Ann Inst Stat Math (2015) 67:897–915
DOI 10.1007/s10463-014-0477-6

Estimation of copula-based models for lifetime medical costs

Xiao Bing Zhao · Xian Zhou

Received: 12 December 2012 / Revised: 16 May 2014 / Published online: 13 August 2014

Abstract Medical cost data are recorded through medical care and the cost is always related to some sojourn in the health status of the patient. The total medical cost accumulated in the entire lifetime of a life is of great interest to the health insurance industry and government policy makers. In this paper, we develop a new method to assess the lifetime medical cost up to the death time by incorporating the dynamics of change in the health status of the patient based on incomplete data. A copula model is proposed to fit the cost lifetime medical data subject to a terminal event (death). A two-stage estimation procedure is applied to draw the statistical inference of the marginals and the copula parameters. The asymptotic properties of the estimators are established, and a simulation is performed to evaluate the proposed model and estimation methods.

Keywords Dynamic change · Medical cost · Sojourn · Copula model · Two-stage estimation · Pseudo-likelihood

1 Introduction

In many medical studies, the occurrence of events over time can be described as the evolution of a finite-state stochastic process, with the state representing the health

X. B. Zhao (✉)
School of Mathematics and Statistics, Zhejiang University of Finance and Economics, 18 Xue-Yuan Street, Xia-Sha District, Hangzhou 310018, Zhejiang Province, China
e-mail: maxbzhao@hotmail.com

X. Zhou (✉)
Department of Applied Finance and Actuarial Studies, Macquarie University, North Ryde, NSW 2109, Australia
e-mail: xian.zhou@mq.edu.au

status or health condition. For example, the classical illness–mortality model describes health status as either “well”, “ill”, or “dead” (Gardiner et al. 2006). A multi-state model is therefore a natural extension of the survival model that considers only two state, “alive” or “dead”, and one possible transition from alive to dead.

The medical cost data can be collected from such a multi-state model in which the cost data are accumulated during each sojourn (the time spent in a state until the next transition) and depend on transition time, health state and patient characteristics (covariates). As common in the longitudinal studies, some patients provide incomplete data since they may be censored (drop out) before the endpoint of observation. In this setting, the usual techniques in survival analysis are not applicable because the accumulated cost at the failure (death) time is generally correlated with the accumulated cost at the censoring time (Castelli et al. 2007).

For these correlated observations, some methods can be found in the literature. For example, Carides (2000) and Polverejan et al. (2003) estimated the medical costs via a function of the failure time, which is based on the grouped cost data. Regression is an alternative in the cost analysis, which has been extensively discussed by many authors. The regression model, however, may be unsuitable for censored costs (Castelli et al. 2007) since the costs at failure and censored times tend to be positively correlated (Lin 2003; Etzioni et al. 1999). To model these correlated observations, in this paper, a copula model is proposed to model the correlation between the lifetime and the corresponding medical cost, where the cost data are collected up to the lifetime (death time) under a Markov process. This model has not been considered so far in the literature.

A Markov process is proposed to describe the dynamic movements between the health states of patients. Our interest is focused on estimating the transition probabilities, the distribution of sojourn and the accumulation rate function of the cost in different states. Since the effects of patient characteristics, such as demographics, treatments and clinical conditions, on these entities can be assessed on the basis of a proportional cost accumulation rate model, the covariate effects can be considered via the cost accumulation rate.

The rest of paper is organized as follows. In Sect. 2, we specify our model. The medical cost function and the copula parameters are estimated in Sect. 3 by the pseudo-likelihood method based on the estimated Markov process. The asymptotic properties of the estimators are established in Sect. 4. Sections 5 reports some simulation results, followed by concluding remarks in Sect. 6. Technical proofs are relegated to Appendix.

2 Model specifications

Let $X = \{X(t), t \geq 0\}$ denote a finite-state *time-inhomogeneous Markov process* with state space $E = \{1, 2, \dots, m\}$, where $X(t)$ represents the state occupied at time t. The transition probabilities $\{p_{kh}(s, t) : k, h \in E\}$ of X from time s to t are defined by

$$\Pr(X(t) = h | X(s) = k, X(u), u < s) = \Pr(X(t) = h | X(s) = k) = p_{kh}(s, t) \quad \text{for all } 0 \leq s \leq t \text{ and } k, h \in E. \tag{1}$$

The associated transition intensities $\alpha = \{\alpha_{kh} : k, h \in E\}$ are given by

$$\alpha_{kh}(t) = \lim_{\Delta t \to 0} \frac{p_{kh}(t, t + \Delta t)}{\Delta t}, \quad k \neq h, \quad \text{and} \quad \alpha_{kk} = -\sum_{h \neq k} \alpha_{kh}.$$

The cumulative (integrated) intensities are given by $\mathcal{A}_{kh}(t) = \int_0^t \alpha_{kh}(u)\,\mathrm{d}u$.

Let $\mathbf{P}(s,t) = (p_{kh}(s,t))_{m\times m}$ and $\mathbf{A}(u) = (\alpha_{kh}(u))_{m\times m}$ denote the transition probability matrix and the intensity matrix, respectively, which are related by the product-integral

$$\mathbf{P}(s,t) = \prod_{s<u\leq t} (\mathbf{I} + \mathbf{A}(u)\,\mathrm{d}u),$$

where $\mathbf{I}$ is the identity matrix (Andersen et al. 1993).

Let Z denote a p-dimensional generic covariate profile. The covariate effects on the transition intensities are modeled by $\alpha_{kh}(t) = \alpha_{kh0}(t)\exp(Z_{kh}^{\top}\mu_{kh})$, where $\alpha_{kh0}(t)$ denotes an unknown baseline intensity and μ_{kh} is an unknown $p \times 1$ parameter vector to be estimated (cf. Gardiner et al. 2006).

Suppose that the costs of patient i are incurred over a time interval $[0, T_i]$, $T_i < \infty$. The cost accumulation rate at time t in state h is denoted by $B_h^i(t)$, so that the cost incurred in interval $[t, t + \mathrm{d}t)$ is $Y_h^i(t)b_h^i(t|Z)\,\mathrm{d}t$, where $Y_h^i(t) = I\{X^i(t-) = h\}$ is the indicator function for occupying state h just before time t and $b_h^i(t|Z) = E[B_h^i(t)|X(t-) = h, Z]$ is the expected rate of the cost accumulation in state h at time t.

Given the initial state $X(0) = k$ and a fixed covariate Z, according to Gardiner et al. (2006), the expected total cost composed by all sojourn costs in state h and interval $[0, T_i]$, conditional on T_i, is given by

$$w_h^i(k, Z|T_i) = \int_0^{T_i} b_h^i(t|Z)\,p_{kh}(0, t|Z)\,\mathrm{d}t. \tag{2}$$

Let $\pi_k(0|Z) = \Pr(X(0) = k|Z)$ be the initial distribution. Given T_i, the expected cost is

$$w_i = \sum_k \sum_h \pi_k(0|Z) \int_0^{T_i} b_h^i(t|Z)\,p_{kh}(0, t|Z)\,\mathrm{d}t. \tag{3}$$

Model (3) has been investigated by Gardiner et al. (2006), in which the accumulation rate $B_h^i(s)$ is estimated by the regression method proposed in Lin et al. (1997) and the transition probabilities are estimated using the partial likelihood. The regression method, however, partitions the observation interval into sub-intervals to weaken the impacts of the correlated observations, which may not be the best way to deal with correlated data.

Suppose that the observed transition times are

$$0 = S_{i,0} < S_{i,1} < \cdots < S_{i,j} < \cdots < S_{i,J_i+1} = T_i \wedge Y_i,$$

where Y_i is a censoring variable. Obviously, the patient occupies the initial state up to the first transition time $S_{i,1}$. Assume that $p_{kh}(0, t)$ is right-continuous and piecewise constant between the observed transition times (which is the case for the nonparametric estimator of $p_{kh}(0, t)$ we will use). Then, $p_{kh}(0, t) = p_{kh}(0, S_{i,j-1})$ for $S_{i,j-1} \leq t < S_{i,j}$, $j = 1, \ldots, J_i + 1$. As a result, (3) implies (cf. Gardiner et al. 2006)

$$w_i = \sum_k \sum_h \sum_{j=1}^{J_i+1} \pi_k(0|Z) p_{kh}(0, S_{i,j-1}) \left(m_h^i(S_{i,j}|Z) - m_h^i(S_{i,j-1}|Z) \right), \tag{4}$$

where $p_{kh}(0, S_{i,j-1})$ and $m_h^i(t|Z) = \int_0^t b_h^i(u|Z)\, \mathrm{d}u$ are to be estimated.

The accumulated cost is correlated with the survival time (Castelli et al. 2007). The correlation of the cost and the failure time, however, is very complicated and cannot be diagnosed easily. In this paper, we propose a copula model to fit the accumulated cost data and the survival time. The main advantage of the copula model is its flexibility to choose the dependent structure of variables and the marginal distributions for the cost and the survival time.

As usual in the cost analysis, the marginal distribution of the lifetime cost can be assumed to be log-normal with mean w_i given by (4) and variance $\sigma^2 > 0$ in the literature (cf. Zhou 1998; Nixon and Thompson 2004). More specifically, conditional on T_i, the density function of the lifetime cost C_i for subject i is given by

$$f_{C_i|T_i}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{ \frac{-(\log x - w_i)^2}{2\sigma^2} \right\}, \quad x > 0. \tag{5}$$

The costs of patient i, however, are observable only in the interval $[0, T_i \wedge Y_i]$, which is not necessarily the whole lifetime interval $[0, T_i]$ due to censoring. Therefore, in this paper, we assume that the incomplete cost is log-normal with mean

$$\tilde{w}_i = \sum_k \sum_h \pi_k(0|Z) \int_0^{Y_i} b_h^i(t|Z) p_{kh}(0, t)\, \mathrm{d}t \tag{6}$$

(i.e., replace T_i by Y_i in (3)). The cost in (6) is referred to as the *total cost*, which differs from the *lifetime cost* in (3).

To best fit the correlated cost, we need the distribution of the variable T_i and Y_i. As common in survival analysis, the failure time T_i and censoring time Y_i are assumed to have proportional hazards as follows:

$$\lambda_{T_i}(t) = \lambda_{01}(t) \exp(Z_i^\top \beta), \quad \lambda_{Y_i}(y) = \lambda_{02}(y) \exp(Z_i^\top \eta), \tag{7}$$

where $\lambda_{01}(t)$ and $\lambda_{02}(y)$ are unspecified baselines hazard rates, and β and η are the coefficients of covariate Z_i to be estimated.

To account for the correlation between the total costs C_i, the failure time T_i, and the censoring time Y_i, we propose to model the joint distribution of (T_i, Y_i, C_i) by a trivariate copula function with bivariate margins. Specifically, the joint survival function of (T_i, Y_i, C_i) is given by

$$\Pr(T_i > t, Y_i > y, C_i > x) = \mathcal{C}_\rho(S_{T_i}(t), S_{Y_i}(y), S_{C_i}(x)), \tag{8}$$

where $\mathcal{C}_\rho(\cdot,\cdot,\cdot)$ is a trivariate copula function on $[0,1]\times[0,1]\times[0,1]$ with parameter ρ, $S_{C_i}(x)$ is the survival function of C_i with density given by (5), and $S_{T_i}(t)$ and $S_{Y_i}(y)$ are, respectively, the survival functions of T_i and Y_i with hazard rates given by (7).

Since $(T_i \wedge Y_i = t, \delta_i = 1) = (T_i = t, Y_i > t)$ and $(T_i \wedge Y_i = y, \delta_i = 0) = (T_i > y, Y_i = y)$, the conditional density of the total medical cost C_i given $(T_i \wedge Y_i = t, \delta_i = 1)$ or $(T_i \wedge Y_i = y, \delta_i = 0)$ can be defined, respectively, by (see Shih and Louis 1995)

$$f_{C_i|(T_i\wedge Y_i=t,\delta_i=1)}(x|t) = \frac{\mathcal{C}_\rho^{(13)}(u_1(t), u_2(t), u_3(x))}{\mathcal{C}_\rho^{(1)}(u_1(t), u_2(t))} f_{C_i}(x) \tag{9}$$

or

$$f_{C_i|(T_i\wedge Y_i=y,\delta_i=0)}(x|y) = \frac{\mathcal{C}_\rho^{(23)}(u_1(y), u_2(y), u_3(x))}{\mathcal{C}_\rho^{(2)}(u_1(y), u_2(y))} f_{C_i}(x), \tag{10}$$

where $(u_1(\cdot), u_2(\cdot), u_3(\cdot)) = (S_T(\cdot), S_Y(\cdot), S_C(\cdot))$, $\mathcal{C}_\rho^{(kl)} = \partial^2\mathcal{C}_\rho/\partial u_k \partial u_l$ for $(k,l) = (1,3)$ or (2,3), $\mathcal{C}_\rho(u_1, u_2)$ is a bivariate margin of $\mathcal{C}_\rho(u_1, u_2, u_3)$, and $\mathcal{C}_\rho^{(k)}(u_1, u_2) = \partial\mathcal{C}_\rho(u_1, u_2)/\partial u_k$, $k = 1, 2$. In this paper, we will apply a trivariate copula of the following form, which will be further studied in statistical inference and simulation in subsequent sections:

$$\mathcal{C}_\rho(u_1, u_2, u_3) = \left[(u_1 u_2)^{1-\rho} + u_3^{1-\rho} - 1\right]^{1/(1-\rho)}. \tag{11}$$

The copula in (11) implies that the survival time T and the censoring variable Y are independent, and the association between T and the medical cost C is the same as that between Y and C, both measured by the copula parameter ρ.

Remark 1 A more general trivariate copula is given by a trivariate survival function of the form:

$$\mathcal{C}_{\gamma,\rho}(u_1, u_2, u_3) = \left[\left(u_1^{1-\gamma} + u_2^{1-\gamma} - 1\right)^{(1-\rho)/(1-\gamma)} + u_3^{1-\rho} - 1\right]^{1/(1-\rho)}. \tag{12}$$

It has bivariate margins of the Clayton's survival copula (see Example 3.2 and Section 5.3 of Joe 1997): $\mathcal{C}_\gamma(u_1, u_2) = (u_1^{1-\gamma} + u_2^{1-\gamma} - 1)^{1/(1-\gamma)}$ and $\mathcal{C}_\rho(u_i, u_j) = (u_i^{1-\rho} + u_j^{1-\rho} - 1)^{1/(1-\rho)}$ for $(i, j) = (1, 3)$ or (2,3). This trivariate copula can model the association between the survival time T and the censoring variable Y if they are assumed to be dependent, and measure its size by γ. However, since only the minimum of T and Y is observed, it would not be possible to estimate all marginal distributions and the copula parameters ρ, γ from the data under model (12). Model (11), on the other hand, can avoid this difficulty by assuming independence between T and Y, and is a special case of (12) as $\gamma \to 1$. Moreover, when T and Y are independent and only their minimum is observed, it is reasonable to measure the associations between (T, C) and between (Y, C) by the same copula parameter ρ as assumed in model (11).

3 Pseudo-maximum likelihood

3.1 Marginal distributions

To construct the likelihood function of the copula model defined in (10), it is crucial to have explicit cumulative distribution functions (cdf's) of the variables T_i, Y_i and the total costs C_i in the interval $[0, T_i]$ or $[0, Y_i]$ given in (5).

Assume that $F_{T_i}(\cdot)$ and $F_{Y_i}(\cdot)$ are the cdf's of the failure time T_i and the censoring variable Y_i, respectively. Based on (7), $F_{T_i}(t)$ and $F_{Y_i}(y)$ can be expressed as

$$F_{T_i}(t) = 1 - [1 - F_{01}(t)]^{\exp(Z_i^\top \beta)} \quad \text{and} \quad F_{Y_i}(y) = 1 - [1 - F_{02}(y)]^{\exp(Z_i^\top \eta)},$$

where $F_{0k}(t) = 1 - \exp\{\int_0^t \lambda_{0k}(s)\,\mathrm{d}s\}$, $k = 1, 2$, are unspecified baseline cdf's determined by the baseline hazard rates $\lambda_{01}(t)$ and $\lambda_{02}(t)$ in (7).

The marginal cdf $F_{C_i}(x)$ of the total cost of subject i can be obtained from the log-normal density in (5). We incorporate the covariate Z_i into the sojourn cost b_h^i by

$$b_h^i(t|Z) = b_{0h}(t) \exp\left(Z_i^\top \alpha_h\right), \tag{13}$$

where $b_{0h}(t)$ is the baseline accumulation rate function and α_h is the coefficient of covariate Z_i to be estimated.

In this paper, the parameters β, η, α_h, σ^2, μ and copula parameter ρ together with the unspecified components $\alpha_{hk0}(t)$, $b_{0h}(s)$, $F_{01}(t)$, $F_{02}(y)$ are to be estimated. In the next section, the partial likelihood will be used to estimate the marginal cdf's $F_{T_i}(t)$ and $F_{Y_i}(y)$, and the multivariate likelihood will be employed to estimate $F_{C_i}(x)$.

3.2 Two-step parametric estimation

The pseudo-likelihood for (10) is computationally difficult to work with. Hence, we propose a two-step procedure to tackle the estimations required. It first estimates the marginal distributions of the failure time T_i and censoring variable Y_i, and then estimates the copula parameters ρ and the marginal distribution of the cost C_i based on an *ad hoc* estimator. We follow this two-step procedure to draw statistical inference, which differs from the two-step procedure proposed by Joe (2005).

Partial likelihood

To estimate the first margin for the terminal event, we make further specifications to our model. As usual in survival analysis, denote the censoring indicator by $\delta_i = 1$ if $T_i \le Y_i$, and 0 otherwise. For the observations $\{T_i, Y_i, \delta_i, i = 1, \ldots, n\}$, the distributions of T_i and Y_i can be easily estimated from the partial likelihood as follows.

Let $t_1 < \cdots < t_k$ be the ordered distinct (uncensored) failure times and $R(t)$ the risk set at time t. An estimator of β is obtained from the following partial likelihood:

$$\mathcal{L}_1(\beta) = \prod_{j=1}^{k} \frac{\exp\{Z_j^\top \beta\}}{\sum_{l \in R(t_j)} \exp\{Z_l^\top \beta\}}. \tag{14}$$

The Nelson–Aalen estimator for the cumulative baseline hazard rate $\Lambda_{01}(t) = \int_0^t \lambda_{01}(s)ds$ is given by

$$\hat{\Lambda}_{01}(t) = \sum_{i:t_i<t} \frac{\delta_i}{\sum_{l \in R(t_i)} \exp\{Z_l^\top \hat{\beta}\}}, \tag{15}$$

where $\hat{\beta}$ is the estimate of β from (14). Then, the cdf of T_i is estimated by

$$\hat{F}_T(t) = 1 - \exp\left\{-\hat{\Lambda}_{01}(t) \exp\left(Z^\top \hat{\beta}\right)\right\}. \tag{16}$$

Proceeding along the same line as (14)–(16), the cdf of the censoring variable Y_i can be estimated as well.

Pseudo-maximum likelihood

Note that the cost C_i is incomplete if $Y_i < T_i$. However, as the cost is log-normal with parameters σ and α_h defined in (5), the likelihood function of C_i conditional on T_i and Y_i of (10) can be written as:

$$\mathcal{L}_2(\cdot) = \prod_{i=1}^{n} f_{C_i}(x_i) \left[\frac{\mathcal{C}_\rho^{(13)}(u_i(t_i), u_2(t_i), u_3(x_i))}{\mathcal{C}_\rho^{(1)}(u_1(t_i), u_2(t_i))}\right]^{\delta_i} \left[\frac{\mathcal{C}_\rho^{(23)}(u_1(y_i), u_2(y_i), u_3(x_i))}{\mathcal{C}_\rho^{(2)}(u_1(y_i), u_2(y_i))}\right]^{1-\delta_i}. \tag{17}$$

The mean $\tilde{w}_i$ of C_i, however, involves unknown baseline functions $b_{h0}(s)$ and the transition probabilities $p_{kh}(0, t)$, which are to be estimated. The pseudo-likelihood estimation is used to complete the statistical inference in a two-step procedure. The first step is to estimate the transition probability function $p_{kh}(0, t)$ semi-parametrically as in Andersen et al. (1993). Then, in the second step, the likelihood is used to estimate the baselines b_{h0} together with some constants in $\tilde{w}_i$, with the transition probabilities $\mathbf{P}(s, t)$ replaced by their consistent estimators $\hat{\mathbf{P}}(s, t)$, and the cdf's $F_T(t)$, $F_Y(y)$ and their densities $f_T(t)$, $f_Y(y)$ replaced by their estimators defined in (14)–(16).

Since the Nelson–Aalen estimator $\hat{\mathbf{P}}(0, t) = (\hat{p}_{kh}(0, t))_{m \times m}$ of $\mathbf{P}(0, t)$ remains constant between transitions, we need an estimator of the total costs between consecutive transition times for (2). If $0 < S_{i,1} < S_{i,2} < \cdots < S_{i,J_i} < T_i \wedge Y_i$ denote all *observed* transition times in $(0, T_i \wedge Y_i)$, an estimator of $w_h^i(k, Z|T_i)$ in (2) is given by

$$\hat{w}_h^i(k, Z|T_i) = \sum_{j=1}^{J_i+1} \hat{p}_{kh}(0, S_{i,j-1}|Z) \left(m_h^i(S_{i,j}|Z) - m_h^i(S_{i,j-1}|Z)\right) \tag{18}$$

(see Gardiner et al. 2006), where $S_{i,0} = 0$, $S_{i,J_i+1} = T_i \wedge Y_i$ and $m_h^i(t|Z) = \int_0^t b_h^i(u)\,\mathrm{d}u$.

In Gardiner et al. (2006), an estimator of $b_h^i(t)$ is given by the standard regression method proposed by Lin et al. (1997). In this paper, for ease of representation of the model and methods, we make a parametric assumption of the baseline $h_{0h}(s)$ (some discussions will be given in Sect. 6). Together with the ad hoc estimator of $\mathbf{P}(s,t)$, $F_T(t)$ and $F_Y(y)$, the likelihood of (17) is referred to as the *pseudo-maximum likelihood*.

4 Asymptotic properties for estimators

We now proceed to establish the asymptotic properties of the estimators in this paper. For the marginal distribution functions $F_T(t)$ and $F_Y(y)$, the asymptotic properties of their estimators given in (16) depend on the asymptotic properties of the estimators of (β, η) and the baseline cdf's $F_{01}(t)$ and $F_{02}(y)$ defined in (14)–(15). To show the asymptotic properties of the estimators for the other marginal cdf $F_C(x)$ and the copula parameter ρ, two steps are needed since the estimation is based on the pseudo-likelihood. This is a new two-step procedure: first show that the estimator of the transition probabilities $\mathbf{P}(s,t)$ is consistent, and then the consistency and asymptotic normality of the estimators of $F_T(t)$ and $F_Y(y)$.

To show the asymptotic properties of $\hat{F}_T(t)$, $\hat{F}_Y(y)$ and $\hat{\mathbf{P}}(s,t)$ is tedious. We omit the details here, because the method is standard, such as: develop the convergence of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$ to Gaussian processes $\mathcal{G}_1$ and $\mathcal{G}_2$, say (this can be easily done by the existing results; cf. Andersen et al. 1993), express $\sqrt{n}(\hat{F}_T(t) - F_T(t))$ as a function of $\mathcal{G}_1$, $\mathcal{G}_2$ plus a remainder, and show that the remainder tends to zero.

In this section, we focus on the asymptotic properties of the pseudo-likelihood estimators of the accumulation rate function, such as $b_h^i(s)$, based on a consistent estimator $\hat{\mathbf{P}}(s,t)$. Our results are based on independent but non-identical distributions (see Hu 1998; Shorack and Wellner 1986). We first introduce some notations that are convenient in the theory of an analogy of empirical process (see Huang 1996). In the sequel, we will denote $P(t) = (p_{11}(0,t), \ldots, p_{mm}(0,t), F_T(t), F_Y(y))^\top$ for simplicity.

Throughout the rest of this section, we will denote by $F^i_{\phi,P(t)}$ the cdf of the cost C_i for every $(\phi, P(t))$ and its density by $f^i_{\phi,P(t)}$, where $\phi = (\nu_{0h}, \alpha_h : h \in E; \rho)^\top$ with true value $\phi_0 = (\nu^0_{0h}, \alpha^0_h : h \in E; \rho_0)^\top$ and ν_{0h} is the collection of all the parameters of $b_h^i(s)$.

Let F_n be the analogy of the empirical distribution of $\{X_i, i = 1, 2, \ldots, n\}$ such that $F_n M = n^{-1}\sum_{i=1}^n M(X_i)$ for any function $M(x)$. Similarly, $F_0 M = n^{-1}\sum_{i=1}^n \int M(x)\,\mathrm{d}F^i_{\phi,P(t)}(x)$. Then, the log-likelihood for a single observation is defined as

$$l(\phi, P(t)|X) = \log f_{\phi,P(t)}(X). \tag{19}$$

We also need the score function for ϕ, denoted by $\dot{l}_\phi(\phi, P(t)|X) = \partial l(\phi, P(t)|X) / \partial\phi$. The true values of $(\phi, P(t))$ are denoted by $(\phi_0, P_0(t))$. The true cdf $f_{\phi_0, P_0(t)}$ is sometimes simplified to f_0. For the rest of this paper, we assume that the following limits exist:

$$F_0\left[\dot{l}_\phi(\phi, P(t)|X)^j\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E\left[\dot{l}_\phi(\phi, P(t)|X)^j\right] < \infty, \quad j = 1, 2. \tag{20}$$

Further assume that the Lyapunov condition holds for random variables $\dot{l}_\phi(\phi_0, P_0(t)|X)$ and $F_0\ddot{l}_{\cdot\cdot}(\phi, P(t)|X) < \infty$, where the double dots on the top stand for twice-differentiation and each dot in the subscript represents ϕ or $P(t)$. Hence, for independent observations $X_1, \ldots, X_n$, the score function can be written as

$$F_n\dot{l}_\phi(\phi, P(t)|\cdot) = \frac{1}{n}\sum_{i=1}^{n} \dot{l}_\phi(\phi, P(t)|X_i). \tag{21}$$

We define the parameter space as follows. Denote the set of all monotone functions $f : R \to [0, 1]$ by Q. For some large $\tilde{A} > 0$ and small $\zeta > 0$, let

$$\mathcal{S}_0 = \left\{\phi : \phi^\top\phi < \tilde{A}\right\} \quad \text{and} \quad \mathcal{S}_\zeta = \{P(t) : \|P(t) - P_0(t)\| \le \zeta\}, \tag{22}$$

where $\|\cdot\|$ is the supremum norm.

We first state some conditions from Hu (1998), which correspond to Conditions 3, 1, 4, 2, 5 of Huang (1996), respectively. Note that $o_{p^*}(1)$ in the following representation indicates convergence to zero in outer probability in case that the term involved is not Borel measurable. The notation $O_{p^*}(\cdot)$ denotes bounded in probability F_0. In the sequel, we denote $\dot{l}_\phi(\phi_0, P_0(t)|X)$ by $\dot{l}_\phi(\phi_0, P_0(t))$ sometimes for simplicity.

Now, we return to the asymptotic properties of the pseudo-likelihood estimator $\hat{\phi}$. In the sequel, the definition of $F_0[\dot{l}_\phi(\phi, P(t))]$ is referred to (20). From (17), we can see that the log-likelihood function, say $l(\phi, P(t)|x)$, for a single observation is given by

$$\log(f_C(x)) + \delta\log\frac{\mathcal{C}^{(13)}(u_1, u_2, u_3)}{\mathcal{C}_\rho^{(1)}(u_1, u_2)} + (1-\delta)\log\frac{\mathcal{C}^{(23)}(u_1, u_2, u_3)}{\mathcal{C}_\rho^{(2)}(u_1, u_2)}.$$

Note that (20) implies $F_0\dot{l}_\phi(\phi, P(t)|X) < \infty$ (this holds under a bounded parameter space). Hence, by Lebesgue's Dominated Convergence Theorem,

$$F_0\left|\dot{l}_\phi(\phi, P(t))) - \dot{l}_\phi(\phi_0, P_0(t)))\right|^2 = o_p(1), \tag{23}$$

where $|\phi - \phi_0| \le \eta_n \downarrow 0$ and $\|P(t) - P_0(t)\| \le \kappa n^{-1/2}$ for some constant κ. Note that ϕ_0 is the unique point such that $\dot{l}_\phi(\phi, P_0(t)|x) = 0$. Thus, we obtain $\hat{\phi}$ by solving $\dot{l}_\phi(\phi, \hat{P}(t)|x) = 0$. We now give the asymptotic properties of the estimators. The proofs of these Theorems can be found in the Appendix.

Theorem 1 *Suppose that covariates satisfy $|z| \leq D$ for some positive constant D, the parameter space Φ ($\phi \in \Phi$) is bounded, and $\hat{P}(t)$ is a consistent estimator of $P(t)$ given in Sect. 3. Then, $\hat{\phi}$ solves equation $F_n \dot{l}_\phi(\hat{\phi}, \hat{P}(t)) = o_{p^*}(1/\sqrt{n})$ almost surely and converges in outer probability to ϕ_0.*

To prove Theorem 2, we need Lemma 1 below, which is a restatement of Vaart and Weller (1996) for the semiparametric model with an infinite-dimensional parameter space. Lemma 1 is an extension of Lemma 3.1.1 of Hu (1998) from the i.i.d. case to independent but non-identical distributions. The conditions of Lemma 1 provide a set of simple sufficient conditions for Condition 1 of Hu (1998), so we will only verify these conditions in Theorem 3 below.

Lemma 1 *Suppose the class of functions $\left\{\psi(\phi, P(t)) : |\phi - \phi_0| < \xi, \|P(t) - P_0(t)\| < \xi\right\}$ to be F_0-Donsker for some $\xi > 0$, and $F_0|\psi(\phi, P(t)|X) - \psi(\phi_0, P_0(t)|X)|^2 \to 0$ as $|\phi - \phi_0| \to 0$ and $\|P(t) - P_0(t)\| \to 0$. If $\hat{\phi} \overset{p^*}{\to} \phi_0$ and $\|\hat{P}(t) - P_0(t)\| \overset{p^*}{\to} 0$, then*

$$\left|\sqrt{n\tilde{h}}\,(F_n - F_0)\left[\psi(\hat{\phi}, \hat{P}(t)) - \psi(\phi_0, P_0(t))\right]\right| = o_{p^*}(1). \tag{24}$$

The next theorem gives the convergence rate of the pseudo-likelihood estimator $\hat{\phi}$.

Theorem 2 *Under the conditions of Theorem 1, $\sqrt{n}(\hat{\phi} - \phi_0) = O_{p^*}(1)$.*

Therefore, the asymptotic representation of $\hat{\phi}$ is an immediate consequence of the above results. Similar to Theorem 3.1.4 in Hu (1998), it can be shown that

$$\begin{aligned}\sqrt{n}(\hat{\phi} - \phi_0) = \left(-F_0 \ddot{l}_{\phi\phi}(\phi_0, P_0(t)\right)^{-1} \sqrt{n}\left\{\left(F_n \dot{l}_\phi(\phi_0, P_0(t)) - F_0 \dot{l}_\phi(\phi_0, P_0(t))\right)\right. \\ \left. + F_0 \ddot{l}_{\phi P(t)}(\phi_0, P_0(t)\left[\hat{P}(t) - P_0(t)\right]\right\} + o_{p^*}(1).\end{aligned} \tag{25}$$

Define $\Sigma(x) = E\left[n^{-1}(\ddot{l}_2)_{\phi\phi}(\phi_0, P_0(t))\right]$ and $\Gamma(x) = \mathrm{Var}(n^{-1}(\dot{l}_2)_\phi(\phi_0, P_0(t)))$. Under some classic regularities, we have the normal distribution of the estimator $\hat{\phi}$ below.

Theorem 3 *Under the conditions of Theorem 1, and some regularities, we have*

$$\begin{aligned}&\sqrt{n}\{[(\hat{\phi}_1 - \phi_1), \ldots, (\hat{\phi}_m - \phi_m)]^\top - \Sigma^{-1}(x) E[(n)^{-1}(\dot{l}_2)_\phi(\phi_0, P_0(t))]\} \\ &\underset{\to}{d}\, (n)^{-1}\Sigma(x)^{-1} N_q\,(0, V),\end{aligned} \tag{26}$$

where

$$V = \mathrm{Var}(\tilde{\Lambda}_1 + F_0 \ddot{l}_{\phi P(t)}(\phi_0, P_0(t))\Lambda_2) \;\; \textit{and} \;\; \tilde{\Lambda}_1 \sim N(\mathbf{0}, \Gamma(x)). \tag{27}$$

Remark 2 For the asymptotic variance V of $\sqrt{n}(\hat{\phi} - \phi_0)$ in Theorem 3, a precise expression can be found in Corollary 3.1.4 of Hu (1998) for the i.i.d. setup. In such a case, there exists a zero-mean $\alpha(X, \phi_0, P_0(t))$ such that

$$\sqrt{n} F_0 \ddot{l}_{\phi\phi}(\phi_0, P_0(t))[\hat{P}(t) - P(t)] = \sqrt{n} F_n \alpha(\cdot, \phi_0, P_0(t)) + o_p(1),$$

where $\alpha(\cdot, \phi_0, P_0(t))$ is defined by (3.1.21) in Hu (1998), and the variance of $\sqrt{n}(\hat{\phi} - \phi_0)$ is given by $V = \text{Var}[\dot{l}_\phi(\phi_0, P_0(t)|X)] + \text{Var}[\alpha(X, \phi_0, P_0(t))]$. Without such an $\alpha(X, \phi_0, P_0(t))$, however, a closed form of V is not available. In that case, we can estimate V by bootstrap, which will be illustrated by the simulations in the next section.

5 Simulation results

To assess the proposed model and methodology in this paper, a simulation is carried out as follows. We consider the case of a three-state model with the baseline transition intensity matrix:

$$Q_0 = \begin{pmatrix} -\alpha_{110} & \alpha_{120} & \alpha_{130} \\ \alpha_{210} & -\alpha_{220} & \alpha_{230} \\ \alpha_{310} & \alpha_{320} & -\alpha_{330} \end{pmatrix} = \begin{pmatrix} -2/3 & 1/3 & 1/3 \\ 0 & -1/2 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}. \tag{28}$$

That is, the change in health status is irreversible, which is realistic when the change is due to aging, and more flexible than the existing literature such as Hsieh et al. (2002). In this simulation, the covariate in $\alpha_{kh}(t)$ is ignored, so that $\alpha_{kh}(t) = \alpha_{kh0}(t), k, h = 1, 2, 3$.

The transition probabilities can be obtained as follow (Andersen et al. 1993). First define accumulative hazards as $H_{ij}(t_1, t_2) = \int_{t_1}^{t_2} \alpha_{ij}(t)\, \mathrm{d}t, i = 1, 2, 3, j = 1, 2, 3$. Then,

$$p_{11}(t_1, t_2) = \exp\{-H_{11}(t_1, t_2)\}. \tag{29}$$

This is the probability of remaining in state 1 between time t_1 and t_2 (referred to as *sojourn* in state 1). The probability of observing state 1 at time t_1 and state 2 at time t_2 is obtained by solving a linear differential equation, which leads to

$$p_{12}(t_1, t_2) = \int_{t_1}^{t_2} \exp\{-H_{11}(t_1, t)\}\, \alpha_{12}(t) \exp\{-H_{12}(t, t_2)\}\, \mathrm{d}t \tag{30}$$

and $p_{13}(t_1, t_2) = 1 - p_{11}(t_1, t_2) - p_{12}(t_1, t_2)$. Note that numerical integration may be needed to calculate (30) and get its estimator. Similarly, $p_{22}(t_1, t_2)$, $p_{23}(t_1, t_2)$ can be obtained.

The initial state $(\pi_1(0)\ \ \pi_2(0)\ \ \pi_3(0))$ is generated from a multinomial distribution with probabilities $(1/2\ \ 1/2\ \ 0)$ for each subject i at entry time $t_{i0} = 0$. For each subject i, the observed terminal time T_i and Y_i are assumed to have proportional hazards

$$\lambda_{T_i}(t) = \lambda_{01}(t) \exp\left(Z_i^\top \beta\right), \quad \lambda_{Y_i}(t) = \lambda_{02}(t) \exp\left(Z_i^\top \eta\right), \tag{31}$$

where the covariates are generated from uniform [0, 1] with coefficients $\beta = 0.5$ and $\eta = 1.2$. The baseline functions are taken as $\lambda_{01}(t) = 0.05t$ and $\lambda_{02}(t) = 0.05$, which correspond to the Weibull and exponential distributions of the baselines. We also assume that all the observations of the terminal events are censored beyond the observation endpoint $\tau = 6$.

For each state, the baseline accumulative rate functions b_{01}, b_{02} and b_{03} are defined as $b_{01} = \theta_1 t = 0.2t$, $b_{02} = \theta_2 = 6$ and $b_{03} = \exp(-\theta_3 t) = \exp(-3t)$. For subject i, the covariate Z_i is normally distributed with mean 0 and variance 2. The coefficient of covariate is $\alpha_h = 0.25$ for all states. Furthermore, we assume a trivariate copula given in (11) with parameter $\rho = 2.0$ for simplicity. The medical cost is generated from the log-normal marginal distribution with parameter $\sigma = 0.5$ and density

$$f_{C_i}(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left\{-\frac{(\log x - w_i)^2}{2\sigma^2}\right\}, \quad x > 0, \tag{32}$$

where w_i is defined in (4).

One can find a special algorithm for sampling from a trivariate Archimedean copula, such as McNeil (2008), Hofert (2008a, b). The partial derivatives for model (11), however, have a simple form, and hence a general algorithm using their conditional copulas (cf. Nelsen 2005) is available for generating random variables from the proposed model, which can proceed as follows:

Step 1 Generate independent random variables V_1, V_2 and V_3 from $U[0, 1]$.

Step 2 Calculate $W_1 = V_1$ and $W_2 = \mathcal{C}_{2|1}^{-1}(V_2|W_1)$, where $\mathcal{C}_{2|1}(u_2|u_1) = \partial\mathcal{C}_\rho(u_1, u_2)/\partial u_1$.

Step 3 Calculate $W_1 = V_1,\ W_2 = V_2$ and $W_3 = \mathcal{C}_{3|1,2}^{-1}(V_3|W_1, W_2)$, where

$$\mathcal{C}_{3|1,2}(u_3|u_1, u_2) = \frac{\partial^2\mathcal{C}_\rho(u_1, u_2, u_3)/\partial u_1 \partial u_2}{\partial^2\mathcal{C}_\rho(u_1, u_2)/\partial u_1 \partial u_2}. \tag{33}$$

Let $\tilde{\mathcal{C}}(u_1, u_2, u_3) = (u_1 u_2)^{1-\rho} + u_3^{1-\rho} - 1$. Then $\mathcal{C}_{3|1,2}$ has a close form under model (11):

$$\begin{aligned}\mathcal{C}_{3|1,2}(u_3|u_1, u_2) &= \rho(u_1 u_2)^{1-2\rho}\tilde{\mathcal{C}}^{(1-2\rho)/(\rho-1)}(u_1, u_2, u_3)\\ &\quad -(\rho-1)(u_1 u_2)^{-\rho}\tilde{\mathcal{C}}^{-\rho/(1-\rho)}(u_1, u_2, u_3).\end{aligned}$$

Step 4 Calculate $C_i = S_{C_i}^{-1}(W_3)$, $Y_i = S_{Y_i}^{-1}(W_2)$ and $T_i = S_{T_i}^{-1}(W_1)$, where $S_{C_i}(\cdot)$, $S_{Y_i}(\cdot)$ and $S_{T_i}(\cdot)$ are the cdf's of Y_i, C_i and T_i, respectively. Then, a dependent triplet (T_i, Y_i, C_i) is generated from the model $\mathcal{C}_\rho(S_{T_i}(t), S_{Y_i}(y), S_{C_i}(x))$.

In each simulation, we replicate 1000 times with sample sizes of 200 or 1000. The simulation results are presented in Tables 1 and 2 below for samples of sizes 200 and 1000, respectively. In Tables 1 and 2, $\hat{\beta}$ denotes the estimator of β, and so on; "Mean" represents the average of estimates and "Std" is the standard error over 1000 replicate samples.

The estimated transition probabilities $p_{11}(0, t)$, $p_{13}(0, t)$, $p_{22}(0, t)$ and $p_{33}(0, t)$ are listed in Figs. 1, 2, 3 and 4 as well. From Tables 1 and 2 and Figs. 1, 2, 3 and 4, we can see that the estimated parameters and curves perform reasonably well based on the proposed semiparametric model.

Table 1 Summaries of parameter estimators with $n = 200$

Estimator	$\hat{\beta}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\rho}$
Mean	0.465	1.241	0.255	0.463	0.154	6.301	3.570	2.378
Std	0.374	0.507	0.009	0.031	0.266	0.489	0.535	0.320

Table 2 Summaries of parameter estimators with $n = 1000$

Estimator	$\hat{\beta}$	$\hat{\eta}$	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\rho}$
Mean	0.512	1.208	0.246	0.442	0.272	6.019	3.142	2.110
Std	0.103	0.329	0.009	0.015	0.129	0.174	0.381	0.107

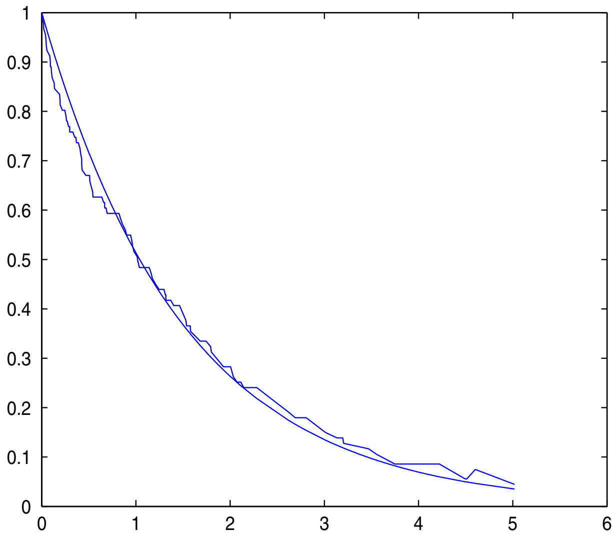

Fig. 1 The true and estimated $p_{11}(t)$

6 Concluding remark

In this paper, we investigated the modeling and analysis of medical cost data of the survival time. A copula model is proposed to capture the dependence between the medical cost and the terminal event, in which the occurrence of the medical cost over time can be described as the evolution of a finite-state stochastic process. A semiparametric model is proposed to describe the medical cost, and the pseudo-likelihood method is employed to estimate the parametric and nonparametric components of the proposed model. The asymptotic properties have been established as well. The model with copula function to fit the medical cost and the terminal event makes a new contribution to the literature.

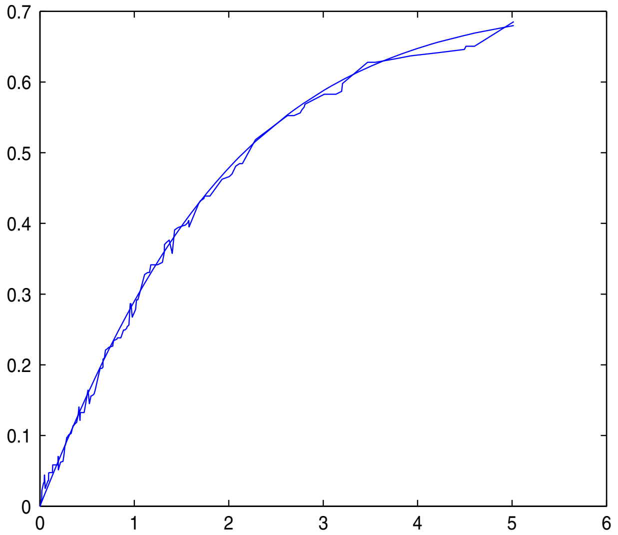

Fig. 2 The true and estimated $p_{13}(t)$

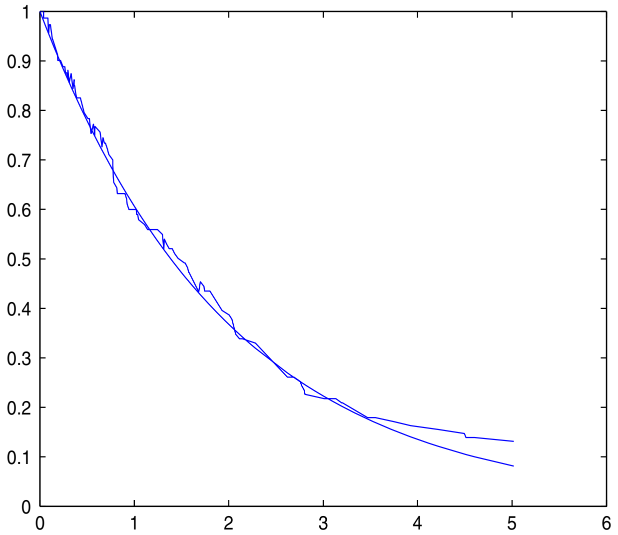

Fig. 3 The true and estimated $p_{22}(t)$

Our focus in this paper is based on the total cost of the subject. We may also consider the case with interest in the sojourn cost and model the relationship between the sojourn and its cost by a copula function. Then, the recurrent event model can be applied to fit these data. The sojourn and its related medical cost might be modeled as nonparametric or semiparametric such as proportional hazards. The marginal proportional hazards assumption, however, is different from the medical cost data with copula model, since the information censoring can be considered through this dependence structure. The

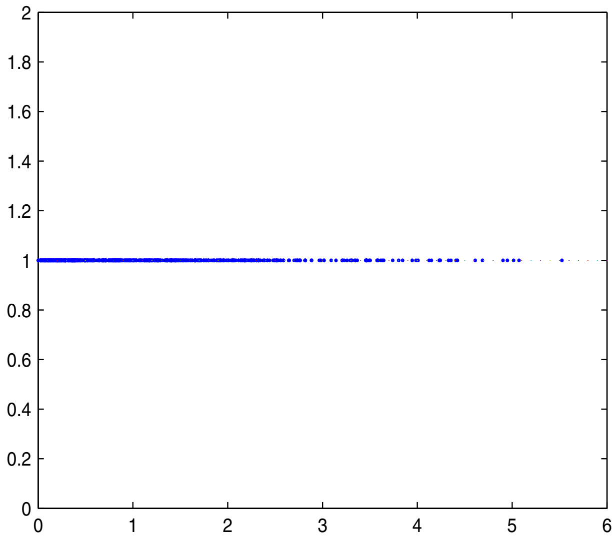

Fig. 4 The true and estimated $p_{33}(t)$

inference function of the margins for the two-step method (Joe 2005) can be used to draw statistical inference. These issues are of interest in further studies.

Appendix

Proof of Theorem 1

Proof Since ϕ_0 is the unique solution to $F_0\dot{l}_\phi(\phi, P(t)) = 0$, this implies that for any fixed $\varepsilon > 0$, there is a $\delta > 0$ such that $F_0[|\hat{\phi} - \phi_0| > \varepsilon] \leq F_0[|F_0\dot{l}_\phi(\hat{\phi}, P_0(t))| > \delta]$. Then, the consistency of $\hat{\phi}$ will follow from $F_0\dot{l}_\phi(\hat{\phi}, P_0(t))| \rightarrow_{p^*} 0$.

To do this, first note that since $\|\hat{P}(t) - P(t)\| = o_{P^*}(1)$, there exists a sequence $\{\delta_n\} \downarrow 0$ such that $\|\hat{P}(t) - P(t)\| \leq \delta_n$ with probability approaching 1. Hence, we have the inequality $|F_0\dot{l}_\phi(\hat{\phi}, P_0(t))| \leq |F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))| + |F_0\dot{l}_\phi(\hat{\phi}, P_0(t)) - F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))|$ by taking $P(t) = \hat{P}(t)$. The first term $|F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))|$ is $o_{p^*}(1)$ by the definition of estimator from the pseudo-likelihood. The second term is bounded by

$$\left|(F_n - F_0)\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right| + \left|F_0(\dot{l}_\phi(\hat{\phi}, \hat{P}(t)) - \dot{l}_\phi(\hat{\phi}, P_0(t)))\right|, \tag{34}$$

which obviously tends to zero when $|P(t) - P_0(t)| \leq \eta_n \downarrow 0$ by (23). Thus, it suffices to show that the class of functions $\Psi_\zeta \equiv \{\dot{l}_\phi(\phi, P(t)) : \phi \in \mathcal{S}_0, |P(t) - P_0(t)| \leq \zeta\}$ is a VC-class for some $\zeta > 0$, where $\mathcal{S}_0$ is defined in (22). This implies that the uniform strong law of large numbers holds, i.e., $\sup_{f \in \Psi_\zeta} |(F_n - F_0)f| \overset{p}{\rightarrow} 0$. It can be easily verified by Example 2.2.21 of Vaart and Weller (1996), Lemma 2.6.18 of Vaart and

Weller (1996) and Lemma 5.1.1 in Hu (1998) that Ψ_ζ is a VC-class by the boundness of the parameter space. The consistency of $\hat{\phi}$ then follows from (34). □

Proof of Theorem 2

Proof Since $|\hat{\phi} - \phi_0| \to o_{p^*}(1)$ and $\|\hat{P}(t) - P_0(t)\| \to O_{p^*}(n^{-\varepsilon})$ with $\varepsilon > 0$, there exists a sequence $\{\delta_n\} \downarrow 0$ and $\kappa > 0$ such that $|\hat{\phi} - \phi_0| \le \delta_n$ and $\|\hat{P}(t) - P_0(t)\| \le \kappa n^{-\varepsilon}$ with probability approaching one.

First, we can show the smoothness condition via direct calculations

$$\begin{aligned}
&\Big|\sqrt{n}\left[F_0\dot{l}_\phi(\phi, P(t)) - F_0\dot{l}_\phi(\phi_0, P_0(t))\right] - \sqrt{n}F_0\ddot{l}_{\phi\phi}(\phi_0, P_0(t))(\phi - \phi_0) \\
&\qquad - \sqrt{n}F_0\ddot{l}_{\phi P(t)}(\phi_0, P_0(t))[P(t) - P_0(t)]\Big| \\
&\quad = o_{p^*}(\sqrt{n}|\phi - \phi_0|) + O_{p^*}(\sqrt{n}|P(t) - P_0(t)|) \\
&\quad = o_{p^*}(1 + \sqrt{n}|\phi - \phi_0|). \qquad (35)
\end{aligned}$$

Hence, (35) is equal to $o(|\phi - \phi_0|)$ for $|\phi - \phi_0| < \delta_n \downarrow 0$.

On the other hand, we verify the stochastic equicontinuity condition:

$$\left|\sqrt{n}(F_n - F_0)\left[\dot{l}_\phi(\hat{\phi}, \hat{P}(t)) - \dot{l}_\phi(\phi_0, P_0(t))\right]\right| = o_{p^*}(1). \qquad (36)$$

$\dot{l}_\phi(\phi, P(t)) - \dot{l}_\phi(\phi_0, P_0(t))$ can be easily found by their definitions. Let

$$F_\zeta = \left\{\dot{l}_\phi(\phi, P(t)) - \dot{l}_\phi(\phi_0, P_0(t)) : |\phi - \phi_0| \le \zeta, \|P(t) - P_0(t)\| \le \zeta\right\}. \qquad (37)$$

Similar to Theorem 1, we can show that F_ζ is F_0-Donsker for some $\zeta > 0$. Thus, (36) follows from the VC-class (37) and (23) by Lemma 1.

By (36), we have

$$\begin{aligned}
&\sqrt{n}\left|(F_n - F_0)\dot{l}_\phi(\hat{\phi}, \hat{P}(t)) - (F_n - F_0)\dot{l}_\phi(\phi_0, P_0(t))\right| \\
&\quad = o_{p^*}(1) + o_{p^*}\left(\sqrt{n}\left|F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right|\right) + \left(\sqrt{n}\left|F_0\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right|\right). \qquad (38)
\end{aligned}$$

By the triangular inequality $-|a| + |b| - |c| \le |a - b - c|$ and the fact that $F_0\dot{l}_\phi(\phi_0, P_0(t)) = 0$,

$$\begin{aligned}
&\sqrt{n}\left|F_0\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right| - \sqrt{n}\left|F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right| - \sqrt{n}\left|F_n\dot{l}_\phi(\phi_0, P_0(t))\right| \\
&\quad = \sqrt{n}\left|(F_n - F_0)\left(\dot{l}_\phi(\hat{\phi}, \hat{P}(t)) - \dot{l}_\phi(\phi_0, P_0(t))\right)\right| \quad \text{(see (38))} \\
&\quad = o_{p^*}(1) + o_{p^*}\left(\sqrt{n}\left|F_n\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right|\right) + \left(\sqrt{n}\left|F_0\dot{l}_\phi(\hat{\phi}, \hat{P}(t))\right|\right), \qquad (39)
\end{aligned}$$

which implies

$$\begin{aligned}
&\sqrt{n}\left|F_0\dot{l}_\phi(\hat{\phi},\hat{P}(t))\right|[1-o_{p^*}(1)]\\
&\quad\leq o_{p^*}(1)+\sqrt{n}\left|F_n\dot{l}_\phi(\hat{\phi},\hat{P}(t))\right|[1-o_{p^*}(1)]+\sqrt{n}\left|F_n\dot{l}_\phi(\phi_0,P_0(t))\right|\\
&\quad= o_{p^*}(1)+o_{p^*}(1)+O_{p^*}(1),
\end{aligned}\tag{40}$$

where the last equation follows from (36) and (20). Hence,

$$\sqrt{n}\left|F_0\dot{l}_\phi(\hat{\phi},\hat{P}(t))\right|=O_{p^*}(1).\tag{41}$$

By (41) together with $F_0\dot{l}_\phi(\phi_0,P_0(t))=0$ and $-|a|+|b|-|c|\leq|a-b-c|$, (35) implies

$$\begin{aligned}
&-O_{p^*}(1)+\left|\sqrt{n}F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))(\hat{\phi}-\phi_0)\right|-\left|\sqrt{n}F_0\ddot{l}_{\phi P(t)}(\phi_0,P_0(t))(\hat{P}(t)-P_0(t))\right|\\
&\quad\leq\Big|\sqrt{n}[F_0\dot{l}(\hat{\phi},\hat{P}(t))-F_0\dot{l}(\phi_0,P_0(t))]-\sqrt{n}F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))(\hat{\phi}-\phi_0)\\
&\qquad-\sqrt{n}F_0\ddot{l}_{\phi P(t)}(\phi_0,P_0(t))(\hat{P}(t)-P_0(t))\Big|\\
&\quad=O_{p^*}\left(1+\sqrt{n}|\hat{\phi}-\phi|\right).
\end{aligned}\tag{42}$$

Since the $m\times m$ matrix $F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))$ is nonsingular, there exists a constant $\kappa_1>0$ such that as $|\phi-\phi_0|\to 0$,

$$\left|F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))\,(\phi-\phi_0)\right|\geq\kappa_1|\phi-\phi_0|.\tag{43}$$

On the other hand, by the following equation

$$F_0\dot{l}(\phi,P_0(t))-F_0\dot{l}(\phi_0,P_0(t))=F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))(\phi-\phi_0)+o(|\phi-\phi_0|),\tag{44}$$

we have $F_0\ddot{l}_{\phi P(t)}(\phi_0,P_0(t))(\hat{P}(t)-P_0(t))=O_{p^*}(1)$. Combining this with inequality (42) yields

$$\begin{aligned}
O_{p^*}(1)&\geq\left|\sqrt{n}F_0\ddot{l}_{\phi\phi}(\phi_0,P_0(t))(\hat{\phi}-\phi_0)\right|-\left|\sqrt{n}F_0\ddot{l}_{\phi P(t)}(\phi_0,P_0(t))(\hat{P}(t)-P_0(t))\right|\\
&\quad-O_{p^*}(1+\sqrt{n}|\hat{\phi}-\phi|)\geq c_1\sqrt{n}|\hat{\phi}-\phi|-O_{p^*}(1)-o_{p^*}(1+\sqrt{n}|\hat{\phi}-\phi|)\\
&=[O_{p^*}(1)-o_{p^*}(1)]\sqrt{n}|\hat{\phi}-\phi|-O_{p^*}(1).
\end{aligned}\tag{45}$$

Hence, $\sqrt{n}|\hat{\phi}-\phi|=O_{p^*}(1)$ in outer probability. □

Proof of Theorem 3

Proof By (25), we need to show

$$\sqrt{n}\,(F_n - F_0)\,\dot{l}_\phi(\phi_0, P_0(t)) \xrightarrow{d} \Lambda_1 \tag{46}$$

where $\Lambda_1 = (\Lambda_{11}, \ldots, \Lambda_{1q})^\top \sim N_q\,(0, \Gamma(x))$ with a $q \times q$ positive definite covariance matrix $\Gamma(x) = \Sigma_{11}$. This can be easily done by calculating its mean and variance. Then, (46) follows from the Lyapounov Central Limit Theorem. On the other hand, by the asymptotic results of transition probabilities in Andersen et al. (1993), we have $\sqrt{n}(\hat{P}(t) - P_0(t)) \xrightarrow{d} \Lambda_2(t)$, where $\{\Lambda_2(t), t \in T\}$ is a Gaussian process with mean zero and auto-covariance given by $\Sigma_{22}(t, t') = \mathrm{Cov}(\Lambda_2(t), \Lambda_2(t'))$ for any $t, t' \in T$. The cross-covariance function between Λ_{1i} and $\Lambda_2(t)$ for $t \in T$ is denoted by $\Sigma_{1i2} = \mathrm{Cov}(\Lambda_{1i}, \Lambda_2(t))$. The conclusion of the theorem then follows from (25). □

Acknowledgments We are truly grateful to the Associate Editor and the reviewers for their valuable and constructive comments and suggestions that helped to improve our paper substantially. This work was partially supported by the NSFC under Grant No. 11271317, Zhejiang Provincial Natural Science Foundation under Grant No. LY12A01017, Zhejiang Provincial Planning Projects of Philosophy and Social Science under Grant No. 12JCJJ17YB, and the Institute of Actuaries of Australia Research Grant. The authors are grateful to Miss Cuiliu Xiao for her support of the simulation in this paper.

References

Andersen, P. K., Borgan, O., Gill, R. D., Keiding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

Carides, G. W. (2000). A regression-based method for estimating mean treatment cost in the presence of right-censoring. *Biostatistics*, *1*, 299–313.

Castelli, C., Combescure, C., Foucher, Y., Daures, J. P. (2007). Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. *Statistics in Medicine*, *26*, 5557–5571.

Etzioni, R. D., Feuer, E. J., Sullivan, S. D., Lin, D. Y., Hu, C. C., Ramsey, S. D. (1999). On the use of survival analysis techniques to estimate medical care costs. *Journal of Health Economics*, *18*, 365–380.

Gardiner, J. C., Luo, Z. H., Bradley, C. J., Sirbu, C. U., Given, C. W. (2006). A dynamic model for estimating change in health status and costs. *Statistics in Medicine*, *25*, 3648–3667.

Hofert, M. (2008a). Sampling Archimedean copulas. *Computational Statistics and Data Analysis*, *52*, 5163–5174.

Hofert, M. (2008b). Efficiently sampling nested Archimedean copulas. *Computational Statistics and Data Analysis*, *55*, 57–70.

Hsieh, H. J., Chen, T. H. H., Chang, S. H. (2002). Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Statistics in Medicine*, *21*, 3369–3382.

Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, *24*, 540–568.

Hu, H. L. (1998). Pseudo Maximum Likelihood Estimation for Semiparametric Models. Ph.D. Thesis, University of Washington.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation methods for copula-based models. *Journal of Multivariate Analysis*, *94*, 401–419.

Lin, D. Y., Feuer, E. J., Etzioni, R., Wax, Y. (1997). Estimating medical costs from incomplete follow-up data. *Biometrics*, *53*, 419–434.

Lin, D. Y. (2003). Regression analysis of incomplete medical cost data. *Statistics in Medicine*, *22*, 1181–1200.

McNeil, A. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation*, *78*, 567–581.

Nelsen, B. (2005). *An Introduction to Copulas* (2nd ed.). New York: Springer.

Nixon, R. M., Thompson, S. G. (2004). Parametric modeling of cost data in medical studies. *Statistics in Medicine*, *23*, 1311–1331.

Polverejan, E., Gardiner, J. C., Bradley, C. J., Rovner, M. H., Rovner, D. (2003). Estimating mean hospital cost as a function of length of stay and patient characteristics. *Health Economics*, *12*, 935–947.

Shih, J. H., Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, *51*, 1384–1399.

Shorack, G. R., Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. New York: John Wiley and Sons.

Van Der Vaart, A. W., Weller, J. A. (1996). *Weak Convergence and Empirical Processes* (2nd ed.). New York: Springer.

Zhou, X. H. (1998). Estimation of the log-normal mean. *Statistics in Medicine*, *17*, 2251–2264.