Additive transformation models for clustered failure time data

Donglin Zeng and
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, dzeng@bios.unc.edu

Jianwen Cai
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA, cai@bios.unc.edu

Abstract

We propose a class of additive transformation risk models for clustered failure time data. Our models are motivated by the usual additive risk model for independent failure times incorporating a frailty with mean one and constant variability which is a natural generalization of the additive risk model from univariate failure time to multivariate failure time. An estimating equation approach based on the marginal hazards function is proposed. Under the assumption that cluster sizes are completely random, we show the resulting estimators of the regression coefficients are consistent and asymptotically normal. We also provide goodness-of-fit test statistics for choosing the transformation. Simulation studies and real data analysis are conducted to examine the finite-sample performance of our estimators.

Keywords

Additive model; Estimating equation; Goodness-of-fit; Robustness; Transformation

1 Introduction

Clustered failure time data are common in biomedical studies. Such data usually arise when study subjects come from the same cluster, for example, failure time can be time to disease occurrence for the patients in the same clinic or time to blindness for the two eyes in the same patient. One of the commonly used statistical models to analyze clustered failure time data is the gamma-frailty model, which accommodates the intra-cluster dependence by incorporating an unobserved random effect, the so-called frailty, into the Cox (1972) proportional hazards model. The consistency and asymptotic distribution of the maximum likelihood estimator for the gamma frailty model have been rigorously studied by Murphy (1994, 1995) for the case of no covariates and by Parner (1998) for the case with covariates.

The multiplicative hazards model focuses on estimating hazard ratios and its multiplicative structure may not model the real data well. In some cases, an additive effect could be more reasonable. Such an effect can be modeled using the so-called additive hazards model (Aalen 1989; Huffer and McKeague 1991; Lin and Ying 1994; and McKeague and Sasieni 1994; among others), which assumes an additive structure of a baseline hazards function and a covariate effect. Particularly, for univariate failure time, the additive models assume the following expression

$$d\Lambda(t|\mathbf{X})=d\Lambda(t)+\mathbf{X}(t)^T\beta dt, \tag{1}$$

where $\Lambda(t|\mathbf{X})$ denotes the cumulative hazard function for the given, possibly time-dependent, covariates $\mathbf{X}$ and $\Lambda(t)$ is the baseline hazard function. Through the additive structure, the additive model interprets $\boldsymbol{\beta}$ as risk difference and it has been eloquently advocated and successfully utilized for right-censored independent survival data in many papers, e.g., Andersen et al. (1993, pp. 563–566), Lin and Ying (1994), McKeague and Sasieni (1994), and Shen and Cheng (1999). Gandy and Jensen (2005a, b) developed a goodness-of-fit test for checking additive risk model. Additionally, the multivariate version of the above additive model (1) has been used to model clustered failure time data in Yin and Cai (2004):

$$d\Lambda_{ij}(t|\mathbf{X})=d\Lambda(t)+\mathbf{X}_{ij}(t)^T\beta dt,$$

where $\Lambda_{ij}(t|\mathbf{X})$ denotes the cumulative hazards function for subject j in cluster i and $\mathbf{X}_{ij}$ is the associated covariate. Yin (2007) further developed a test for checking the additive structure using clustered data. Recently, Yin et al. (2008) considered an additive hazards model with varying coefficients, where they allowed the effect of $\mathbf{X}_{ij}(t)$ to depend on other covariates.

All the previous additive models except Yin et al. (2008) assume that $\mathbf{X}$ affects the hazard rate linearly. However, in practical applications, this linear assumption may not always be appropriate. To relax the linear assumption, alternative links other than the linear link can be considered.

Especially, in this paper, we propose the following general class of additive transformation models:

$$\Lambda_{ij}(t|\mathbf{X}_{ij})=\Lambda(t)+Q\left(\int_0^t \mathbf{X}_{ij}(s)^T\beta ds\right), \tag{2}$$

where $Q(\cdot)$ is some known and monotone transformation function. Clearly, when $Q(x) = x$, we obtain the previous additive model. However, Q can be other positive link function, for example, $\exp\{x\}$ and $\log(1 + x)$ etc. For transformation $Q(x)$ which satisfies $E[\exp\{-\xi x\}] = \exp\{-Q(x)\}$ for some positive frailty ξ, model (2) can actually be motivated from a frailty model for clustered failure time data: when failure times are clustered, in parallel to generalizing the proportional hazards model to gamma-frailty model, we can generalize the univariate additive hazards model to incorporate some cluster-specific frailty. One natural generalization of the linear additive model in (1) is the following model

$$d\Lambda_{ij}(t|\mathbf{X}_{ij},\xi_i)=d\Lambda(t)+\xi_i\mathbf{X}_{ij}(t)^T\beta dt,$$

where $\Lambda_{ij}(t|\mathbf{X}, \xi_i)$ denotes the cumulative hazards function for subject j in cluster i given ξ_i and we introduce cluster-specific random effect ξ_i to account for within-cluster correlation in cluster i. Especially, when ξ_i has mean 1 and the cluster size is completely random, the above model yields that the effects of $\mathbf{X}(t)$ for all clusters have the same average $\boldsymbol{\beta}$ but allow random variation from cluster to cluster. From this general model, by the definition of $Q(x)$, we obtain that the marginal survival function for subject j in cluster i is

$$P(T_{ij} > t | \mathbf{X}_{ij}) = \exp\{-\Lambda(t)\} E_\xi \left[\exp\left\{ -\xi \int_0^t \mathbf{X}_{ij}(s)^T \beta ds \right\} \right]$$

$$= \exp\left\{ -\Lambda(t) - Q\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds \right) \right\}.$$

Hence, the marginal hazard function for T_{ij} is given by model (2). Particulary, if ξ follows a gamma-distribution with shape and scale parameters (θ_1, θ_2), then

$$Q(x) = \theta_1 \log(1 + \theta_2 x).$$

When $\theta_1 = \theta_2^{-1}$, $Q(x) = \theta_2^{-1} \log(1 + \theta_2 x)$ and we define $Q(x) = x$ at the special case of $\theta_2 = 0$, which is the limit as θ_2 goes to zero. However, for transformation $Q(x)$ which is not given by equation $E[\exp\{-\xi x\}] = \exp\{-Q(x)\}$, model (2) cannot be introduced by any frailty model any more.

The general model (2) gives a class of transformed additive model for marginal hazards function for clustered failure times and provides a very flexible way of modeling hazards models, in parallel to transformation models for multiplicative models. Our goal of this paper is to provide inference results for this general class of models. Especially, in Sect. 2, we give a simple inference procedure for model (2) and provide asymptotic properties for the proposed estimators. In Sect. 3, we discuss in detail some issues on how to select transformation function Q. Sect. 4 provides some numerical results from simulation studies and our approach is applied to data from Diabetic Retinopathy Study (Huster et al. 1989) in Sect. 5.

2 Model and inference

Assume that there are n independent clusters and for cluster i, there are n_i subjects. We denote T_{ij} and $\mathbf{X}_{ij}$ as the failure time and the covariates associated with subject j in cluster i, respectively. We let $\Lambda_{ij}(t | \mathbf{X}_{ij})$ be the cumulative hazard function for T_{ij} given covariates $\mathbf{X}_{ij}$. Our general additive risk model is given by Eq. 2.

Our goal is to estimate parameters specified in Eq. 2. In the presence of right-censoring, we only observe data

$$(Y_{ij} = T_{ij} \wedge C_{ij}, \Delta_{ij} = I(T_{ij} \leq C_{ij}), \mathbf{X}_{ij}), j = 1, \ldots, n_i,$$

where C_{ij} is the censoring time and assumed to be independent of T_{ij} given $\mathbf{X}_{ij}$ and the cluster sizes are assumed to be completely random. According to Eq. 2, it is clear

$$E[dN_{ij}(t) | \mathbf{X}_{ij}, Y_{ij} \geq t] = R_{ij}(t) \left\{ d\Lambda(t) + \mathbf{X}_{ij}(t)^T \beta Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds \right) dt \right\},$$

where $N_{ij}(t) = I(Y_{ij} \leq t)\Delta_{ij}$ is the observed event process, $R_{ij}(t) = I(Y_{ij} \geq t)$ is the at-risk process, and $Q'(\cdot)$ denotes the derivative of $Q(\cdot)$. Hence, using the fact that

$$dN_{ij}(t) - \left[R_{ij}(t)\left\{d\Lambda(t) + \mathbf{X}_{ij}(t)^T \beta Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds\right)dt\right\}\right]$$

has conditional mean zero and following the work of Lin and Ying (1994), we can estimate β and $\Lambda(\cdot)$ by solving the following equations

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \int I(t \le s) R_{ij}(t)$$

$$\left[dN_{ij}(t) - R_{ij}(t)\left\{d\Lambda(t) + \mathbf{X}_{ij}(t)^T \beta Q'\left(\int_0^t \mathbf{X}_{ij}(u)^T \beta du\right)dt\right\}\right] = 0,$$

(3)

for all $s \ge 0$, and

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \int \mathbf{X}_{ij}(t) R_{ij}(t)$$

$$\left[dN_{ij}(t) - R_{ij}(t)\left\{d\Lambda(t) + \mathbf{X}_{ij}(t)^T \beta Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds\right)dt\right\}\right] = 0.$$

(4)

Note that each term of the above equations has mean zero when the parameters take the true values. Particularly, we let $\Lambda(t)$ be a step function only jumping at the observed event time. Then from Eq. 2, we obtain that for given β,

$$d\Lambda(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n_i} R_{ij}(t)\left[dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T \beta Q'(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds)dt\right]}{\sum_{i=1}^n \sum_{j=1}^{n_i} R_{ij}(t)}.$$

(5)

After substituting back into Eq. 3 and some re-organization, we obtain that the estimator for β solves equation

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \int R_{ij}(t)\left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^n \sum_{l=1}^{n_k} R_{kl}(t)\mathbf{X}_{kl}(t)}{\sum_{k=1}^n \sum_{l=1}^{n_k} R_{kl}(t)}\right]$$

$$\times \left\{dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T \beta Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds\right)dt\right\} = 0.$$

The Newton–Raphson iterations can be used to solve the above equation. We denote the estimator for β as $\hat{\beta}$ and denote the estimator for $\Lambda(\cdot)$ given in Eq. 3 as $\hat{\Lambda}(\cdot)$.

The following results provide the asymptotic properties of the proposed estimators. To this end, we define

$$
\begin{aligned}
\Sigma = \quad & \lim_{n \to \infty} n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right]^T \\
& \times \left\{ \mathbf{X}_{ij}(t) \, Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds \right) + \mathbf{X}_{ij}(t)^T \beta_0 Q'' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds \right) \right. \\
& \left. \int_0^t \mathbf{X}_{ij}(s) ds \right\} \, dt
\end{aligned}
$$

and assume the limit exists and Σ is non-singular. Let $\boldsymbol{\beta}_0$ and $\Lambda_0(\cdot)$ denote the true values of $\boldsymbol{\beta}$ and $\Lambda(\cdot)$ respectively and cluster sizes n_k are random and identically distributed for each cluster. We obtain the following theorem.

Theorem 1

Assume $Q(\cdot)$ is a twice-continuously differentiable function. Moreover, $\mathbf{X}_{ij}(t)$ has a bounded total variation with probability one and $P(T_{ij} > \tau | \mathbf{X}_{ij}) > \delta$ and $P(C_{ij} > \tau | \mathbf{X}_{ij}) > \delta$ hold for some positive constant δ. Then $\sqrt{n}(\widehat{\beta} - \beta_0, \widehat{\Lambda}(t) - \Lambda_0(t))$ converges in distribution to a zero-mean Gaussian process in $R^d \times l^{\infty}[0, \tau]$.

The outline of the proof is given in the appendix. From the proof of Theorem 1, we further obtain that the influence function for $\widehat{\boldsymbol{\beta}}$ is given by

$$
\begin{aligned}
-\Sigma^{-1} & \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right. \\
& \left. \times \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}(t)^T \beta_0 Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds \right) dt \right\} \right].
\end{aligned}
$$

Therefore, a consistent estimator for the asymptotic covariance of $\widehat{\boldsymbol{\beta}}$ is

$$
\begin{aligned}
\widehat{\Sigma}^{-1} \sum_{i=1}^{n} & \left\{ \sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right] \right. \\
& \left. \times \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}(t)^T \widehat{\beta} Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) dt \right\} \right\}^{\otimes 2} \widehat{\Sigma}^{-1},
\end{aligned}
$$

where $a^{\otimes 2} = a a^T$ and $\hat{\Sigma}$ is a consistent estimator for Σ given by

$$
\begin{aligned}
n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} & \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right]^T \\
& \times \left\{ \mathbf{X}_{ij}(t) \, Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) + \mathbf{X}_{ij}(t)^T \widehat{\beta} Q'' \left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) \int_0^t \mathbf{X}_{ij}(s) ds \right\} \, dt.
\end{aligned}
$$

Additionally, the asymptotic covariance function for $\hat{\Lambda}(t)$ and $\hat{\Lambda}(s)$ can be consistently estimated by the empirical covariance of random variables $S(\cdot, t)$ and $S(\cdot, s)$, where "$\cdot$" denotes the observed data and

$$
S(\cdot, t) = \left\{ \int_0^t \frac{\sum_{j=1}^{n_i} R_{ij}(u) \left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T \widehat{\beta} Q'(\int_0^u \mathbf{X}_{ij}(s)^T \widehat{\beta} ds) du \right]}{\sum_{i=1}^n [\sum_{j=1}^{n_i} R_{ij}(u)]} \right\}
$$
$$
- \left\{ \int_0^t \frac{\sum_{i=1}^n \left[\sum_{j=1}^{n_i} R_{ij}(u) \left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T \widehat{\beta} Q'(\int_0^u \mathbf{X}_{ij}(s)^T \widehat{\beta} ds) du \right] \right] \sum_{j=1}^{n_i} R_{ij}(u)}{\sum_{i=1}^n [\sum_{j=1}^{n_i} R_{ij}(u)]^2} \right\}
$$
$$
- \left\{ \int_0^t \frac{\sum_{i=1}^n [\sum_{j=1}^{n_i} R_{ij}(u)\mathbf{X}_{ij}(u)^T \widehat{\beta} \nabla_\beta Q(\int_0^u \mathbf{X}_{ij}(s)^T \widehat{\beta} ds) du]}{\sum_{i=1}^n [\sum_{j=1}^{n_i} R_{ij}(u)]} \right\} S_\beta(\cdot),
$$

and S_β is the influence function for $\hat{\beta}$ but with β_0 replaced by $\hat{\beta}$ in its expression.

3 Selection of link function

3.1 Test of goodness-of-fit

One critical issue in model (2) is the choice of the transformation function $Q(\cdot)$. However, in practice, since one does not know the underlying distribution for the frailty, one question is whether the choice of some Q is sufficient to fit the data well. To address this question, we propose the following statistic for testing the goodness of fit with model (2):

$$
G_n(x) = \sum_{i=1}^n \sum_{j=1}^{n_i} \int I(\mathbf{X}_{ij}(t)^T \widehat{\beta} \le x) R_{ij}(t)
$$
$$
\times \left[dN_{ij}(t) - R_{ij}(t) \left\{ d\widehat{\Lambda}(t) + \mathbf{X}_{ij}(t)^T \widehat{\beta} Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) \right\} dt \right].
$$

In other words, our test statistic is the cumulative summation of the residuals given by

$$
dN_{ij}(t) - R_{ij}(t) \left\{ d\widehat{\Lambda}(t) + \mathbf{X}_{ij}(t)^T \widehat{\beta} Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) \right\} dt,
$$

where the cumulation is taken over the values of $\mathbf{X}_{ij}(t)^T \hat{\beta}$. The similar statistics have been given before, e.g., Yin (2007).

Using Eq. 3, $G_n(x)$ can also be written as

$$
\sum_{i=1}^n \sum_{j=1}^{n_i} \int R_{ij}(t) \left[I(\mathbf{X}_{ij}(t)^T \widehat{\beta} \le x) - \frac{\sum_{k=1}^n \sum_{l=1}^{n_k} R_{kl}(t) I(\mathbf{X}_{kl}(t)^T \widehat{\beta} \le x)}{\sum_{k=1}^n \sum_{l=1}^{n_k} R_{kl}(t)} \right]
$$
$$
\times \left\{ dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T \widehat{\beta} Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \widehat{\beta} ds \right) dt \right\}.
$$

The proposed test statistic $G_n(x)$ can be treated as examining how the mean event process relates to linear predictor $\mathbf{X}_{ij}(t)^T\boldsymbol{\beta}$. Thus, if the transformation function Q is misspecified, we would expect $G_n(x)$ departs from zero. Particulary, the following theorem holds.

Theorem 2—*Under model (2), $n^{-1/2}G_n(x)$ converges in distribution to a zero-mean Gaussian process in $l^\infty(-\infty,\infty)$.*

Using Theorem 2, we can simulate the limiting Gaussian process to test the goodness of fit. Especially, the resampling approach as given in Su and Wei (1991) can be used to simulate this process: we simulate n i.i.d observations, say $Z_1,\ldots,Z_n$, from standard normal distribution. We then generate a new $\boldsymbol{\beta}$ as

$$
\begin{aligned}
\beta^{new}= \; & \widehat{\beta} - \widehat{\Sigma}^{-1} n^{-1} \sum_{i=1}^{n} \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \right. \\
& \left. \times \left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^{n}\sum_{l=1}^{n_k} R_{kl}(t)\mathbf{X}_{kl}(t)}{\sum_{k=1}^{n}\sum_{l=1}^{n_k} R_{kl}(t)} \right] d\widehat{M}_{ij}(t) \right] Z_i,
\end{aligned}
$$

where $\widehat{\Sigma}$ is a consistent estimate for Σ given in the previous section, and

$$
d\widehat{M}_{ij}(t) = dN_{ij}(t) - R_{ij}(t)d\widehat{\Lambda}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T\widehat{\beta} Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T\widehat{\beta}ds\right) dt.
$$

With $\boldsymbol{\beta}^{new}$, we calculate the process

$$
\begin{aligned}
G_n^{new}(x)= \; & \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n_i} \int R_{ij}(t) \right. \\
& \times \left[I(\mathbf{X}_{ij}(t)^T\widehat{\beta} \le x) - \frac{\sum_{k=1}^{n}\sum_{l=1}^{n_k} R_{kl}(t) I(\mathbf{X}_{kl}(t)^T\widehat{\beta} \le x)}{\sum_{k=1}^{n}\sum_{l=1}^{n_k} R_{kl}(t)} \right] \\
& \times \left\{ dN_{ij}(t) - R_{ij}(t)d\widehat{\Lambda}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T\beta^{new} Q' \right. \\
& \left. \left. \times \left(\int_0^t \mathbf{X}_{ij}(s)^T\beta^{new}ds\right) dt \right\} \right\} Z_i.
\end{aligned}
$$

We repeatedly generate $G_n^{new}(x)$ for many times and the obtained simulated process should approximate the limiting process of $n^{-1/2}G_n(x)$.

The rationale behind this resampling approach is below. Following the exact expansion as in proving Theorem 2, we can show that conditional on the observed data, $n^{-1/2}G_n^{new}(x)$ is equivalent to

$$n^{-1/2} \sum_{i=1}^{n} \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \right.$$

$$\left[I(\mathbf{X}_{ij}(t)^T \beta_0 \leq x) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I(\mathbf{X}_{kl}(t)^T \beta_0 \leq x)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] dM_{ij}(t) \right] Z_i$$

$$-E \left\{ \sum_{j=1}^{n_i} \int R_{ij}(t) \left[I(\mathbf{X}_{ij}(t)^T \beta_0 \leq x) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I(\mathbf{X}_{kl}(t)^T \beta_0 \leq x)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.$$

$$\left. \times R_{ij}(t) \nabla_\beta (\mathbf{X}_{ij}(t)^T \beta_0 Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds \right)) dt \right\} \sqrt{n}(\widehat{\beta}^{new} - \beta_0) + o_p(1).$$

Therefore, conditioning on the observed data, $n^{-1/2} G_n^{new}(x)$ is approximately a Gaussian process with the same covariance as the asymptotic covariance of $n^{-1/2} G_n(x)$.

Using the resampling approach, we can obtain the approximate distribution of many useful testing statistics, including $\sup_x |G_n(x)|$, $\int |G_n(x)| dx$, and etc.

3.2 Robustness for null test

One interesting question is whether testing for the null hypothesis: $\boldsymbol{\beta} = 0$ is valid even if Q is misspecified. Particularly, we look into this issue under the following conditions:

(C.1) $\mathbf{X}_{ij}$ is time-independent and has mean zero;

(C.2) C_{ij} is independent of $\mathbf{X}_{ij}$ and has the same distribution for $j = 1, \ldots, n_i$;

(C.3) The misspecified transformation function Q is continuously differentiable and strictly increasing with $Q(0) = 0$.

Among these conditions, Condition (C.1) holds for any time-independent covariates after proper shifting. Condition (C.2) assumes the common distribution for the censoring time and independence between C_{ij} and X_{ij}. This condition is used mostly for the purpose of simplification; however, it should be carefully justified in practice. For example, when $C_{ij} = C_{ij'}$ for $j, j' = 1, \ldots, n_i$ are the administrative censoring time, the condition holds when X_{ij} are patient treatment allocations. However, the condition is not satisfied when patients in different treatment groups quit the study at systematically different times. Condition (C.3) is satisfied by transformations $Q(x) = \theta_1 \log(1 + \theta_2 x)$. These conditions are sufficient but not necessary conditions.

Under the assumptions in Theorem 1 and from its proof, we know $\hat{\boldsymbol{\beta}}$ with misspecified transformation function Q is consistent with some parameter $\boldsymbol{\beta}^*$ which satisfies

$$E \left[\sum_{j=1}^{n_i} \int R_{ij}(t)(\mathbf{X}_{ij} - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]}) \right.$$
$$\left. \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}^T \beta^* Q'(t \mathbf{X}_{ij}^T \beta^*) dt \right\} \right] = 0.$$

Since

$$E[dN_{ij}(t) | \mathbf{X}_{ij}, Y_{ij} \geq t] = R_{ij}(t) \left\{ d\Lambda_0(t) + \mathbf{X}_{ij}^T \beta_0 Q'_0(t \mathbf{X}_{ij}^T \beta_0) dt \right\},$$

we obtain

$$E\left[\sum_{j=1}^{n_i}\int I\left(Y_{ij}\geq t\right)\left(\mathbf{X}_{ij}-\frac{E\left[\sum_{l=1}^{n_k}R_{kl}\left(t\right)\mathbf{X}_{kl}\right]}{E\left[\sum_{l=1}^{n_k}R_{kl}\left(t\right)\right]}\right)\right.$$
$$\left.\left\{\mathbf{X}_{ij}^T\beta_0 Q_0'\left(t\mathbf{X}_{ij}^T\beta_0\right)-\mathbf{X}_{ij}^T\beta^* Q'\left(t\mathbf{X}_{ij}^T\beta^*\right)\right\}\,dt\right]=0. \tag{6}$$

This is the starting point for us to examine the robustness when Q is a misspecified transformation function.

When $\beta_0 = 0$, we note T_{ij} is independent of $\mathbf{X}_{ij}$ and has the same survival function, so from (C.1) and (C.2),

$$\frac{E\left[\sum_{l=1}^{n_k}R_{kl}\left(t\right)\mathbf{X}_{kl}\right]}{E\left[\sum_{l=1}^{n_k}R_{kl}\left(t\right)\right]}=0.$$

Thus, Eq. 6 gives

$$E\left[\sum_{j=1}^{n_i}\mathbf{X}_{ij}\,Q(Y_{ij}\mathbf{X}_{ij}^T\beta^*)\right]=0.$$

As the result, we have

$$E\left[\sum_{j=1}^{n_i}\mathbf{X}_{ij}Q'(Y_{ij}\mathbf{X}_{ij}^T\tilde{\beta})Y_{ij}\mathbf{X}_{ij}^T\right]\beta^*=0,$$

where $\tilde{\beta}$ is between 0 and β^*. Since $Q(\cdot)$ is strictly increasing from (C.3), the above equation holds if and only if $\beta^* = 0$. That is, we have

Theorem 3—*Under conditions (C.1)–(C.3)*, $\beta^* = 0$.

Theorem 3 indicates that under the null hypothesis, inference for β_0 using misspecified transformation function is still valid. In other words, one need not worry about using incorrect transformation when performing the hypothesis testing for $\beta = 0$.

4 Simulation studies

We have conducted simulations to examine the small-sample performance of the proposed method. In the first simulation study, we generated clustered survival data using the following true model

$$d\Lambda_{ij}\left(t|\mathbf{X}_{ij},\xi_i\right)=d\Lambda(t)+\xi_i\mathbf{X}_{ij}\left(t\right)^T\beta dt,$$

where each cluster contained two subjects and cluster-specific random effect ξ was generated from some gamma distribution with mean one and variance θ. Clearly, such data generation implied transformation additive model (2), where the different choices of θ gave

different transformations $Q(x)$. Two covariates were generated for each subject: for the first covariate, one subject in the cluster took value 0 and the other one had value 1; the second covariate was randomly generated from the uniform distribution in [0, 1]. We let $\Lambda(t) = t^2/2$ and generated censoring time from the uniform distribution in [0, 3], which yielded censoring rate around 30%. We set covariates effects to be 0.2 and 0.8, respectively.

We applied our method to estimate covariate effects based on the working model (2). Additionally, the sandwiched variance estimates was used for inference. Table 1 summarizes the results from 1000 repetitions with sample size 200 and 400. Column "Bias" denotes the average bias of the point estimates; column "SE" denotes the empirical standard error of the point estimates; column "ESE" is the average of the estimated standard errors; column "CP" refers to the coverage probability of 95% confidence interval based on the normal approximation. Table 1 shows that for each of the three transformation models, the proposed method works reasonably well with approximately unbiased estimates and correct inference.

In the second simulation study, we used the same set-up as before. However, we focused on the small-sample behavior for goodness-of-fit statistics in model (2). To generate the stochastic process in the goodness-of-fit statistics, we let x go through all the grid points in the range of $\mathbf{X}^T\boldsymbol{\beta}$. To obtain the distribution of such process, we adopted the resampling approach described in Sect. 3.1. Additionally, we focused on the supreme statistics $\sup_x |G_n(x)|$ and used 500 resamplings to simulate its distribution. Table 2 reports the probability that the observed statistics are larger than the 95% percentile of the simulated statistics from 1000 repetitions. It shows that the tail probability using our approach is very close to the theoretical value 5%, especially for $n = 400$. In other words, when the transformation model is specified correctly, using our goodness-of-fit statistics should provide a valid inference.

In the third simulation study, we aimed to confirm the robustness results as given in Sect. 3.2. In this simulation, we considered only one uniformly distributed covariate which had the same value for the two subjects in the same cluster. The true model corresponded to $Q(x) = 2\log(1 + x/2)$ and assumed that the covariate had no effect. However, we considered estimating $\boldsymbol{\beta}$ using the working model (2) with each of three transformations $Q(x) = x$, $Q(x) = 2\log(1 + x/2)$, and $Q(x) = \log(1 + x)$. Thus, we misspecified the transformation in the first and third cases. Table 3 gives the results from 1000 repetitions with column "α" being the type I error when testing the null hypothesis. Not surprisingly, we conclude that no matter what transformation is used, the inference for the null effects is always correct.

5 Application

We apply the proposed methods to the well-known Diabetic Retinopathy Study (Kupfer and ETDRS research group 1976; Huster et al. 1989), which was conducted to assess the effectiveness of laser photocoagulation in delaying visual loss among patients with diabetic retinopathy. One eye of each patient was randomly selected to receive the laser treatment while the other eye was used as a control. The failure time of interest is the time (in months) to visual loss as measured by visual acuity less than 5/200. Following previous authors, we confine our attention to a subset of 197 high-risk patients, and consider three covariates: $\mathbf{X}_{1ij}$ indicates, by the values 1 versus 0, whether or not the jth eye ($j = 1$ for the left eye and $j = 2$ for the right eye) of the ith patient was treated with laser photocoagulation, $\mathbf{X}_{2i1} \equiv \mathbf{X}_{2i2}$ indicates, by the values 1 versus 0, whether the ith patient had adult-onset or juvenile-onset diabetics, and $\mathbf{X}_{3i1} = \mathbf{X}_{3i2}$ is the age of the subject. Furthermore, we include the interaction term between treatment and diabetic type in the regression.

We fit model (2) with these covariates and consider a class of transformations $Q(x) = \theta^{-1}\log(1 + \theta x)$ by varying θ from 0 to 1 with $Q(x) = x$ when $\theta = 0$. To select the best fit, we calculate the goodness-of-fit statistic for each θ and the final θ is chosen as the one

minimizing the L_2-norm of such statistic. The best θ is determined to be 0.79. For comparison, Table 4 gives the estimated results from $\theta = 0$, $\theta = 0.79$, and $\theta = 1$. Table 4 shows that for different transformations, the point estimates are very close to each other but there is slightly increasing variation with increasing θ. Based on the selected model, both the treatment indicator and the interaction term are significant, whereas the diabetic type is not. The findings of both the direction and the significance of the covariate's effect agree with the results using different models (Cai et al. 2002;Zeng et al. 2005). Figure 1 gives the estimated cumulative hazards function for $\Lambda(t)$ for the untreated eye in a patient with juvenile diabetics and age 16 and it appears pretty linear.

In Fig. 2, we also plot the process for testing goodness-of-fit along with a few random paths from the limiting process, where the thick curve is the observed process. The plot shows that the selected model fits data pretty well. The p-value using the supreme statistics is 0.42. For comparison, we also fit the proportional hazards model with gamma frailty. Since the marginal hazard model is given by $\theta^{-1} \log\{1 + \Lambda(t)e^{\mathbf{X}^T\beta}\theta\}$ under the gamma frailty model where θ is the variance of the gamma frailty, we then construct the same goodness-of-fit statistic

$$
\begin{aligned}
G_n^M(x) = & \sum_{i=1}^{n}\sum_{j=1}^{n_i} \int I(\mathbf{X}_{ij}(t)^T\widehat{\beta} \le x)R_{ij}(t) \\
& \times \left[dN_{ij}(t) - R_{ij}(t)d\left\{\theta^{-1}\log\left(1+\widehat{\Lambda}(t)e^{\mathbf{X}_{ij}(t)^T\widehat{\beta}}\theta\right)\right\}\right].
\end{aligned}
$$

It turns out that the observed process for $G_n^M(x)$ has an increasing drift (Fig. 3) indicating not a good fit. Additionally, testing the equivalence of this model with our previously selected model using the supreme value of $G_n^M(x)$ yields the p-value close to zero. Thus, we conclude that the multiplicative model is not suitable for this data set. The additive model with $\theta = 0.79$ is a good choice.

6 Remarks

We have proposed a class of additive transformation models for multivariate failure time. We have particularly discussed the issue of selecting the transformation by proposing a valid goodness-of-fit test. Even though we considered the multivariate failure time, it is noted that all the procedures and approaches apply to the univariate failure time. However, the latter lacks the motivation for transformation model as given in the introduction. For the additive risk model, one drawback is that it may yield the decreasing cumulative hazards function when predicting $\Lambda(t|X)$. In this case, one usually modifies the predicted value, say $\hat{\Lambda}(t|X)$, as $\max_{s \le t} \hat{\Lambda}(s|X)$. It is easy to show that such modification still yields a consistent estimator.

When transformation $Q(x)$ is specifically from some gamma-frailty, we can work on the full likelihood function instead of marginal transformation hazards model as given in the paper. The observed likelihood function has form

$$
\prod_{i=1}^{n} \int \prod_{j=1}^{n_i} \left[\left(\lambda(t)+\xi e^{\mathbf{X}_{ij}(t)^T\beta}\right)^{\Delta_{ij}} e^{-\Lambda(t)-\xi\int_0^t \mathbf{X}_{ij}(s)^t\beta ds}\right] f(\xi)d\xi,
$$

where $f(\xi)$ is the gamma-frailty density. The inference for maximizing the likelihood function becomes much harder. It would be interesting to see the efficiency gain over our estimators by using this approach. Furthermore, one may even allow frailty variability to vary from cluster to cluster.

The transformation models can also be extended to other cases, for example, multiple types of events (Cai and Schaubel 2004) by allowing different baseline hazards function and different effects for different events. Additionally, when neither additive model nor multiplicative model is a sufficiently good approximation to model data structure, some combination of additive and multiplicative effects as given by Lin and Ying (1995) and Zeng et al. (2005), along with transformations, may be useful.

Acknowledgments

Appendix: outline of the proofs

Proof of Theorem 1 Let

$$
l_n(\beta) = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right]
$$
$$
\times \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}(t)^T \beta Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds \right) dt \right\}.
$$

First, we note

$$
\frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \to_{a.s.} \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]}.
$$

Thus, in any neighborhood of $\boldsymbol{\beta}_0$, $l_n(\boldsymbol{\beta})$ converges almost surely and uniformly in $\boldsymbol{\beta}$ to

$$
l(\beta) = E \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.
$$
$$
\times \left. \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}(t)^T \beta Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta ds \right) dt \right\} \right].
$$

Moreover, $\nabla_{\boldsymbol{\beta}} l_n(\boldsymbol{\beta})$ uniformly converges to $\nabla_{\boldsymbol{\beta}} l(\boldsymbol{\beta})$. Since

$$
l(\beta_0) = E \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.
$$
$$
\times \left. \left\{ dN_{ij}(t) - R_{ij}(t) \mathbf{X}_{ij}(t)^T \beta_0 Q' \left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds \right) dt \right\} \right]
$$
$$
= E \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) \mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] d\Lambda(t) \right] = 0
$$

and $-\nabla_{\boldsymbol{\beta}} l(\boldsymbol{\beta}_0) = \Sigma$ is non-singular by assumption, from the inverse function theorem, there exists some $\hat{\boldsymbol{\beta}}$ such that $l_n(\hat{\boldsymbol{\beta}}) = 0$. Therefore, we have shown $\hat{\boldsymbol{\beta}} \to_p \boldsymbol{\beta}_0$.

By the Taylor expansion, it is easy to show

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \Sigma^{-1}\mathbf{G}_n \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t)\mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.$$
$$\left. \times \left\{ dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T\beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T\beta_0 ds \right) dt \right\} \right] + o_p(1),$$

where $\mathbf{G}_n$ denotes the empirical process from n i.i.d observations. Furthermore, from Eq. 3 and the fact that

$$\Lambda_0(t) = \int_0^t \frac{E[\sum_{j=1}^{n_i} R_{ij}(u)[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T\beta_0 Q'(\int_0^u \mathbf{X}_{ij}(s)^T\beta_0 ds)du]]}{E[\sum_{j=1}^{n_i} R_{ij}(u)]},$$

we obtain

$$\sqrt{n}(\widehat{\Lambda}(t) - \Lambda_0(t))$$
$$= \mathbf{G}_n \left\{ \int_0^t \frac{\sum_{j=1}^{n_i} R_{ij}(u)\left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T\widehat{\beta}Q'(\int_0^u \mathbf{X}_{ij}(s)^T\widehat{\beta}ds)du \right]}{E[\sum_{j=1}^{n_i} R_{ij}(u)]} \right\}$$
$$- \mathbf{G}_n \left\{ \int_0^t \frac{E\left[\sum_{j=1}^{n_i} R_{ij}(u)\left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T\widehat{\beta}Q'(\int_0^u \mathbf{X}_{ij}(s)^T\widehat{\beta}ds)du \right]\right] \sum_{j=1}^{n_i} R_{ij}(u)}{E[\sum_{j=1}^{n_i} R_{ij}(u)]^2} \right\}$$
$$- E \left\{ \int_0^t \frac{E[\sum_{j=1}^{n_i} R_{ij}(u)\mathbf{X}_{ij}(u)^T\tilde{\beta}\nabla_\beta(\int_0^u \mathbf{X}_{ij}(s)^T\tilde{\beta}ds)du]}{E[\sum_{j=1}^{n_i} R_{ij}(u)]} \right\} \sqrt{n}\left(\widehat{\beta} - \beta_0\right),$$

where $\tilde{\beta}$ is between $\widehat{\beta}$ and β_0. Since $\mathbf{X}_{ij}(t)$, $R_{ij}(t)$ and $N_{ij}(t)$ are stochastic process with bounded variation, they belong to some Donsker class indexed by t. The same holds for $\mathbf{X}_{ij}(t)^T\beta$, $\int_0^t \mathbf{X}_{ij}(s)^T\beta ds$ and $\int_0^t \mathbf{X}_{ij}(s)\mathbf{X}_{ij}(s)^T\beta ds$ as stochastic process indexed by both t and β. As the result, we have

$$\sqrt{n}(\widehat{\Lambda}(t) - \Lambda_0(t))$$
$$= \mathbf{G}_n \left\{ \int_0^t \frac{\sum_{j=1}^{n_i} R_{ij}(u)\left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T\beta_0 Q'(\int_0^u \mathbf{X}_{ij}(s)^T\beta_0 ds)du \right]}{E[\sum_{j=1}^{n_i} R_{ij}(u)]} \right\}$$
$$- \mathbf{G}_n \left\{ \int_0^t \frac{E\left[\sum_{j=1}^{n_i} R_{ij}(u)\left[dN_{ij}(u) - R_{ij}(u)\mathbf{X}_{ij}(u)^T\beta_0 Q'(\int_0^u \mathbf{X}_{ij}(s)^T\beta_0 ds)du \right]\right] \sum_{j=1}^{n_i} R_{ij}(u)}{E[\sum_{j=1}^{n_i} R_{ij}(u)]^2} \right\}$$
$$- E \left\{ \int_0^t \frac{E[\sum_{j=1}^{n_i} R_{ij}(u)\mathbf{X}_{ij}(u)^T\beta_0 \nabla_\beta Q(\int_0^u \mathbf{X}_{ij}(s)^T\beta_0 ds)du]}{E[\sum_{j=1}^{n_i} R_{ij}(u)]} \right\} \sqrt{n}\left(\widehat{\beta} - \beta_0\right) + o_p(1).$$

Then theorem 1 holds using the Donsker theorem (c.f., van der Vaart and Wellner 1996)

Proof of Theorem 2 First note

$$
\frac{1}{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int R_{ij}(t)\left[I\left(\mathbf{X}_{ij}(t)^T\widehat{\beta}\le x\right)-\frac{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)\,I\left(\mathbf{X}_{kl}(t)^T\widehat{\beta}\le x\right)}{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)}\right]\right.
$$

$$
\times\left.\left\{dN_{ij}(t)-R_{ij}(t)\mathbf{X}_{ij}(t)^T\beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T\beta_0 ds\right)dt\right\}\right\}
$$

$$
=\frac{1}{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int R_{ij}(t)\right.
$$

$$
\times\left.\left[I\left(\mathbf{X}_{ij}(t)^T\widehat{\beta}\le x\right)-\frac{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)\,I\left(\mathbf{X}_{kl}(t)^T\widehat{\beta}\le x\right)}{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)}\right]dM_{ij}(t)\right\},
$$

where

$$
dM_{ij}(t)=dN_{ij}(t)-R_{ij}(t)d\Lambda_0(t)-R_{ij}(t)\mathbf{X}_{ij}(t)^T\beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T\beta_0 ds\right)dt.
$$

From the martingale property of $d\,M_{ij}$, we obtain

$$
\frac{1}{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int R_{ij}(t)\left[I\left(\mathbf{X}_{ij}(t)^T\widehat{\beta}\le x\right)-\frac{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)\,I\left(\mathbf{X}_{kl}(t)^T\widehat{\beta}\le x\right)}{\sum_{k=1}^{n}\sum_{l=1}^{n_k}R_{kl}(t)}\right]\right.
$$

$$
\times\left.\left\{dN_{ij}(t)-R_{ij}(t)\mathbf{X}_{ij}(t)^T\beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T\beta_0 ds\right)dt\right\}\right\}
$$

$$
=\frac{1}{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int R_{ij}(t)\right.
$$

$$
\times\left.\left[I\left(\mathbf{X}_{ij}(t)^T\beta_0\le x\right)-\frac{E[\sum_{l=1}^{n_k}R_{kl}(t)\,I\left(\mathbf{X}_{kl}(t)^T\beta_0\le x\right)]}{E[\sum_{l=1}^{n_k}R_{kl}(t)]}\right]dM_{ij}(t)\right\}
$$

$$
+o_p(n^{-1/2}).
$$

Therefore, we can rewrite $G_n(x)$ as

$$G_n(x)$$

$$= n^{-1/2} \mathbf{G}_n \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \right.$$

$$\times \left[I\left(\mathbf{X}_{ij}(t)^T \widehat{\beta} \le x\right) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \widehat{\beta} \le x\right)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right]$$

$$\times \left\{ dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T \beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds\right) dt \right\} \right]$$

$$- \frac{1}{n} \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n_i} \int R_{ij}(t) \left[I\left(\mathbf{X}_{ij}(t)^T \widehat{\beta} \le x\right) - \frac{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \widehat{\beta} \le x\right)}{\sum_{k=1}^{n} \sum_{l=1}^{n_k} R_{kl}(t)} \right] \right.$$

$$\left. \times R_{ij}(t)\nabla_\beta \left(\mathbf{X}_{ij}(t)^T \beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds\right)\right) dt \right\} (\widehat{\beta} - \beta_0) + o_p(|\widehat{\beta} - \beta_0|).$$

Since $I\left(\mathbf{X}_{ij}(t)^T \boldsymbol{\beta} \le x\right)$, as a bounded process index by t, $\boldsymbol{\beta}$ and x, belongs to some Donsker class, we obtain

$$n^{-1/2} G_n(x)$$

$$= \mathbf{G}_n \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \right.$$

$$\times \left[I\left(\mathbf{X}_{ij}(t)^T \beta_0 \le x\right) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \beta_0 \le x\right)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] dM_{ij}(t) \right]$$

$$- E \left\{ \sum_{j=1}^{n_i} \int R_{ij}(t) \left[I\left(\mathbf{X}_{ij}(t)^T \beta_0 \le x\right) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \beta_0 \le x\right)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.$$

$$\left. \times R_{ij}(t)\nabla_\beta \left(\mathbf{X}_{ij}(t)^T \beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds\right)\right) dt \right\} \sqrt{n}(\widehat{\beta} - \beta_0) + o_p(1).$$

Using the proof of Theorem 1, we have

$$n^{-1/2} G_n(x) = \mathbf{G}_n \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \right.$$

$$\left[I\left(\mathbf{X}_{ij}(t)^T \beta_0 \le x\right) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \beta_0 \le x\right)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] dM_{ij}(t) \right]$$

$$- E \left\{ \sum_{j=1}^{n_i} \int R_{ij}(t) \left[I\left(\mathbf{X}_{ij}(t)^T \beta_0 \le x\right) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t) I\left(\mathbf{X}_{kl}(t)^T \beta_0 \le x\right)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.$$

$$\left. \times R_{ij}(t)\nabla_\beta \left(\mathbf{X}_{ij}(t)^T \beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds\right)\right) dt \right\} \Sigma^{-1}$$

$$\times \mathbf{G}_n \left[\sum_{j=1}^{n_i} \int R_{ij}(t) \left[\mathbf{X}_{ij}(t) - \frac{E[\sum_{l=1}^{n_k} R_{kl}(t)\mathbf{X}_{kl}(t)]}{E[\sum_{l=1}^{n_k} R_{kl}(t)]} \right] \right.$$

$$\left. \times \left\{ dN_{ij}(t) - R_{ij}(t)\mathbf{X}_{ij}(t)^T \beta_0 Q'\left(\int_0^t \mathbf{X}_{ij}(s)^T \beta_0 ds\right) dt \right\} \right] + o_p(1).$$

Theorem 2 thus follows from the Donsker theorem.

## References

Aalen OO. A linear regression model for the analysis of lifetimes. Stat Med 1989;8:907–925. [PubMed: 2678347]

Andersen, PK.; Borgan, O.; Gill, RD.; Keiding, N. Statistical models based on counting processes. New York: Springer-Verlag; 1993.

Cai J, Schaubel DE. Marginal means/rates models for multiple type recurrent event data. Lifetime Data Analysis 2004;10:121–138. [PubMed: 15293628]

Cai T, Cheng SC, Wei LJ. Semiparametric mixed-effects models for clustered failure time data. J Am Stat Assoc 2002;97:514–522.

Cox DR. Regression models and life tables (with discussion). J R Stat Soc B 1972;34:187–220.

Gandy A, Jensen U. Checking a semiparametric additive risk model. Lifetime Data Anal 2005;11:451–472. [PubMed: 16328571]

Gandy A, Jensen U. On goodness-of-fit tests for Aalen's additive risk model. Scand J Stat 2005;32:425–445.

Huffer FW, McKeague IW. Weighted least squares estimation for Aalen's additive risk model. J Am Stat Assoc 1991;86:114–129.

Huster WJ, Brookmeyer R, Self SC. Modelling paired survival data with covariates. Biometrics 1989;45:145–156. [PubMed: 2655727]

Kupfer C. ETDRS research group. The diabetic retinopathy vitrectomy study (editorial). Am J Ophthalmol 1976;81:687–690. [PubMed: 1275051]

Lin DY, Ying ZL. Semiparametric analysis of the additive risk model. Biometrika 1994;81:61–71.

Lin DY, Ying ZL. Semiparametric analysis of general additive-multiplicative intensity models for counting processes. Annal Stat 1995;23:1712–1734.

McKeague IW, Sasieni PD. A partly parametric additive risk model. Biometrika 1994;81:501–514.

Murphy SA. Consistency in a proportional hazards model incorporating a random effect. Annal Stat 1994;22:712–731.

Murphy SA. Asymptotic theory for the frailty model. Annals Stat 1995;23:182–198.

Parner E. Asymptotic theory for the correlated gamma-frailty model. Annal Stat 1998;26:183–214.

Shen Y, Cheng SC. Confidence bands for cumulative incidence curves under the additive risk model. Biometrics 1999;55:1093–1100. [PubMed: 11315053]

Su J, Wei LJ. A lack-of-fit test for the mean function in a generalized linear model. J Am Stat Assoc 1991;86:420–426.

van der Vaart, AW.; Wellner, JA. Weak convergence and empirical processes. New York: Springer; 1996.

Yin G. Model checking for additive hazards model with multivariate survival data. J Multivariate Anal 2007;98:1018–1032.

Yin G, Cai J. Additive hazards model with multivariate failure time data. Biometrika 2004;91:801–818.

Yin G, Li H, Zeng D, Zhou Y. Partially linear additive hazards regression with varying coefficients. J Am Stat Assoc 2008;103:1200–1213.

Zeng D, Lin DY, Yin G. Maximum likelihood estimation in proportional odds model with random effects. J Am Stat Assoc 2005;100:470–483.

Zeng D, Yin G, Ibrahim J. Inference for a class of transformed hazards models. J Am Stat Assoc 2005;100:1000–1008.
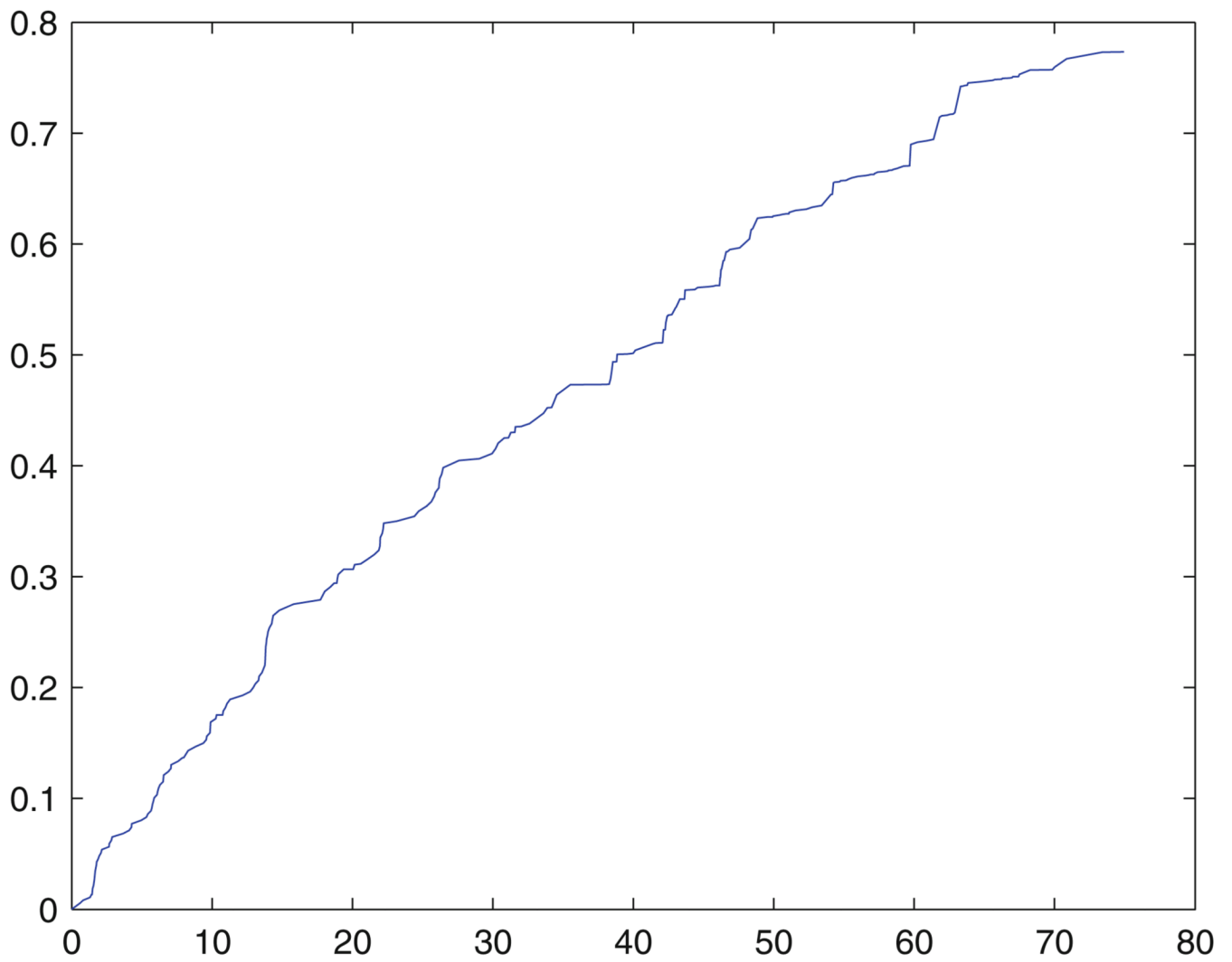
Fig. 1.
Estimated cumulative hazards function for θ = 0.79 in an untreated eye for a patient with
juvenile diabetics and age 16. The unit of x-axis is month.
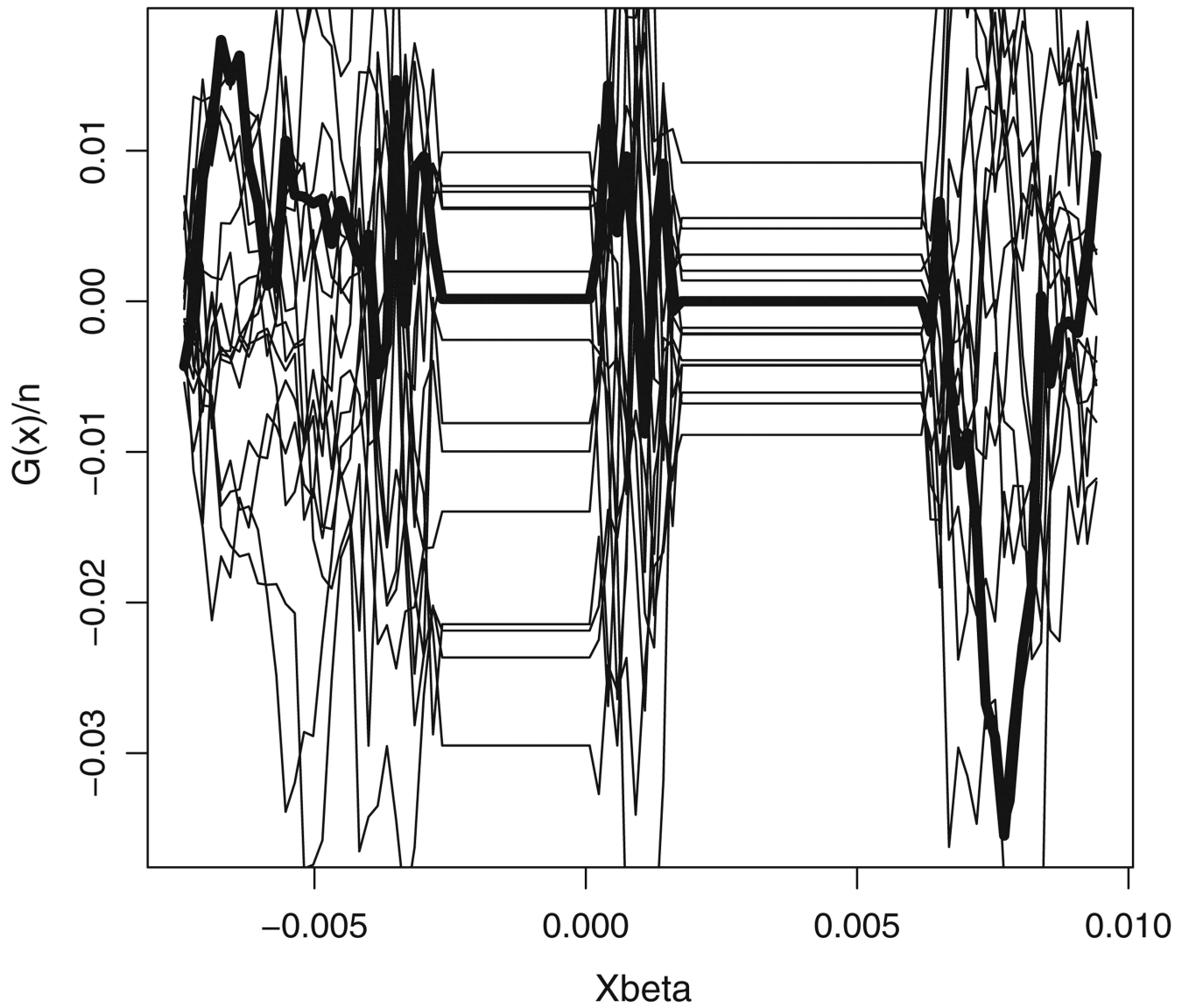
Fig. 2.
The goodness-of-fit statistic for $\theta = 0.79$ (in *thick solid line*) and simulated random paths from its limiting process.
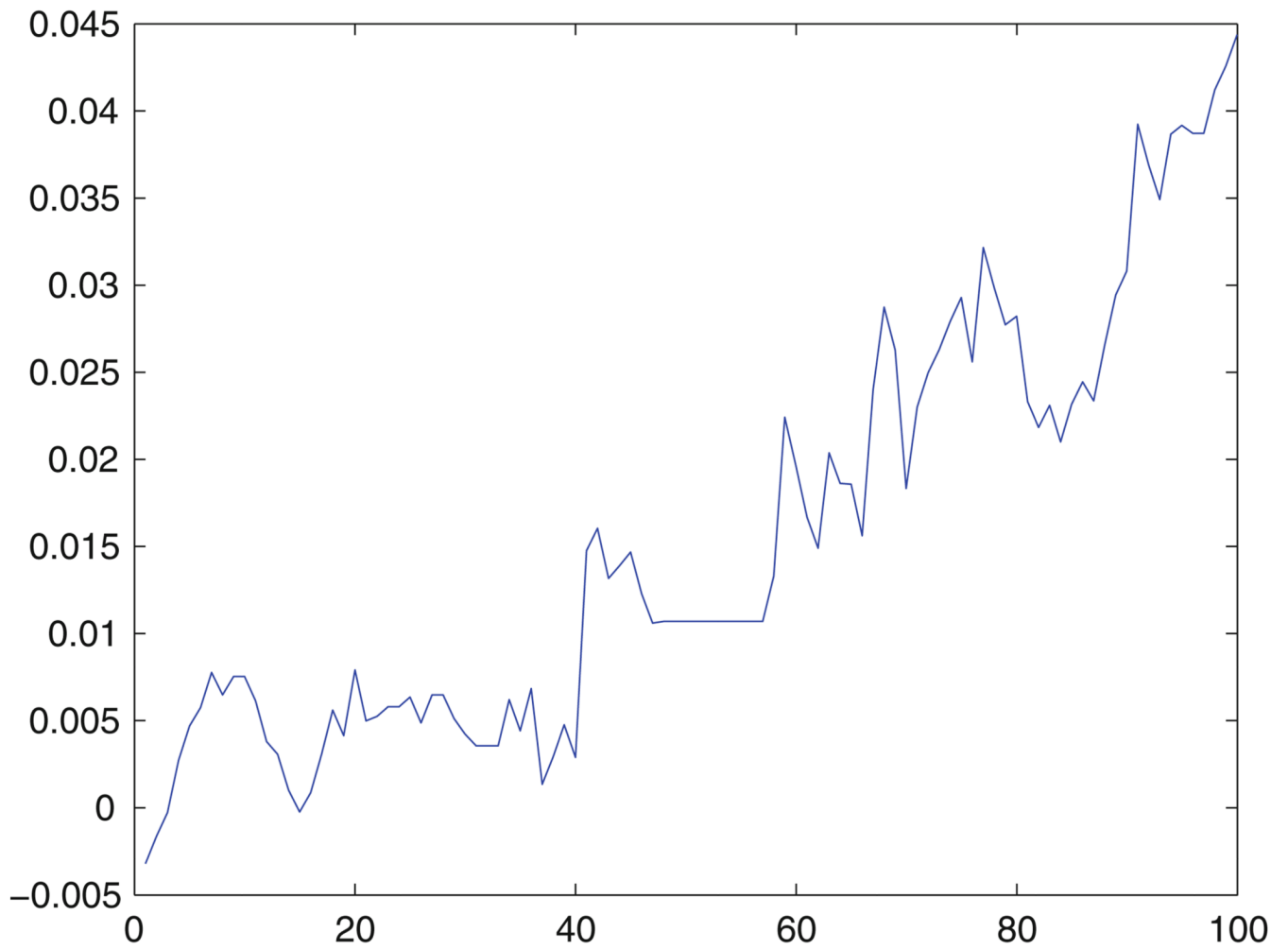
Fig. 3.
The goodness-of-fit statistic for the proportional hazards model with frailty: *x*-axis is the linear predictor from the proportional hazards model.

Table 1

Simulation results of estimates with transformed additive models

$Q(x)$	Par.	n=200				n=400			
		Bias	SE	ESE	CP	Bias	SE	ESE	CP
$Q(x) = x$	β_1	.013	.121	.124	.949	.013	.083	.086	.952
	β_2	.030	.230	.219	.936	.009	.158	.153	.950
$Q(x) = 2\log(1 + x/2)$	β_1	.029	.150	.154	.963	.006	.105	.105	.957
	β_2	.075	.322	.315	.955	.026	.217	.215	.946
$Q(x) = \log(1 + x)$	β_1	.025	.185	.185	.957	.010	.126	.126	.945
	β_2	.101	.439	.419	.947	.040	.279	.283	.955

Table 2

Empirical tail probability for goodness-of-fit statistics

Transformation	n=200	n=400
$Q(x) = x$	.054	.050
$Q(x) = 2 \log(1 + x/2)$	.033	.052
$Q(x) = \log(1 + x)$	.052	.052

Table 3

Inference with misspecified transformation under the null hypothesis

Transformation	n=200				n=400			
	Bias	**SE**	**ESE**	**α**	**Bias**	**SE**	**ESE**	**α**
$Q(x) = x$	-.004	.145	.149	.041	.001	.099	.104	.052
$Q(x) = 2\log(1 + x/2)$	.020	.147	.151	.051	.006	.105	.104	.062
$Q(x) = \log(1 + x)$	.010	.145	.151	.043	.008	.107	.104	.053

Table 4

Application to Diabetic Retinopathy Study

Covariate	$\theta = 0$			$\theta = 0.79$			$\theta = 1$		
	Est	**SE**	**P**	**Est**	**SE**	**P**	**Est**	**SE**	**P**
Treatment	–.0046	.0020	.0238	–.0044	.0019	.0235	–.0043	.0020	.0269
Type	.0035	.0040	.3866	.0045	.0044	.2968	.0049	.0045	.2791
Age	.0001	.0001	.4336	.0001	.0001	.4698	.0001	.0001	.4843
Treat*Type	–.0092	.0034	.0078	–.0093	.0043	.0299	–.0094	.0046	.0406