)$ *converges almost surely to* 0, *uniformly in* $t\in[0,\tau]$, *while* $n^{\frac{1}{2}}\{\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t)\}$ *converges weakly to a zero-mean Gaussian process with covariance function* $\psi(s,t)=E[\Psi_1(s)\Psi_1(t)]$, *where*

$$\Psi_i(t)=\int_0^t\frac{\sum_{k=1}^K dM_{ik}(u)}{\sum_{k=1}^K \mathbf{s}_k^{(0)}(u,\theta_0)}-\int_0^t \mathbf{e}(v,\theta_0)^{\mathrm{T}}dvA^{-1}\sum_{k=1}^K\left[\int_0^\tau\{Q_{ik}(u,\theta_0)-\bar{q}(u,\theta_0)\}dM_{ik}(u)\right].$$

The covariance function $\psi(s,t)$ can be estimated by replacing limiting quantities in $\Psi_i(t)$ with their respectively empirical counterparts. Specifically, $\widehat{\psi}(s,t)=n^{-1}\sum_{i=1}^n\widehat{\Psi}_i(s)\widehat{\Psi}_i(t)$, where

$$\widehat{\Psi}_i(t)=\int_0^t\frac{\sum_{k=1}^K d\widehat{M}_{ik}(u)}{\sum_{k=1}^K \mathbf{S}_k^{(0)}(u,\widehat{\theta})}-\int_0^t \mathbf{E}(v,\widehat{\theta})^{\mathrm{T}}dv(\widehat{A})^{-1}\sum_{k=1}^K\left[\int_0^\tau\{Q_{ik}(u,\widehat{\theta})-\bar{Q}(u,\widehat{\theta})\}d\widehat{M}_{ik}(u)\right].$$

Theorem 3.2 can be proved based on a decomposition and using the uniform strong law of large numbers (Pollard [16]) and various empirical process results (van der Vaart and Wellner [18]). The proof of Theorem 3.2 is given in Section 8.

4 Model Checking Techniques

In this section, we develop methods for assessing the adequacy of the model (1.2). Denote $W_{ik}(t)=\{Z_{ik}(t)^{\mathrm{T}},X_{ik}(t)^{\mathrm{T}}\}^{\mathrm{T}}$. Mimicking idea proposed by Lin et al. [9], we consider the following multiparameter stochastic process, which involves various functional forms of the cumulative sums of residual $\widehat{M}_{ik}(s;\widehat{\theta})$:

$$\mathbf{G}_n(t,x;\theta)=\sum_{i=1}^n\sum_{k=1}^K\int_0^t \mathbf{f}(W_{ik}(s))I(W_{ik}(s)\le x)d\widehat{M}_{ik}(s;\theta),$$

where $\mathbf{f}(\cdot)$ is a known vector-valued bounded function and the event $I(W_{ik}(s)\le x)$ means that each of the components of $W_{ik}(s)$ is no larger than the respective component of x.

Define

$$\mathbf{g}_n(t,x)=\frac{\sum_{i=1}^n\sum_{k=1}^K \mathbf{f}(W_{ik}(t))I(W_{ik}(t)\le x)Y_{ik}(t)}{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(t)}$$

and

$$\begin{aligned}&\mathbf{h}_n(t,x;\theta)\\&=n^{-1}\sum_{i=1}^n\sum_{k=1}^K\int_0^t \mathbf{f}(W_{ik}(s))I(W_{ik}(s)\le x)Y_{ik}(s)\left[\begin{Bmatrix} Z_{ik}(s)\\ h_\alpha(\bar{X}_{ik}(s))X_{ik}(s)\end{Bmatrix}-\mathbf{E}(s,\theta)\right]ds.\end{aligned}$$

Denote the limits of $\mathbf{g}_n(t,x)$ and $\mathbf{h}_n(t,x;\theta)$ respectively by $\tilde{\mathbf{g}}(t,x)$ and $\tilde{\mathbf{h}}(t,x;\theta)$. Using the Taylor series expansions of $\mathbf{G}_n(t,x;\widehat{\theta})$ and $\mathbf{U}(\widehat{\theta},\tau)$ around θ_0, we have that $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ is asymptotically equivalent to $n^{-\frac{1}{2}}\tilde{\mathbf{G}}(t,x;\theta_0)$, where

$$\tilde{\mathbf{G}}(t,x;\theta)=\sum_{i=1}^n\Phi_i(t,x;\theta)$$

and

$$\Phi_i(t,x;\theta) = \sum_{k=1}^{K}\int_0^t \{\mathbf{f}(W_{ik}(s))I(W_{ik}(s)\le x) - \tilde{\mathbf{g}}(s,x)\}\,dM_{ik}(s,\theta)$$
$$-\tilde{\mathbf{h}}(t,x;\theta)A^{-1}\sum_{k=1}^{K}\int_0^\tau \{Q_{ik}(s,\theta)-\bar{q}(s,\theta)\}dM_{ik}(s,\theta).$$

Theorem 4.1 *Under conditions* (C1)–(C6), $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ *converges weakly to a zero-mean Gaussian random field with the covariance function between* (t,x) *and* (t^*,x^*) *given by* $E\{\Phi_1(t,x;\theta_0)\Phi_1^{\mathrm{T}}(t^*,x^*;\theta_0)\}$.

The key steps in the proof are to verify the finite-dimensional distribution convergence and the tightness condition as outlined in Section 8. Furthermore, the covariance function can be consistently estimated by $n^{-1}\sum_{i=1}^{n}\widehat{\Phi}_i(t,x;\widehat{\theta})\widehat{\Phi}_i^{\mathrm{T}}(t^*,x^*;\widehat{\theta})$, where

$$\widehat{\Phi}_i(t,x;\theta) = \sum_{k=1}^{K}\int_0^t \{\mathbf{f}(W_{ik}(s))I(W_{ik}(s)\le x) - \mathbf{g}_n(s,x)\}\,dM_{ik}(s,\theta)$$
$$-\mathbf{h}_n(t,x;\theta)A^{-1}\sum_{k=1}^{K}\int_0^\tau \{Q_{ik}(s,\theta)-\bar{Q}(s,\theta)\}dM_{ik}(s,\theta).$$

Using Theorem 4.1, we can simulate the limiting distribution of $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ to test the goodness of fit through the resampling approach. Specifically, we simulate n i.i.d. observations, say $\eta_1,\dots,\eta_n$, from the standard normal distribution and then obtain the perturbed version of the stochastic process $\widehat{\mathbf{G}}(t,x;\theta)=\sum_{i=1}^{n}\widehat{\Phi}_i(t,x;\theta)\eta_i$.

The next theorem provides the theoretical justification for this perturbing procedure.

Theorem 4.2 *Given the observed data* $\{(N_{ik}(t),Y_{ik}(t),Z_{ik}(t),X_{ik}(t))\colon t\in[0,\tau]; i=1,2,\dots,n; k=1,2,\dots,K\}$, $n^{-\frac{1}{2}}\widehat{\mathbf{G}}(t,x;\widehat{\theta})$ *converges weakly to the same zero-mean Gaussian random field as that of* $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$.

This theorem was also discussed by Lin et al. [9] and Yin [19]. The critical argument is that conditional on the observed data, $\widehat{\mathbf{G}}(t,x;\widehat{\theta})$ can be viewed as a sum of the independent random variables for each fixed time t and x. To save space, we omit the proof of Theorem 4.2.

We illustrate that how $\mathbf{G}_n(t,x;\widehat{\theta})$ can be used for different purposes of model checking in the following derivations. To check the functional forms of a covariate, e.g., the j-th component of W_{ik}, denoted by W_{ikj}, we take $\mathbf{f}(W_{ik})=1,\ t=\tau$, and let every component except the j-th component of x be ∞, then the resulting testing statistic is

$$\mathbf{G}_n^j(\tau,x;\widehat{\theta}) \equiv \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau I(W_{ikj}(s)\le x_j)d\widehat{M}_{ik}(s;\widehat{\theta}).$$

In order to construct an omnibus test for checking the overall fit of the model, one can take $\mathbf{f}(W_{ik})=1$, then the resulting testing process is

$$\mathbf{G}_n^o(t,x;\widehat{\theta}) \equiv \sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t I(W_{ik}(s)\le x)d\widehat{M}_{ik}(s;\widehat{\theta}).$$

To approximate the distribution of $\mathbf{G}_n^o(t,x;\widehat{\theta})$ one can obtain a large number of realizations from $\widehat{\mathbf{G}}_n^o(t,x;\widehat{\theta})$, by repeatedly generating the standard normal random sample $(\eta_1,\dots,\eta_n)$

while fixing the data $\{(N_{ik}(t), Y_{ik}(t), Z_{ik}(t), X_{ik}(t))\colon t \in [0,\tau]; i = 1,2,\ldots,n; k = 1,2,\ldots,K\}$ at their observed values. Graphically, one can plot a few realizations of $\widehat{\mathbf{G}}_n^o(t,x;\widehat{\theta})$ along with the observed $\mathbf{G}_n^o(t,x;\widehat{\theta})$ to see if they can be regarded as arising from the same population. More formally and objectively, we can apply the supremum test statistic $\sup_{0\le t\le\tau,x}|\mathbf{G}_n^o(t,x;\widehat{\theta})|$ to test the overall fit of the model. The p-value of this test is obtained by generating a large number of realizations from $\sup_{0\le t\le\tau,x}|\widehat{\mathbf{G}}_n^o(t,x;\widehat{\theta})|$ and comparing them with the observed value of $\sup_{0\le t\le\tau,x}|\mathbf{G}_n^o(t,x;\widehat{\theta})|$. A similar resampling approach can be also applied to test statistic $\mathbf{G}_n^j(\tau,x;\widehat{\theta})$.

5 Simulation Study

We conducted simulation studies to assess the adequacy of the asymptotic approximation of the proposed method. Specifically, we generated clustered survival time data using the following model

$$\lambda(t|Z_{ik}, X_{ik}; \xi_i) = 1 + \beta^{\mathrm{T}} Z_{ik} + \xi_i^{\mathrm{T}} X_{ik}.$$

The covariates of Z_{ik} and X_{ik} for each subject in the same cluster were both generated from Unif$(0,1)$. The frailty variables ξ_i were generated from the Gamma or the Inverse Gaussian (IG) distribution both with mean μ and variance σ^2. For simplicity, we take $\alpha = \mu$ while letting σ^2 be fixed. Censoring times were set to be the minimum of a Unif$(0,c)$ distributed variable and τ, where c and τ were chosen to yield an approximately censoring rate of 35% or 60% in all the considered configurations. The estimating function $\mathbf{U}(\theta,\tau)$ was derived by choosing $Q_{ik}(t,\theta)$ according to (2.4).

To investigate the performance of the proposed estimator in practical sample size and the cluster size, we set $n = 200$ and 400 and $K = 2$ and 5 and consider two different censoring rates of 35% and 60% and the covariate effect size $(\beta_0, \alpha_0) = (1,2)$ with σ^2 fixed at 1. The corresponding results of 1000 replications are summarized in Table 1. The sample mean and sample standard deviation of the 1000 estimates are given in the Mean and SD columns, respectively. The SE columns give the average of the estimated standard errors and the CP columns give the coverage probability of the nominal 95% confidence interval for the true parameter using the estimated standard error. When the censoring rate is 60%, it can be seen from Table 1 in both the Gamma and IG frailties that the biases of the proposed estimators are essentially negligible, the estimated standard errors agree well with the sample standard errors, and the coverage probabilities are around the nominal level 95%. More precise estimators are obtained when the cluster size, K, increases from 2 to 5 or the censoring rate decreases to 35%.

Table 2 is presented to summarize the effect of σ^2 on the parameter estimation. It can be seen that the proposed estimator works well in the considered scenarios. When σ^2 is misspecified to be 1, the corresponding results are reported in Table 3. When the frailty is Gamma distributed, the estimate for β_0 performs well and does not show the negative effects arising from the misspecification of σ^2. On the other hand, the estimate for α_0 is still of satisfactory when the cluster size is small. The performance is attenuated along with the cluster size increasing, especially in the case of $\sigma^2 = 2$. This is partly due to that more error is cumulated when more correlations among more cluster members are misspecified. Similar conclusions can be drawn from the IG frailty. Thus, our proposed method is to some extent robust against to the misspecification of σ^2 in the case of small cluster size, which is frequently encountered in practice.

				$\beta_0 = 1$				$\alpha_0 = 2$			
Frailty	Cen.	n	K	Mean	SD	SE	CP	Mean	SD	SE	CP
Gamma	35%	200	2	0.988	0.510	0.516	0.951	2.036	0.521	0.529	0.954
			5	1.004	0.317	0.324	0.950	2.004	0.330	0.336	0.947
		400	2	0.986	0.355	0.363	0.966	2.010	0.373	0.372	0.947
			5	0.990	0.224	0.228	0.952	2.010	0.231	0.237	0.945
	60%	200	2	1.050	0.672	0.670	0.952	2.010	0.667	0.676	0.958
			5	1.000	0.414	0.421	0.961	2.017	0.422	0.428	0.950
		400	2	1.020	0.469	0.471	0.949	1.993	0.474	0.474	0.958
			5	0.985	0.294	0.296	0.956	2.012	0.290	0.302	0.961
IG	35%	200	2	0.993	0.509	0.515	0.952	2.048	0.669	0.673	0.952
			5	0.999	0.323	0.324	0.950	2.043	0.439	0.423	0.945
		400	2	0.995	0.364	0.363	0.950	2.007	0.445	0.465	0.960
			5	0.987	0.226	0.229	0.952	2.004	0.305	0.295	0.951
	60%	200	2	1.028	0.656	0.665	0.955	2.019	0.780	0.783	0.948
			5	0.993	0.417	0.421	0.955	2.009	0.490	0.492	0.946
		400	2	1.012	0.485	0.472	0.933	2.040	0.546	0.550	0.952
			5	0.985	0.307	0.297	0.946	2.008	0.350	0.346	0.944

Table 1 Simulation results for the proposed estimator with $\sigma^2 = 1$

			$\beta_0 = 1$				$\alpha_0 = 2$			
Frailty	σ^2	K	Mean	SD	SE	CP	Mean	SD	SE	CP
Gamma	0.5	2	0.996	0.505	0.523	0.949	2.002	0.541	0.530	0.950
		5	1.004	0.325	0.328	0.955	1.999	0.327	0.335	0.952
		7	0.999	0.279	0.275	0.944	1.996	0.282	0.284	0.942
	2	2	1.009	0.514	0.506	0.937	2.021	0.527	0.532	0.957
		5	1.021	0.312	0.317	0.955	2.005	0.343	0.338	0.947
		7	0.999	0.260	0.268	0.956	1.999	0.301	0.288	0.941
IG	0.5	2	1.013	0.524	0.519	0.944	2.053	0.622	0.607	0.952
		5	0.990	0.334	0.326	0.948	2.028	0.390	0.380	0.945
		7	0.991	0.275	0.276	0.938	2.016	0.327	0.321	0.938
	2	2	1.004	0.532	0.505	0.937	2.160	0.789	0.825	0.972
		5	0.994	0.317	0.318	0.949	2.056	0.537	0.500	0.933
		7	0.998	0.273	0.268	0.949	2.019	0.419	0.420	0.952

Table 2 Simulation results for the proposed estimator with $n = 200$ and a censoring rate of 35%

Frailty	σ^2	K	$\beta_0 = 1$ Mean	SD	SE	CP	$\alpha_0 = 2$ Mean	SD	SE	CP
Gamma	0.5	2	0.996	0.505	0.522	0.950	2.081	0.543	0.533	0.947
		5	1.004	0.325	0.327	0.956	2.079	0.329	0.336	0.952
		7	0.999	0.278	0.275	0.944	2.076	0.283	0.285	0.939
	2	2	1.009	0.514	0.507	0.937	1.881	0.517	0.522	0.943
		5	1.021	0.312	0.317	0.956	1.863	0.336	0.332	0.919
		7	0.999	0.260	0.268	0.956	1.858	0.295	0.283	0.906
IG	0.5	2	1.009	0.523	0.518	0.943	2.137	0.659	0.685	0.966
		5	0.990	0.333	0.326	0.947	2.117	0.432	0.429	0.951
		7	0.991	0.275	0.276	0.938	2.102	0.364	0.360	0.944
	2	2	1.003	0.533	0.506	0.935	1.940	0.655	0.654	0.945
		5	0.994	0.318	0.318	0.950	1.899	0.449	0.409	0.901
		7	0.998	0.273	0.268	0.948	1.872	0.353	0.347	0.924

Table 3 Simulation results for the proposed estimator with $n = 200$ and a censoring rate 35% when σ^2 is misspecified as 1

6 A Real Example

We now apply the methods to a data set from the Diabetic Retinopathy Study (Kupfer and ET-DRS Research Group [8]; Huster et al. [6]). The study was conducted to assess the effectiveness of laser photocoagulation in delaying visual loss among patients with diabetic retinopathy. Each patient had one eye randomized to laser treatment and the other eye receiving no treatment was a control. The failure time of interest is the time (in months) to visual loss as measured by visual acuity less than $\frac{5}{200}$. As in previous analysis of this study we confine our attention to a subset of 197 patients between risk group 6–12, and consider two covariates where Z_{1ik} indicates, by the value 1 or 0, whether or not the k-th eye ($k = 1$ for the left eye and $k = 2$ for the right eye) of the i-th patient was treated with laser photocoagulation and $Z_{2i1} = Z_{2i2}$ indicates, by the value 1 or 0, whether the i-th patient had adult-onset or juvenile-onset diabetics. Considering that different patients within different risk groups have very different courses of disease progression, we further evaluate the potential random effect of risk group on the time of visual loss by defining covariate $X_{i1} = X_{i2}$ indicating, by the value 0 or 1, whether the i-th patient is in risk group 6–9 or risk group 10–12.

The fitted model is

$$\lambda(t|Z_{ik}, X_{ik}; \xi_i) = \lambda_0(t) + \beta^{\mathrm{T}} Z_{ik} + \xi_i X_{ik},$$

where $Z_{ik} = (Z_{1ik}, Z_{2ik})^{\mathrm{T}}$. For simplicity, we take $\alpha = E\xi_i$ and assume that ξ_i is from the Gamma or IG distribution with σ^2 fixed at $0.5, 1$, or 2. The analysis results are summarized in Table 4. It can be seen that the laser treatment could significantly delay the visual loss. The adult-onset patients with diabetics tended to lost visual sooner. Furthermore, using our proposed random effect additive hazards model, we found that the patients associated with higher risk (risk group 10–12) were more likely to suffer from visual loss compared with these patients in risk group 6–9, although the sizes of the covariate effects of the risk groups seem substantially different for the Gamma and IG frailties. Additionally, these findings seem not to depend on the values of σ^2. As a comparison, we also fitted the data set with the Cox proportional hazards model (1.1), where ξ_i is Gamma distributed. We can draw similar conclusions even though the magnitude of the covariate effect estimates is different. This is mainly due to the different interpretation of the two model parameters: one is on the hazards difference and the other on the hazards ratio.

Model	Frailty	σ^2	Treatment		Age-onset		Risk groups	
			Est.	SE	Est.	SE	Est.	SE
Additive	Gamma	0.5	−0.1041	0.0197	0.0098	0.0237	0.2983	0.0443
		1	−0.1038	0.0197	0.0100	0.0237	0.3668	0.0529
		2	−0.1034	0.0197	0.0100	0.0237	0.4499	0.0648
	IG	0.5	−0.1056	0.0202	0.0088	0.0239	0.1007	0.0240
		1	−0.1056	0.0202	0.0088	0.0239	0.1027	0.0248
		2	−0.1056	0.0200	0.0089	0.0240	0.1072	0.0267
Cox	Gamma	-	−0.8050	0.1690	0.0697	0.1620	0.7416	0.1670

Table 4 Application to the Diabetic Retinopathy Study

7 Remarks

We propose a semiparametric regression method for the clustered failure time data when the intraclass dependence among the subjects from the same cluster is of interest, as well as the effects of covariates on the failure times. Estimating equations for the model parameters have been proposed. The resultant estimators were shown to be consistent and asymptotically normal.

The statistical challenge for the model (1.1) is mainly that, when marginalized over the random effect, the model (1.1) usually no longer possesses the proportional form (Hougaard [5]). The related inference procedures typically rely on the EM algorithm (Zeng and Lin [21]), which is usually complicated to be implemented in practice. Our proposed additive hazards model with random effect provides a useful alternative and features that, when marginalized over the random effect, it still maintains the structure of the additive hazards regression. The corresponding computational procedure is less demanding.

Note that the baseline hazard function is the same for all the subjects, which is more suitable for the real example analyzed in current paper. However, to incorporate the distinct baseline functions, the following model could be considered,

$$\lambda(t|Z_{ik}, X_{ik}; \xi_i) = \lambda_{0k}(t) + \beta^{\mathrm{T}} Z_{ik}(t) + \xi_i^{\mathrm{T}} X_{ik}(t),$$

where $\lambda_{0k}(\,\cdot\,)$ is the unspecified subject-specific baseline hazard function for the k-th subjects in the clusters. Our proposed inference procedures can be slightly modified for this model.

8 Proofs of the Asymptotics

The limit is taken as $n \to \infty$ unless otherwise indicated. We restate a useful lemma which is adapted from Spiekerman and Lin [17].

Lemma 8.1 *Assume that $f_n\,(n=1,2,\dots)$ is a sequence of random functions on $[0,\tau]$ that satisfies $\int_0^\tau |df_n(u)| = O_p(1)$ and $\sup_{t\in[0,\tau]} |f_n(t)| = o_p(1)$. Then for $k=1,2,\dots,K$,*

$$\sup_{t\in[0,\tau]} \left| n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t f_n(u) dM_{ik}(u) \right| \rightarrow_p 0.$$

Proof See Spiekerman and Lin [17, p. 1172]. □

Let $\mathbf{U}_k(\theta,t) = \sum_{i=1}^n \int_0^t \{Q_{ik}(u,\theta) - \bar{Q}(u,\theta)\} dM_{ik}(u)$. Then $\mathbf{U}(\theta,t) = \sum_{k=1}^K \mathbf{U}_k(\theta,t)$. From Lemma 8.1, we have that $n^{-\frac{1}{2}}\mathbf{U}_k(\theta_0,t)$ is asymptotically equivalent to $n^{-\frac{1}{2}}\mathcal{U}_k(\theta_0,t)$, where

$$\mathcal{U}_k(\theta_0,t) = \sum_{i=1}^{n} \int_0^t \{Q_{ik}(u,\theta_0) - \bar{q}(u,\theta_0)\} dM_{ik}(u).$$

Lemma 8.2 $-n^{-1}\frac{\partial \boldsymbol{U}(\widehat{\theta}_n,\tau)}{\partial\theta}$ *converges almost surely to A for any estimator $\widehat{\theta}_n \rightarrow_{\text{a.s.}} \theta_0$.*

Proof Note that for $k=1,2,\dots,K$,

$$\begin{aligned} n^{-1}\frac{\partial \mathbf{U}_k(\theta,\tau)}{\partial\theta} &= n^{-1}\sum_{i=1}^{n}\int_0^\tau Y_{ik}(u)\left\{Q_{ik}(u,\theta) - \bar{Q}(u,\theta)\right\} \begin{pmatrix} Z_{ik}(u)du \\ h_\alpha(\bar{X}_{ik}(u))X_{ik}(u)du \end{pmatrix}^{\mathrm{T}} \\ &\quad - n^{-1}\sum_{i=1}^{n}\int_0^\tau \left\{\dot{Q}_{ik}(u,\theta) - \dot{\bar{Q}}(u,\theta)\right\} dM_{ik}(u,\theta) \\ &\equiv a_k(\theta) - b_k(\theta), \end{aligned}$$

where $\dot{Q}_{ik}(t,\theta) = \frac{\partial Q_{ik}(t,\theta)}{\partial\theta}$ and $\dot{\bar{Q}}(t,\theta) = \frac{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(t)\dot{Q}_{ik}(t,\theta)}{\sum_{i=1}^n\sum_{k=1}^K Y_{ik}(t)}$.

Under conditions (C3) and (C5), select δ sufficiently small such that $n^{-1}\partial\mathbf{U}_k(\theta,\tau)/\partial\theta$ sufficiently closed to $n^{-1}\partial\mathbf{U}_k(\theta_0,\tau)/\partial\theta$ uniformly in n whenever $\|\theta-\theta_0\| < \delta$. Using the strong law of large numbers, we have that $a_k(\theta_0)$ converges almost surely to

$$E\left\{\int_0^\tau Y_{1k}(u)\{Q_{1k}(u,\theta_0) - \bar{q}(u,\theta_0)\}\begin{pmatrix} Z_{1k}(u)du \\ h_{\alpha_0}(\bar{X}_{1k}(u))X_{1k}(u)du \end{pmatrix}^{\mathrm{T}}\right\}.$$

It follows from Lemma 8.1 that $b_k(\theta_0)$ is asymptotically negligible. Thus, the desired result follows from a straightforward calculation. □

Proof of Theorem 3.1 Under Lemma 8.2 and the condition (C4), denote $d = (4\|A^{-1}\|)^{-1}$ and $d_n = [4\|\{n^{-1}\partial\mathbf{U}(\theta_0,\tau)/\partial\theta\}^{-1}\|]^{-1}$ whenever $n^{-1}\partial\mathbf{U}(\theta_0,\tau)/\partial\theta$ is nonsingular. Select δ sufficiently small such that $\|n^{-1}\partial\mathbf{U}(\theta,\tau)/\partial\theta - n^{-1}\partial\mathbf{U}(\theta_0,\tau)/\partial\theta\| < d$ whenever $\|\theta-\theta_0\| < \delta$ for all n. Since d_n almost surely converges to d by Lemma 8.2, we conclude that

$$\left\| n^{-1}\frac{\partial\mathbf{U}(\theta,\tau)}{\partial\theta} - n^{-1}\frac{\partial\mathbf{U}(\theta_0,\tau)}{\partial\theta}\right\| < 2d_n \tag{8.1}$$

for n large enough, where n does not depend on θ. In other words, one can find a commonly large n such that the inequality (8.1) holds for all θ.

Let $O_\delta = \{\theta\colon \|\theta-\theta_0\| < \delta\}$ and it follows from the inverse function theorem (see Foutz [4]) that $n^{-1}\mathbf{U}(\,\cdot,\tau)$ is a one-to-one mapping from O_δ onto $n^{-1}\mathbf{U}(O_\delta,\tau)$ and the image set $n^{-1}\mathbf{U}(O_\delta,\tau)$ contains an open neighborhood of $n^{-1}\mathbf{U}(\theta_0,\tau)$ with radius $d_n\delta$. Hence, when n is taken sufficiently large, the image set $n^{-1}\mathbf{U}(O_\delta,\tau)$ contains the open neighborhood $n^{-1}\mathbf{U}(\theta_0,\tau)$ with radius $\frac{d\delta}{2}$. On the other hand, the convergence of $n^{-1}\mathbf{U}(\theta_0,\tau)$ to zero can be derived obviously from straightforward extension of Lemma 8.1. Therefore, $\widehat{\theta}$ exists and is unique in O_δ and $\widehat{\theta}$ converges to θ_0 almost surely since δ can be taken arbitrarily small. Moreover, the arguments in Jacobsen [7] can be used to demonstrated the global uniqueness of $\widehat{\theta}$ for large n.

It follows from the Taylor expansion and Lemmas 8.1 and 8.2 that

$$\begin{aligned} n^{\frac{1}{2}}(\widehat{\theta}-\theta_0) &= \left\{-n^{-1}\frac{\partial \mathbf{U}(\theta^*,\tau)}{\partial\theta}\right\}^{-1} n^{-\frac{1}{2}}\mathbf{U}(\theta_0,\tau) \\ &= A^{-1}\sum_{k=1}^{K}\left\{n^{-\frac{1}{2}}\mathcal{U}_k(\theta_0,\tau)\right\} + o_p(1), \end{aligned}$$

where θ^* is on the line segment between $\widehat{\theta}$ and θ_0. Obviously, $n^{-\frac{1}{2}}\mathcal{U}_k(\theta_0,\tau)$ converges in distribution to a zero-mean normal with covariance $\Sigma_{kk}(\tau,\tau)$. This result, combined with the Slutsky theorem, concludes the proof of the asymptotic normality. □

Proof Theorem 3.2 Some manipulation entails that

$$\begin{aligned} \widehat{\Lambda}_n(t,\widehat{\theta}) - \Lambda_0(t) &= \{\widehat{\Lambda}_n(t,\theta_0) - \Lambda_0(t)\} + \{\widehat{\Lambda}_n(t,\widehat{\theta}) - \widehat{\Lambda}_n(t,\theta_0)\} \\ &= n^{-1}\int_0^t \frac{d\sum_{i=1}^n\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{S}_k^{(0)}(u,\theta_0)} + \left\{\frac{\partial\widehat{\Lambda}_n(t,\theta_0)}{\partial\theta}\right\}^{\mathrm{T}}(\widehat{\theta}-\theta_0) + o_p(|\widehat{\theta}-\theta_0|) \\ &= n^{-1}\int_0^t \frac{d\sum_{i=1}^n\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{S}_k^{(0)}(u,\theta_0)} - \int_0^t \mathbf{E}(v,\theta_0)^{\mathrm{T}}dv(\widehat{\theta}-\theta_0) + o_p(|\widehat{\theta}-\theta_0|). \end{aligned}$$

Using Lemma 8.1, we obtain

$$\sup_{t\in[0,\tau]}\left|\int_0^t \frac{d\sum_{i=1}^n\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{S}_k^{(0)}(u,\theta_0)} - \int_0^t \frac{d\sum_{i=1}^n\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{s}_k^{(0)}(u,\theta_0)}\right| = o_p(n^{\frac{1}{2}}).$$

This, coupled with the almost sure convergence of $\widehat{\theta}$ to θ_0, concludes that $\widehat{\Lambda}_n(t,\widehat{\theta})$ converges almost surely to $\Lambda_0(t)$, uniformly in $t\in[0,\tau]$. On the other hand, we could display the following asymptotic approximation

$$\begin{aligned} & n^{\frac{1}{2}}\{\widehat{\Lambda}_n(t,\widehat{\theta}) - \Lambda_0(t)\} \\ &= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t \frac{d\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{S}_k^{(0)}(u,\theta_0)} - \int_0^t \mathbf{E}(v,\theta_0)^{\mathrm{T}}dvn^{\frac{1}{2}}(\widehat{\theta}-\theta_0) + o_p(n^{\frac{1}{2}}|\widehat{\theta}-\theta_0|) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^n\int_0^t \frac{d\sum_{k=1}^K M_{ik}(u)}{\sum_{k=1}^K \mathbf{s}_k^{(0)}(u,\theta_0)} - \int_0^t \mathbf{e}(v,\theta_0)^{\mathrm{T}}dvA^{-1}\sum_{k=1}^K\left\{n^{-\frac{1}{2}}\mathcal{U}_k(\theta_0,\tau)\right\} + o_p(1) \\ &= n^{-\frac{1}{2}}\sum_{i=1}^n \Psi_i(t) + o_p(1), \end{aligned}$$

where $o_p(1)$ is uniformly in $t\in[0,\tau]$.

Obviously, the convergence of $n^{\frac{1}{2}}\{\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t)\}$ in finite-dimensional distribution follows from the central limit theorem. On the other hand, we can write $\Psi_i(t)$ as a sum of some monotone functions in t and monotone functions have pseudodimension 1 (see Pollard [16, p. 15]). Thus, the process $\{\Psi_i(t)\colon i=1,2,\ldots,n\}$ are manageable (see Pollard [16, p. 38]). It then follows from the functional central limit theorem (see Pollard [16, p. 53]) that $\{\Psi_i(t)\colon i=1,2,\ldots,n\}$ is tight, which concludes the proof of the weak convergence. □

Proof of Theorem 4.1 Using the functional central limit theorem, $n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}M_{ik}(t)$ converges weakly to a zero-mean Gaussian process with continuous sample paths. By the strong embedding theorem (see van der Vaart and Wellner [18]) and Lemma A.3 (see Bilias et al. [1]), we have that as $n\to\infty$,

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\{\mathbf{g}_n(s,x)-\tilde{\mathbf{g}}(s,x)\}dM_{ik}(s)\to 0,$$
$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\{\bar{Q}(s,x)-\bar{q}(s,x)\}dM_{ik}(s)\to 0,$$

in probability. Some calculations yield that $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ is asymptotically equivalent to $n^{-\frac{1}{2}}\tilde{\mathbf{G}}(t,x;\theta_0)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\Phi_i(t,x;\theta_0)$. For fixed t and x, $\{\Phi_i(t,x;\theta_0)\colon i=1,2,\dots,n\}$ are i.i.d. random vectors. Consequently, it follows from the central limits theorem that $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ converges in finite-dimensional distribution to a zero-mean Gaussian process.

Without loss of generality, we assume that the covariates are bounded in $[-1,1]$. Following the arguments in Spiekerman and Lin [17], we next show the tightness of $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ in $\mathcal{D}([0,\tau]\times[-1,1]^{p+q})$. Rewrite

$$n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})=n^{-\frac{1}{2}}\Phi^{(1)}(t,x;\theta_0)+n^{-\frac{1}{2}}\Phi^{(2)}(t,x;\theta_0)+n^{-\frac{1}{2}}\Phi^{(3)}(t,x;\theta_0)+o_p(1),$$

where

$$\Phi^{(1)}(t,x;\theta_0)=\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\mathbf{f}(W_{ik}(s))I(W_{ik}(s)\le x)dM_{ik}(s,\theta_0),$$
$$\Phi^{(2)}(t,x;\theta_0)=\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^t\tilde{\mathbf{g}}(s,x)dM_{ik}(s,\theta_0),$$
$$\Phi^{(3)}(t,x;\theta_0)=\tilde{\mathbf{h}}(t,x;\theta_0)n^{\frac{1}{2}}(\widehat{\theta}-\theta_0).$$

Note that $n^{-\frac{1}{2}}\Phi^{(1)}(t,x;\theta_0)$ is tight by using Example 2.11.16 in van der Vaart and Wellner [18]. It follows from the weak convergence of $n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{k=1}^{K}M_{ik}(t,\theta_0)$ that $n^{-\frac{1}{2}}\Phi^{(2)}(t,x;\theta_0)$ converges weakly to a zero-mean Gaussian random field. Thus, $n^{-\frac{1}{2}}\Phi^{(2)}(t,x;\theta_0)$ is tight by Theorem 10.2 (Pollard [16]). The tightness of $n^{-\frac{1}{2}}\Phi^{(3)}(t,x;\theta_0)$ follows the uniform boundedness of $\tilde{\mathbf{h}}(t,x;\theta_0)$ and the asymptotic normality of $n^{\frac{1}{2}}(\widehat{\theta}-\theta_0)$. Hence, we have proved the tightness of $n^{-\frac{1}{2}}\mathbf{G}_n(t,x;\widehat{\theta})$ and then its weak convergence property. □

Acknowledgements The authors thank the associate editor and the referees for their careful reading and many insightful suggestions, which strengthened the work immensely.

References

[1] Bilias, Y., Gu, M., Ying, Z.: Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.*, **25**, 662–682 (1997)

[2] Fine, J. P., Glidden, D. V., Lee, K. E.: A simple estimation for a shared frailty regression model. *J. Royal Statist. Soc., Ser. B*, **65**, 317–329 (2003)

[3] Fleming, T. R., Harrington, D. P.: Counting Processes and Survival Analysis, Wiley, New York, 1991

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

[5] Hougaard, P.: Analysis of Multivariate Survival Data, Springer, New York, 2000

[6] Huster, W. J., Brookmeyer, R., Self, S. C.: Modelling paired survival data with covariates. *Biometrics*, **45**, 145–156 (1989)

[7] Jacobsen, M.: Existence and unicity of MLEs discrete exponential family distributions. *Scand. J. Statist.*, **16**, 335–349 (1989)

[8] Kupfer, C., ETDRS Research Group: The diabetic retinopathy vitrectomy study (editorial). *Amer. J. Ophthalmology*, **81**, 687–690 (1976)

[9] Lin, D. Y., Wei, L. J., Yang, I., et al.: Semiparametric regression for the mean and rate functions of recurrent events. *J. Royal Statist. Soc.*, *Ser., B*, **62**, 711–730 (2000)

[10] Lin, D. Y., Ying, Z.: Semiparametric analysis of the additive risk model. *Biometrika*, **81**, 61–71 (1994)

[11] Martinussen, T., Pipper, C. B.: Estimation in the positive stable shared frailty Cox proportional hazards model. *Lifetime Data Analysis*, **11**, 99–115 (2005)

[12] McKeague, I. W., Sasieni, P. D.: A partly parametric additive risk model. *Biometrika*, **81**, 501–514 (1994)

[13] Murphy, S. A.: Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.*, **22**, 712–731 (1994)

[14] Murphy, S. A.: Asymptotic theory for the frailty model. *Ann. Statist.*, **23**, 182–198 (1995)

[15] Parner, E.: Asymptotic theory for the correlated gamma-frailty model. *Ann. Statist.*, **26**, 183–214 (1998)

[16] Pollard, D.: Empirical Processes: Theories and Applications, Institute of Mathematical Statistics, Hayward, 1990

[17] Spiekerman, C. F., Lin, D. Y.: Marginal regression models for multivariate failure time data. *J. Amer. Statist. Assoc.*, **93**, 1164–1175 (1998)

[18] van der Vaart, A. W., Wellner, J. A.: Weak Convergence and Empirical Processes, Springer, New York, 1996

[19] Yin, G.: Model checking for additive hazards model with multivariate survival data. *J. Multivariate Anal.*, **98**, 1018–1032 (2007)

[20] Yin, G., Cai, J.: Additive hazards model with multivariate failure time data. *Biometrika*, **91**, 801–818 (2004)

[21] Zeng, D., Lin, D. Y.: Maximum likelihood estimation in semiparametric models with censored data (with discussion). *J. Royal Statist. Soc.*, *Ser. B*, **69**, 507–564 (2007)