Acta Mathematica Sinica, English Series
Sep., 2011, Vol. 27, No. 9, pp. 1831–1842
Published online: August 15, 2011
DOI: 10.1007/s10114-011-8232-x
Http://www.ActaMath.com

Acta Mathematica Sinica, English Series

Semiparametric Additive Intensity Model with Frailty for Recurrent Events

Yan Yan LIU **Yuan Shan WU**

School of Mathematics and Statistics, Wuhan University, Wuhan 430072, *P. R. China*
E-mail: *liuyy@whu.edu.cn* *wuyuanshan*1102*@yahoo.cn*

Abstract The seminal Cox's proportional intensity model with multiplicative frailty is a popular approach to analyzing the frequently encountered recurrent event data in scientific studies. In the case of violating the proportional intensity assumption, the additive intensity model is a useful alternative. Both the additive and proportional intensity models provide two principal frameworks for studying the association between the risk factors and the disease recurrences. However, methodology development on the additive intensity model with frailty is lacking, although would be valuable. In this paper, we propose an additive intensity model with additive frailty to formulate the effects of possibly time-dependent covariates on recurrent events as well as to evaluate the intra-class dependence within recurrent events which is captured by the frailty variable. The asymptotic properties for both the regression parameters and the association parameters in frailty distribution are established. Furthermore, we also investigate the large-sample properties of the estimator for the cumulative baseline intensity function.

Keywords Additive intensity model, counting process, Donsker class, frailty, recurrent events

MR(2000) Subject Classification 62N02, 62N01, 62G20

1 Introduction

Since the fundamental paper of Andersen and Gill [1], the proportional intensity model has been the popular choice for the regression analysis of recurrent event data. Let $N^*(t)$ denote the number of events that the subject has experienced by time t, and let $Z(t)$ be a vector of possibly time-dependent covariates. The proportional intensity model specifies that the intensity function for $N^*(t)$ associated with $Z(t)$ takes the form

$$\lambda(t|Z(s), s \le t) = \lambda_0(t)\mathrm{e}^{\beta^T Z(t)}, \tag{1.1}$$

where $\lambda_0(\cdot)$ is an unspecified baseline intensity function and β is a vector of unknown regression parameters. In this model, the risks that are induced by the covariates are relative to the baseline $\lambda_0(\cdot)$. Elegant martingale method was adopted by Andersen and Gill [1] to drive the large-sample properties for β and $\Lambda_0(t) \equiv \int_0^t \lambda_0(u)du$.

In some case, however, it may be more appropriate to use the additive intensity model (Lin and Ying [2]; McKeague and Sasieni [3]). In this model the intensity function for $N^*(t)$ is

Received May 22, 2008, revised December 9, 2009, accepted April 2, 2010
Supported by National Natural Science Foundation of China (Grant No. 10771163)

related to a vector of covariates $Z(t)$ through

$$\lambda(t|Z(s), s \leq t) = \lambda_0(t) + \beta^T Z(t), \tag{1.2}$$

where $\lambda_0(\cdot)$ is again an unspecified baseline function. Thus, in this model the risks that are induced by the covariates are excess risks. Lin and Ying [2] derived a large-sample theory paralleling the martingale method developed by Andersen and Gill [1].

It should be noted that methodology development on both models (1.1) and (1.2) always implies that the occurrence of an event is independent of any earlier events that occurred to the same subject and usually ignores the intra-class correlation among the same subjects when estimating the covariate effects. When one is interested in the dependence of the recurrent event times within the same subject, a useful approach to accommodating it is to incorporate a frailty or random effect into the considered model. In particular, as an extension of the model (1.1), the conditional intensity model is related to a vector of covariates $Z(t)$ and a frailty ξ through

$$\lambda(t|Z(s), s \leq t; \xi) = \xi\lambda_0(t)\mathrm{e}^{\beta^T Z(t)}, \tag{1.3}$$

where ξ is an unobserved frailty representing the intra-class correlation among the recurrent events occurring on the same subject. The presence of frailty poses considerable challenges in statistical inference. This model has been extensively investigated by Nielsen et al. [4] and Oakes [5]. The rigorous asymptotic theory has been established for the special case of gamma frailty without covariates by Murphy [6, 7] and with covariates by Parner [8], respectively. The other choices of frailties such as positive stable distribution are also discussed by Hougaard [9]. Recently, Zeng and Lin [10] proposed a broad class of intensity models with random effects and studied the nonparametric maximum likelihood estimators for parameters of these models based on empirical processes theory.

Despite the progress achieved in the proportional intensity model with frailty, methodology developed on the additive intensity model with frailty is lacking although would be valuable. Pipper and Martinussen [11] proposed marginal additive hazards model with parametric shared frailty, which multiplicatively affects the baseline hazard function and regression function, for analyzing the clustered survival data.

An alternative way to construct frailty model with additive intensity function is through what we call additive frailty, which assumes that the frailty is added, instead of multiplied, to the intensity function. In this article, we propose the statistical inference for the additive intensity model (1.2) which accommodates the dependence of the recurrent event times within the same subject by introducing an unobserved additive frailty variable denoted by ξ. To be specific, the additive intensity model with additive frailty takes the following form:

$$\lambda(t|Z(s), X(s), s \leq t; \xi) = \lambda_0(t) + \beta^T Z(t) + \xi^T X(t), \tag{1.4}$$

where β is a set of unknown regression parameters, ξ is the frailty variable with density function $\phi(\xi; \gamma)$ indexed by parameter γ, $\lambda_0(\cdot)$ is a completely unspecified baseline function, Z and X are the possibly time-dependent p-vector and q-vector covariate processes, associated with fixed and random effects, respectively. Usually, the covariates X may contain 1 and part of Z. In addition, note that the proposed model (1.4) is different from the partly parametric version of Aalen's additive model studied by McKeague and Sasieni [3] in which the influence of some covariates varies nonparametrically over time and that of the remaining is constant across time.

In Section 2, we describe the inference procedures for θ and Λ_0, where $\theta = (\beta^T, \gamma^T)^T$. Section 3 studies the large-sample properties for the derived estimators with proofs relegated to Section 4. Some remarks are provided in Section 5.

2 Inference Procedures

In most applications, individual in the study cohort is subject to censoring, and let C be the censoring time. Define $N(t) = N^*(t \wedge C)$ and $Y(t) = I(C \geq t)$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. Assume that the censoring is conditionally on frailty variable ξ and the covariate processes Z and X is non-informative on θ (Andersen et al. [12]). Let τ denote the duration of the study. For a random sample of n subjects, we assume that $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot), X_i(\cdot), \xi_i\}$ $(i = 1, \ldots, n)$ are independent and identically distributed. We intend to utilize the observable data $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot), X_i(\cdot)\}$ $(i = 1, \ldots, n)$ to make inference about θ and $\Lambda_0(t)$.

Let the observed data filtration be

$$\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s), X_i(s) : i = 1, \ldots, n, 0 \leq s \leq t\},$$

and the full data filtration be

$$\mathcal{G}_t = \sigma\{N_i(s), Y_i(s), Z_i(s), X_i(s), \xi_i : i = 1, \ldots, n, 0 \leq s \leq t\}.$$

It follows from model (1.4) that the intensity function for counting process $N_i(t)$ associated with $\mathcal{G}_t$-filtration is $Y_i(t)\lambda(t|Z_i(s), X_i(s), s \leq t; \xi_i)$, where

$$\lambda(t|Z_i(s), X_i(s), s \leq t; \xi_i) = \lambda_0(t) + \beta^T Z_i(t) + \xi_i^T X_i(t),$$

and then by innovative theorem (Andersen et al. [12]), the intensity function for $N_i(t)$ associated with observed $\mathcal{F}_t$-filtration is $Y_i(t)\lambda(t|Z_i(s), X_i(s), s \leq t)$, where

$$\lambda(t|Z_i(s), X_i(s), s \leq t) = \lambda_0(t) + \beta^T Z_i(t) + E(\xi_i|\mathcal{F}_{t-})^T X_i(t).$$

To compute the conditional expectation $E(\xi_i|\mathcal{F}_{t-})$, we derive the observed likelihood based on the observed data $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot), X_i(\cdot)\}$ $(i = 1, \ldots, n)$ as follows:

$$\begin{aligned}\mathcal{L}(\theta, \lambda_0) = \prod_{i=1}^n \Bigg\{ \int_\xi \prod_{u \leq \tau} \{Y_i(u)(\lambda_0(u) + \beta^T Z_i(u) + \xi^T X_i(u))\}^{\Delta N_i(u)} \\ \times \exp\left\{ -\int_0^\tau Y_i(u)\lambda_0(u)du - \beta^T \int_0^\tau Y_i(u)Z_i(u)du - \xi^T \int_0^\tau Y_i(u)X_i(u)du \right\} \\ \times \phi(\xi; \gamma)d\xi \Bigg\}.\end{aligned}$$

Thus, combing Bayesian rule and some simple calculations, we have

$E(\xi_i|\mathcal{F}_{t-})$

$$
\begin{aligned}
&= \frac{\int_\xi \xi \prod_{u\le t-}\left\{Y_i(u)(\lambda_0(u)+\beta^T Z_i(u)+\xi^T X_i(u))\right\}^{\Delta N_i(u)} \exp\left\{-\xi^T\overline{X}_i(t)\right\}\phi(\xi;\gamma)d\xi}{\int_\xi \prod_{u\le t-}\left\{Y_i(u)(\lambda_0(u)+\beta^T Z_i(u)+\xi^T X_i(u))\right\}^{\Delta N_i(u)} \exp\left\{-\xi^T\overline{X}_i(t)\right\}\phi(\xi;\gamma)d\xi} \\
&= \frac{\int_\xi \xi \exp\{\int_0^{t-}\log\left\{\lambda_0(u)+\beta^T Z_i(u)+\xi^T X_i(u)\right\}dN_i(u)-\xi^T\overline{X}_i(t)\}\phi(\xi;\gamma)d\xi}{\int_\xi \exp\{\int_0^{t-}\log\left\{\lambda_0(u)+\beta^T Z_i(u)+\xi^T X_i(u)\right\}dN_i(u)-\xi^T\overline{X}_i(t)\}\phi(\xi;\gamma)d\xi}, \qquad (2.1)
\end{aligned}
$$

where $\Delta N_i(u)$ denotes the jump of N_i at u and $\overline{X}_i(t)=\int_0^{t-}Y_i(u)X_i(u)du$. To emphasize the dependence of $E(\xi_i|\mathcal{F}_{t-})$ on the unknown parameters θ and λ_0, we denote it by $H_i(t,\theta,\lambda_0)$. In practice, the integrations in this formula can be evaluated by numerical approximations (Evans and Swartz [13]).

Next we intend to drive the inference procedures for θ and λ_0 based on the intensity model

$$\lambda(t|Z_i(s),X_i(s),s\le t)=\lambda_0(t)+\beta^T Z_i(t)+H_i(t,\theta,\lambda_0)^T X_i(t). \qquad (2.2)$$

For notational simplicity, denote $\lambda(t|Z_i(s),X_i(s),s\le t)$ by $\lambda(t|Z_i,X_i)$, and let

$$M_i(t,\theta,\lambda_0)=N_i(t)-\int_0^t Y_i(u)\lambda(u|Z_i,X_i)du, \qquad (2.3)$$

which is a counting process martingale associated with the observed $\mathcal{F}_t$-filtration (Fleming and Harrington [14]).

Obviously, the main challenge to estimate the unknown parameters based on model (2.2) is the dependence of $H_i(t,\theta,\lambda_0)$ on the baseline intensity function λ_0. However, observe that any estimator for $\lambda_0(\cdot)$ depends on the jumps of N_i $(i=1,\dots,n)$ from (2.1). Hence, following the arguments taken by Pipper and Martinussen [11], we estimate $\Lambda_0(\cdot)$ initially as follows:

$$\widetilde{\Lambda}_0^{\rm I}(t)=\int_0^t \frac{\sum_{i=1}^n\{dN_i(u)-Y_i(u)\widetilde{\beta}_{\rm I}^T Z_i(u)du\}}{\sum_{i=1}^n Y_i(u)},$$

where $\widetilde{\beta}_{\rm I}$ is some initial consistent estimator for β_0. Note that $\widetilde{\Lambda}_0^{\rm I}(t)$ is a consistent estimator of $\Lambda_0(t)$ under conditions C1–C4 and C8 listed in the next section. Furthermore, the kernel-smoothed intensity function estimator from Klein and Moeschberger [15] is

$$\widetilde{\lambda}_0^{\rm KS}(t)=\int_0^\tau \frac{1}{b_n}K\left(\frac{t-u}{b_n}\right)d\widetilde{\Lambda}_0^{\rm I}(u),$$

where $K(\cdot)$ is the symmetric kernel function and b_n is the bandwidth such that $b_n\to 0$ and $nb_n^2\to\infty$ as $n\to\infty$. It can be shown that $\widetilde{\lambda}_0^{\rm KS}(\cdot)$ is a consistent estimator for $\lambda_0(\cdot)$ under some regular conditions described in Klein and Moeschberger [15]. Therefore, it is natural to derive the model (2.2) based estimator for cumulative baseline intensity function $\Lambda_0(t)$ with given θ as follows:

$$\widehat{\Lambda}_n(t,\theta)=\int_0^t \frac{\sum_{i=1}^n\{dN_i(u)-Y_i(u)(\beta^T Z_i(u)+H_i(u,\theta,\widetilde{\lambda}_0^{\rm KS})^T X_i(u))du\}}{\sum_{i=1}^n Y_i(u)}.$$

To derive the estimator for θ, suppose that $Q_i(t,\theta,\lambda_0)$ is a smooth $(p+q)$-vector-valued function of $Z_i(t)$, $X_i(t)$ and θ, and also dependent on λ_0. As elucidated by Lin and Ying [16] and Liu et al. [17], we specify the estimating function denoted by $\mathbf{U}(\tau,\theta)$ for parameter θ, where

$$\mathbf{U}(t,\theta)=\sum_{i=1}^n\int_0^t Q_i(u,\theta,\widetilde{\lambda}_0^{\rm KS})\{dN_i(u)-Y_i(u)(\beta^T Z_i(u)du$$

$$+ H_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u)du + d\widehat{\Lambda}_n(u,\theta))\}$$
$$= \sum_{i=1}^n \int_0^t \{Q_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})\}\{dN_i(u)$$
$$- Y_i(u)(\beta^T Z_i(u) + H_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u))du\},$$

with

$$\overline{Q}(t,\theta,\lambda_0) = \frac{\sum_{i=1}^n Y_i(t) Q_i(t,\theta,\lambda_0)}{\sum_{i=1}^n Y_i(t)}.$$

The solution to $\mathbf{U}(\tau,\theta) = 0$, denoted by $\widehat{\theta}$, is used to estimate θ_0, the true value of θ. Therefore, we obtain the estimator $\widehat{\Lambda}_n(t,\widehat{\theta})$ for $\Lambda_0(t)$. 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 to those in Lin and Ying [2], it can be shown that $\widehat{\Lambda}_n^*(t) - \widehat{\Lambda}_n(t,\widehat{\theta}) = o_p(n^{-\frac{1}{2}})$. On the other hand, we may choose

$$Q_i(t,\theta,\lambda_0) = [Z_i^T(t) + X_i^T(t) h_i^{(1)}(t,\theta,\lambda_0), X_i^T(t) h_i^{(2)}(t,\theta,\lambda_0)]^T,$$

for simplicity, where $h_i^{(1)}(t,\theta,\lambda_0) = \frac{\partial}{\partial\beta} H_i(t,\theta,\lambda_0)$ and $h_i^{(2)}(t,\theta,\lambda_0) = \frac{\partial}{\partial\gamma} H_i(t,\theta,\lambda_0)$.

3 Asymptotic Properties

In this section, we establish the large sample properties of the proposed estimators, beginning with the following regularity conditions for $i = 1,\ldots,n$, throughout our discussion.

C1 $\{N_i^*(\cdot), Y_i(\cdot), Z_i(\cdot), X_i(\cdot), \xi_i, Q_i(\cdot,\theta,\lambda_0)\}$ are independent and identically distributed, where θ is in a parameter space Θ, say. Furthermore, Θ is compact, containing true value of parameter θ_0 as its interior point.

C2 $P(C_i \ge \tau) > 0$.

C3 $N_i(\tau)$ is bounded by a constant. $\int_0^\tau \lambda_0(t)dt < \infty$ and $\lambda_0(t)$ is continuous on $[0,\tau]$.

C4 Both $Z_i(\cdot)$ and $X_i(\cdot)$ are of bounded total variations; moreover, $Q_i(\cdot,\theta,\lambda_0)$ has bounded total variation, uniformly in $\theta \in \Theta$, for every λ_0.

C5 The following defined matrix A is nonsingular:

$$A = E\left\{\int_0^\tau Y_1(u)\{Q_1(u,\theta_0,\lambda_0) - \overline{q}(u,\theta_0,\lambda_0)\}\begin{bmatrix} (Z_1(u) + h_1^{(1)}(u,\theta_0,\lambda_0)^T X_1(u))du \\ h_1^{(2)}(u,\theta_0,\lambda_0)^T X_1(u)du \end{bmatrix}^T\right\},$$

where $\overline{q}(t,\theta,\lambda_0) = \frac{E[Y_1(t)Q_1(t,\theta,\lambda_0)]}{E[Y_1(t)]}$.

C6 The class $\{\partial Q_i(t,\cdot,\lambda_0)/\partial\theta, H_i(t,\theta,\lambda_0), h_i(t,\theta,\lambda_0) : t \in [0,\tau], i = 1,\ldots,n\}$ are equicontinuous and bounded uniformly in parameter space Θ, where $h_i(t,\theta,\lambda_0) = \frac{\partial}{\partial\theta} H_i(t,\theta,\lambda_0)$.

C7 $H_i(t,\theta_1,\lambda_0^1) = H_i(t,\theta_2,\lambda_0^2)$ a.s. implies that $\theta_1 = \theta_2$ and $\lambda_0^1 = \lambda_0^2$. $h_i(t,\theta,\lambda_0)v_\gamma = 0$ a.s., then $v_\gamma = 0$.

C8 For $i = 1,\ldots,n$, ξ_i and $X_i(t)$ are orthogonal in the sense that $E\{\xi_i^T X_i(t)\} = 0$ for each t.

Conditions C1, C3, C4 and C6 simplify our derivations of the asymptotic results but are not practical limitations, and Condition C2 can be enforced by choosing τ to be not greater than the maximum observation time. In addition, Condition C5 is a technique assumption.

Note that $\overline{q}(t,\theta,\lambda_0)$ is well defined under Condition C2. Condition C7 is a necessary identifiability condition for model parameters. Actually, Condition C8 states that random effect ξ and covariate $X(t)$ are uncorrelated after centerization.

We describe the asymptotic behaviors of estimates of the regression parameters and the association parameters in the following theorem.

Theorem 3.1 *Under Conditions* C1 *to* C8, $\widehat{\theta}$ *is a consistent estimator for* θ_0, *while* $\sqrt{n}(\widehat{\theta}-\theta_0)$ *converges weakly to a zero-mean normal with covariance matrix* $A^{-1}\Sigma(\tau,\tau)(A^T)^{-1}$, *where*

$$\Sigma(s,t) = E\bigg[\int_0^s \{Q_1(u,\theta_0,\lambda_0) - \overline{q}(u,\theta_0,\lambda_0)\}dM_1(u)\int_0^t \{Q_1(v,\theta_0,\lambda_0) - \overline{q}(v,\theta_0,\lambda_0)\}^T dM_1(v)\bigg]$$

for s *and* t *in* $[0,\tau]$, *with* $dM_1(t) = dM_1(t,\theta_0,\lambda_0)$.

A consistent estimator of the covariance matrix is given by $\widehat{A}^{-1}\widehat{\Sigma}(\tau,\tau)(\widehat{A}^T)^{-1}$, where

$$\begin{aligned}\widehat{A} &= \frac{1}{n}\sum_{i=1}^n \int_0^\tau Y_i(u)\{Q_i(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})\}\\ &\quad\cdot\begin{bmatrix} (Z_i(u) + h_i^{(1)}(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u))du \\ h_i^{(2)}(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u)du \end{bmatrix}^T,\\ \widehat{\Sigma}(s,t) &= \frac{1}{n}\sum_{i=1}^n \int_0^s \{Q_i(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})\}d\widehat{M}_i(u)\\ &\quad\cdot\int_0^t \{Q_i(v,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(v,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})\}^T d\widehat{M}_i(v),\end{aligned}$$

with $d\widehat{M}_i(t) = dN_i(t) - Y_i(t)(d\widehat{\Lambda}_n(t,\widehat{\theta}) + \widehat{\beta}^T Z_i(t)dt + H_i(t,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(t)dt)$. The proof of Theorem 3.1 is deferred to Section 4.

Let $\mathbf{S}^{(0)}(t) = \frac{1}{n}\sum_{i=1}^n Y_i(t)$ and

$$\mathbf{S}^{(1)}(t,\theta,\lambda_0) = \frac{1}{n}\sum_{i=1}^n Y_i(t)\begin{bmatrix} Z_i(t) + h_i^{(1)}(t,\theta,\lambda_0)^T X_i(t) \\ h_i^{(2)}(t,\theta,\lambda_0)^T X_i(t)\end{bmatrix},$$

and write $\mathbf{E}(t,\theta,\lambda_0) = \mathbf{S}^{(1)}(t,\theta,\lambda_0)/\mathbf{S}^{(0)}(t)$. The limiting values of $\mathbf{S}^{(0)}(t), \mathbf{S}^{(1)}(t,\theta,\lambda_0)$, and $\mathbf{E}(t,\theta,\lambda_0)$ are given by $\mathbf{s}^{(0)}(t)$, $\mathbf{s}^{(1)}(t,\theta,\lambda_0)$, and $\mathbf{e}(t,\theta,\lambda_0)$, respectively. The essential asymptotic results for the cumulative baseline intensity function estimator are summarized by the following theorem.

Theorem 3.2 *Under Conditions* C1–C8, $\{\widehat{\Lambda}_n(t,\widehat{\theta}) - \Lambda_0(t)\}$ *converges in probability to* 0, *uniformly in* $t \in [0,\tau]$, *while* $\sqrt{n}(\widehat{\Lambda}_n(t,\widehat{\theta}) - \Lambda_0(t))$ *converges weakly to a zero-mean Gaussian process with covariance function* $\psi(s,t) = E[\Psi_1(s)\Psi_1(t)]$, *where*

$$\Psi_i(t) = \int_0^t \frac{1}{\boldsymbol{s}^{(0)}(u)}dM_i(u) - \int_0^t \boldsymbol{e}(v,\theta_0,\lambda_0)^T dv A^{-1}\int_0^\tau \{Q_i(u,\theta_0,\lambda_0) - \overline{q}(u,\theta_0,\lambda_0)\}dM_i(u).$$

Denote

$$\widehat{\Psi}(t) = \int_0^t \frac{1}{\mathbf{S}^{(0)}(u)}d\widehat{M}_i(u) - \int_0^t \mathbf{E}(v,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})^T dv \widehat{A}^{-1}\int_0^\tau \{Q_i(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(u,\widehat{\theta},\widetilde{\lambda}_0^{\mathrm{KS}})\}d\widehat{M}_i(u).$$

Then the covariance function $\psi(s,t)$ could be consistently estimated by $\frac{1}{n}\sum_{i=1}^n \widehat{\Psi}(s)\widehat{\Psi}(t)$. The detailed proofs are also relegated to Section 4.

4 Proofs

The proofs of Theorems 3.1 and 3.2 are investigated in this section. We first state two useful lemmas.

Lemma 4.1 $\frac{1}{\sqrt{n}}\boldsymbol{U}(\tau,\theta_0)$ *converges weakly to a zero-mean normal with covariance* $\Sigma(\tau,\tau)$.

Proof Note that $\widetilde{\lambda}_0^{\mathrm{KS}}$ is a consistent estimator for λ_0. Furthermore, this convergence can be strengthened to uniform convergence in $t \in [0,\tau]$ by the continuity of λ_0 under Condition C3. Hence, it follows from the dominated convergence theorem and Conditions C3, C4, and C6 that

$$\begin{aligned}\frac{1}{\sqrt{n}}\mathbf{U}(\tau,\theta) = & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \{Q_i(u,\theta,\lambda_0) - \overline{Q}(u,\theta,\lambda_0)\}\\ & \times \{dN_i(u) - Y_i(u)(\beta^T Z_i(u) + H_i(u,\theta,\lambda_0)^T X_i(u))du\}\\ & + o_p(1)\\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \{Q_i(u,\theta,\lambda_0) - \overline{q}(u,\theta,\lambda_0)\}\\ & \times \{dN_i(u) - Y_i(u)(\beta^T Z_i(u) + H_i(u,\theta,\lambda_0)^T X_i(u))du\}\\ & + o_p(1).\end{aligned}$$

Thus, simple calculation yields that $\frac{1}{\sqrt{n}}\mathbf{U}(\tau,\theta_0)$ is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau \{Q_i(u,\theta_0,\lambda_0) - \overline{q}(u,\theta_0,\lambda_0)\}\, dM_i(u),$$

which is essentially a sum of independent and identically distributed random vectors. Then the desirable result follows straightforwardly.

Lemma 4.2 $-\frac{1}{n}\frac{\partial \boldsymbol{U}(\tau,\theta_0)}{\partial\theta}$ *converges in probability to* A.

Proof Analogously to Lemma 4.1, some simple algebraic manipulation yields that

$$\begin{aligned}-\frac{1}{n}\partial\mathbf{U}(\tau,\theta)/\partial\theta = & \frac{1}{n}\sum_{i=1}^n\int_0^\tau Y_i(u)\{Q_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{Q}(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})\}\\ & \cdot \begin{bmatrix} (Z_i(u) + h_i^{(1)}(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u))du \\ h_i^{(2)}(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u)du \end{bmatrix}^T\\ & - \frac{1}{n}\sum_{i=1}^n\int_0^\tau \{\dot{Q}_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}}) - \overline{\dot{Q}}(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})\}\\ & \times \big\{dN_i(u) - Y_i(u)\big(\beta^T Z_i(u) + H_i(u,\theta,\widetilde{\lambda}_0^{\mathrm{KS}})^T X_i(u)\big)du\big\} + o_p(1)\\ = & \frac{1}{n}\sum_{i=1}^n\int_0^\tau Y_i(u)\,\{Q_i(u,\theta,\lambda_0) - \overline{q}(u,\theta,\lambda_0)\}\\ & \cdot \begin{bmatrix} (Z_i(u) + h_i^{(1)}(u,\theta,\lambda_0)^T X_i(u))du \\ h_i^{(2)}(u,\theta,\lambda_0)^T X_i(u)du \end{bmatrix}^T\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\left\{\dot{Q}_i(u,\theta,\lambda_0)-\overline{\dot{q}}(u,\theta,\lambda_0)\right\}\\
&\times\left\{dN_i(u)-Y_i(u)\left(\beta^T Z_i(u)+H_i(u,\theta,\lambda_0)^T X_i(u)\right)du\right\}+o_p(1)\\
\equiv\ & a(\theta)-b(\theta)+o_p(1),
\end{aligned}
$$

where $\dot{Q}_i(t,\theta,\lambda_0) = \frac{\partial \dot{Q}_i(t,\theta,\lambda_0)}{\partial\theta}$, $\overline{\dot{Q}}(t,\theta,\lambda_0) = \frac{\sum_{i=1}^n Y_i(t)\dot{Q}_i(t,\theta,\lambda_0)}{\sum_{i=1}^n Y_i(t)}$, and $\overline{\dot{q}}$ is the limit of $\overline{\dot{Q}}$ as $n\to\infty$.

Obviously, $a(\theta_0)$ converges almost surely to A by Strong Law of Large Numbers. $b(\theta_0)$ is equal to

$$
\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau}\{\dot{Q}_i(u,\theta_0,\lambda_0)-\overline{\dot{q}}(u,\theta_0,\lambda_0)\}dM_i(u),
$$

which is asymptotically negligible. Thus we have the proof done. Furthermore, we can conclude under Condition C6 that $-\frac{1}{n}\partial\mathbf{U}(\tau,\theta^\dagger)/\partial\theta$ converges in probability to A for any estimator $\theta^\dagger$ such that $\theta^\dagger\to_p\theta_0$.

Proof of Theorem 3.1 For a vector b, define $\|b\| = \sup_i |b_i|$, and for a matrix B, define $\|B\| = \sup_{i,j}|b_{ij}|$. In view of Lemma 4.2 and Condition C5, let $d = 1/(4\|A^{-1}\|)$ and $d_n = 1/(4\|(-\frac{1}{n}\partial\mathbf{U}(\tau,\theta_0)/\partial\theta)^{-1}\|)$ whenever $\frac{1}{n}\partial\mathbf{U}(\tau,\theta_0)/\partial\theta$ is nonsingular. Select δ sufficiently small such that $\|\frac{1}{n}\partial\mathbf{U}(\tau,\theta)/\partial\theta-\frac{1}{n}\partial\mathbf{U}(\tau,\theta_0)/\partial\theta)\| < d$ whenever $\|\theta-\theta_0\|<\delta$, for all n. Since d_n converges in probability to d by Lemma 4.2, we can conclude that $\|\frac{1}{n}\partial\mathbf{U}(\tau,\theta)/\partial\theta-\frac{1}{n}\partial\mathbf{U}(\tau,\theta_0)/\partial\theta)\| < 2d_n$ for large n, where n is not dependent on θ, i.e., one can find a commonly large n for all θ under Condition C6.

Write $O_\delta = \{\theta : \|\theta-\theta_0\|<\delta\}$. It follows from the inverse function theorem (Foutz [18]) that $\frac{1}{n}\mathbf{U}(\tau,\cdot)$ is a one-to-one mapping from O_δ onto $\frac{1}{n}\mathbf{U}(\tau,O_\delta)$ and the image set $\frac{1}{n}\mathbf{U}(\tau,O_\delta)$ contains the open neighborhood $\frac{1}{n}\mathbf{U}(\tau,\theta_0)$ with radius $d_n\delta$. Hence, when n is taken sufficiently large, image set $\frac{1}{n}\mathbf{U}(\tau,O_\delta)$ contains the open neighborhood $\frac{1}{n}\mathbf{U}(\tau,\theta_0)$ with radius $d\delta/2$. On the other hand, the convergence of $\frac{1}{n}\mathbf{U}(\tau,\theta_0)$ to zero can be derived obviously from Lemma 4.1. 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.

With respect to asymptotic normality, it follows from the Taylor expansion, Lemmas 4.1 and 4.2 that

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}-\theta_0) &= \left(-\frac{1}{n}\frac{\partial\mathbf{U}(\tau,\theta^*)}{\partial\theta}\right)^{-1}\frac{1}{\sqrt{n}}\mathbf{U}(\tau,\theta_0)\\
&= A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{0}^{\tau}\{Q_i(u,\theta_0,\lambda_0)-\overline{q}(u,\theta_0,\lambda_0)\}dM_i(u)+o_p(1),
\end{aligned}
$$

where θ^* is on the line segment between $\widehat{\theta}$ and θ_0. Consequently, $\sqrt{n}(\widehat{\theta}-\theta_0)$ behaves asymptotically as a sum of independent and identically distributed random vectors, which converges to a zero-mean normal with covariance $A^{-1}\Sigma(\tau,\tau)(A^T)^{-1}$ following from Lemma 4.1.

Furthermore, through repeated applications of the strong law of large numbers and the convergence of $\widehat{\theta}$ to θ_0, the estimator $\widehat{A}^{-1}\widehat{\Sigma}(\tau,\tau)(\widehat{A}^T)^{-1}$ can be shown to converge in probability to $A^{-1}\Sigma(\tau,\tau)(A^T)^{-1}$.

To prove Theorem 3.2, we need to present two lemmas.

Lemma 4.3 *Let* $\mathbf{f}_n$ *be a sequence of bounded vector-valued functions and* g_n *be a sequence of bounded functions such that*

(i) $\sup_{0\le t\le\tau}\|\mathbf{f}_n(t)-\mathbf{f}(t)\|\to 0$ *for some bounded function* $\mathbf{f}$*,*

(ii) $\mathbf{f}_n$ *are elementwisely monotone on* $[0,\tau]$ *and*

(iii) $\sup_{0\le t\le\tau}|g_n(t)-g(t)|\to 0$ *for some continuous function* g *on* $[0,\tau]$*.*

Then

$$\sup_{0\le t\le\tau}\left\|\int_0^t \mathbf{f}_n(u)dg_n(u)-\int_0^t \mathbf{f}(u)dg(u)\right\|\to 0,$$
$$\sup_{0\le t\le\tau}\left\|\int_0^t g_n(u)d\mathbf{f}_n(u)-\int_0^t g(u)d\mathbf{f}(u)\right\|\to 0.$$

This lemma is a simple extension of Lemma 1 of Lin et al. [19] and its proof follows from that of Lin et al. [19] straightforwardly by replacing $|\cdot|$ by $\|\cdot\|$. We now extend Lemma 4.3 to the situation of sequences of stochastic processes.

Lemma 4.4 *Let* $\mathbf{F}_n(t)$ *be a sequence of bounded vector-valued processes and* $G_n(t)$ *be a sequence of processes such that*

(i) $\sup_{0\le t\le\tau}\|\mathbf{F}_n(t)-\mathbf{F}(t)\|\to_p 0$ *for some bounded process* $\mathbf{F}(t)$*,*

(ii) $\mathbf{F}_n(t)$ *are elementwisely monotone on* $[0,\tau]$ *and*

(iii) $G_n(t)$ *converges weakly to a mean-zero process with continuous sample paths.*

Then

$$\sup_{0\le t\le\tau}\left\|\int_0^t \{\mathbf{F}_n(u)-\mathbf{F}(u)\}dG_n(u)\right\|\underset{p}{\to} 0,$$

$$\sup_{0\le t\le\tau}\left\|\int_0^t G_n(u)d\{\mathbf{F}_n(u)-\mathbf{F}(u)\}\right\|\underset{p}{\to} 0.$$

Proof It suffices to prove the first conclusion because the second one follows from it through the integration by parts formula. Let $G(t)$ be the weak convergence limit of $G_n(t)$. It follows from the strong embedding theorem that there exists a new probability space wherein $(\widetilde{\mathbf{F}}_n,\widetilde{G}_n)$ converges almost surely to $(\widetilde{\mathbf{F}},\widetilde{G})$, where $(\widetilde{\mathbf{F}}_n,\widetilde{G}_n)$ has the same distribution as $(\mathbf{F}_n,G_n)$ and $(\widetilde{\mathbf{F}},\widetilde{G})$ has the same distribution as $(\mathbf{F},G)$. Since $\mathbf{F}_n(t)$ is monotone on $[0,\tau]$ and $G(t)$ is continuous on $[0,\tau]$, by applying Lemma 4.3, we have, as $n\to\infty$, in this new probability space

$$\sup_{0\le t\le\tau}\left\|\int_0^t \widetilde{\mathbf{F}}_n(u)d\widetilde{G}_n(u)-\int_0^t \widetilde{\mathbf{F}}(u)d\widetilde{G}(u)\right\|\underset{\text{a.s.}}{\to} 0,$$

while in the original probability space we have

$$\sup_{0\le t\le\tau}\left\|\int_0^t \mathbf{F}_n(u)dG_n(u)-\int_0^t \mathbf{F}(u)dG(u)\right\|\underset{p}{\to} 0.$$

Specifically, by taking the place of $\mathbf{F}_n(u)$ with $\mathbf{F}(u)$ in the above display, one has

$$\sup_{0\le t\le\tau}\left\|\int_0^t \mathbf{F}(u)d\{G_n(u)-G(u)\}\right\|\underset{p}{\to} 0.$$

Thus

$$\sup_{0\le t\le\tau}\left\|\int_0^t\{\mathbf{F}_n(u)-\mathbf{F}(u)\}dG_n(u)\right\| \le \sup_{0\le t\le\tau}\left\|\int_0^t \mathbf{F}_n(u)dG_n(u)-\int_0^t\mathbf{F}(u)dG(u)\right\| + \sup_{0\le t\le\tau}\left\|\int_0^t \mathbf{F}(u)d\{G_n(u)-G(u)\}\right\| \underset{p}{\to} 0.$$

Note that Lemma 4.4 still holds if condition (ii) is extended to more general case in which $\mathbf{F}_n(t)$ could be represented as a sum of finite number of monotone functions on $[0,\tau]$.

After preparations of these two lemmas, we turn back to the proof of Theorem 3.2.

Proof of Theorem 3.2 Taylor expansion and some simple manipulation entail that

$$\begin{aligned}&\sqrt{n}(\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t))\\ &= -\int_0^t \mathbf{E}(u,\theta_0,\lambda_0)^T du\sqrt{n}(\widehat{\theta}-\theta_0)+\int_0^t\frac{\sqrt{n}\sum_{i=1}^n dM_i(u)}{\sum_{i=1}^n Y_i(u)}+o_p(\sqrt{n}\|\widehat{\theta}-\theta_0\|)\\ &= -\int_0^t \mathbf{E}(u,\theta_0,\lambda_0)^T duA^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\{Q_i(u,\theta_0,\lambda_0)-\overline{q}(u,\theta_0,\lambda_0)\}dM_i(u)\\ &\quad+\int_0^t\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(u)}{\mathbf{S}^{(0)}(u)}+o_p(1).\end{aligned}$$

Note that as a function of t,

$$Y(t)\begin{bmatrix} Z(t)+h^{(1)}(t,\theta_0,\lambda_0)^T X(t)\\ h^{(2)}(t,\theta_0,\lambda_0)^T X(t)\end{bmatrix}$$

is of bounded variation over $t\in[0,\tau]$ under Conditions C4 and C6 so it can be represented as a sum of two monotone functions of $t\in[0,\tau]$. Hence, condition (ii) of Lemma 4.4 is satisfied. It follows from Theorem 2.7.5 of van der Vaart and Wellner [20] that the class of functions

$$\left\{Y(t)\begin{bmatrix} Z(t)+h^{(1)}(t,\theta_0,\lambda_0)^T X(t)\\ h^{(2)}(t,\theta_0,\lambda_0)^T X(t)\end{bmatrix}:t\in[0,\tau]\right\}$$

is Donsker, so is the class$\{Y(t):t\in[0,\tau]\}$. Hence, we have

$$\sup_{0\le t\le\tau}\|\mathbf{S}^{(1)}(t,\theta_0,\lambda_0)-\mathbf{s}^{(1)}(t,\theta_0,\lambda_0)\|\underset{p}{\to}0$$

and

$$\sup_{0\le t\le\tau}\|\mathbf{S}^{(0)}(t)-\mathbf{s}^{(0)}(t)\|\underset{p}{\to}0.$$

Then under Condition C2,

$$\sup_{0\le t\le\tau}\|\mathbf{E}(t,\theta_0,\lambda_0)-\mathbf{e}(t,\theta_0,\lambda_0)\|\underset{p}{\to}0.$$

Thus, condition (i) of Lemma 4.4 is verified. Since condition (iii) of Lemma 4.4 is naturally satisfied, from Lemma 4.4, we have

$$\sup_{0\le t\le\tau}\left\|\int_0^t(\mathbf{E}(u,\theta_0,\lambda_0)-\mathbf{e}(u,\theta_0,\lambda_0))du\right\|\underset{p}{\to}0.$$

On the other hand, to show

$$\sup_{0\le t\le\tau}\left\|\int_0^t\left(\frac{1}{\mathbf{S}^{(0)}(u)}-\frac{1}{\mathbf{s}^{(0)}(u)}\right)d\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n M_i(u)\right)\right\|\underset{p}{\rightarrow}0,$$

it suffices to show that the limiting process of $\frac{1}{\sqrt{n}}\sum_{i=1}^n M_i(t)$ has continuous sample paths by using Lemma 4.4 and condition C2. Likewise, it follows from (2.3) that the class $\{M(t):t\in[0,\tau]\}$ is Donsker. Observe that $E(M(t))=0$. Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^n M_i(t)$ converges weakly to a limiting Gaussian process, denoted by $\mathcal{W}(t)$. It can be shown that $E\{\mathcal{W}(t)-\mathcal{W}(s)\}^4\le\widetilde{K}|t-s|^2$ for some constant $\widetilde{K}>0$. It then follows from the Kolmogorov–Centsov theorem (Karatzas and Shereve [21]) that $\mathcal{W}$ has continuous sample paths.

Hence, we can conclude that

$$\sqrt{n}\sup_{0\le t\le\tau}|\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t)|=O_p(1) \tag{4.1}$$

and

$$\begin{aligned}
&\sqrt{n}(\widehat{\Lambda}_n(t,\widehat{\theta})-\Lambda_0(t))\\
&=-\int_0^t\mathbf{e}(u,\theta_0,\lambda_0)^T du A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\{Q_i(u,\theta_0,\lambda_0)-\overline{q}(u,\theta_0,\lambda_0)\}dM_i(u)\\
&\quad+\int_0^t\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(u)}{\mathbf{s}^{(0)}(u)}+o_p(1)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\Psi_i(t)+o_p(1).
\end{aligned} \tag{4.2}$$

Thus, the uniform consistency follows from (4.1) and the weak convergence follows from (4.2), and the class $\{\Psi(t):t\in[0,\tau]\}$ is Donsker. Hence, the proof of Theorem 3.2 is done.

Furthermore, for fixed s and t, the convergence of $\frac{1}{n}\sum_{i=1}^n\widehat{\Psi}(s)\widehat{\Psi}(t)$ to $\psi(s,t)$ can be derived straightforwardly from the uniform strong law of large numbers and the consistency of $\widehat{\theta}$ to θ_0 and $\widehat{\Lambda}_n(t,\widehat{\theta})$ to $\Lambda_0(t)$.

5 Concluding Remarks

We propose a semiparametric additive intensity model with additive frailty for analyzing the recurrent event data. The corresponding estimators for both regression parameters and association parameters are derived together with their large sample properties. Meanwhile, we also discuss the asymptotic properties of the cumulative baseline intensity function estimator.

Since different frailty variables would induce different correlation structures, the more reasonable choice should be based on the nature of data set and the aim of investigator. For univariate frailty, the positive stable frailty should be preferred in situation where the correlated recurrent event data show a decreasing association with time, while the gamma frailty shows an increasing association with time.

The proposed additive intensity model with additive frailty can be extended straightforwardly to accommodate the clustered survival data when one is interested in comparing the lifetimes of individuals within the same cluster, estimating the correlation between lifetimes, as well as evaluating the influence of covariates on lifetime. A separate research is now under way.

Acknowledgements The authors are grateful for the valuable comments and suggestions from the Associate Editor and the referees which drastically improved the appearance of this article.

References

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

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

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

[4] Nielsen, G. G., Gill, R. D., Andersen, P. K., et al.: A counting process approach to maximum likelihood estimation in frailty models. *Scand. J. Statist.*, **19**, 25–43 (1992)

[5] Oakes, D.: Frailty models for multiple event times. In: Survival Analysis: State of the Art (J. P. Klein and P. K. Goel eds.), Kluwer Academic, Amsterdam, 1992, 371–379

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

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

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

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

[10] Zeng, D., Lin, D. Y.: Semiparametric transformation models with random effects for recurrent events. *J. Amer. Statist. Assoc.*, **102**, 167–180 (2007)

[11] Pipper, C. B., Martinussen, T.: An estimating equation for parametric shared frailty models with marginal additive hazards. *J. Roy. Statist. Soc. Ser. B*, **66**, 207–220 (2004)

[12] Andersen, P. K., Borgan, O., Gill, R. D., et al.: Statistical Models Based on Counting Processes, Springer-Verlag, New York, 1993

[13] Evans, M., Swartz, T.: Approximating Integrals via Monte Carlo and Deterministic Methods, Oxford University Press, Oxford, 2000

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

[15] Klein, J. P., Moeschberger, M. L.: Survival Analysis: Techniques for Censored and Truncated Data, 2nd Edition, Springer-Verlag, New York, 2003

[16] Lin, D. Y., Ying, Z.: Semiparametric analysis of general additve-multiplicative hazard models for counting processes. *Ann. Statist.*, **23**, 1712–1734 (1995)

[17] Liu, Y., Wu, Y., Cai, J., et al.: Additive-multiplicative rates model for recurrent events. *Lifetime Data Anal.*, **16**, 353–373 (2010)

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

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

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

[21] Karatzas, I., Shereve, S. E.: Brownian Motion and Stochastic Calculus, Springer-Verlag, New York, 1988