Computational Statistics and Data Analysis 83 (2015) 151–167

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A class of transformed hazards models for recurrent gap times

Fangyuan Kang [a], Liuquan Sun [a], Xingqiu Zhao [b,*]

[a] Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100190, China

[b] Department of Applied Mathematics, The Hong Kong Polytechnic University Shenzhen Research Institute, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 10 July 2013
Received in revised form 17 September 2014
Accepted 5 October 2014
Available online 16 October 2014

Keywords:
Estimating equations
Gap times
Model checking
Recurrent events
Transformed hazards model

ABSTRACT

In this article, a class of transformed hazards models is proposed for recurrent gap time data, including both the proportional and additive hazards models as special cases. An estimating equation-based inference procedure is developed for the model parameters, and the asymptotic properties of the resulting estimators are established. In addition, a lack-of-fit test is presented to assess the adequacy of the model. The finite sample behavior of the proposed estimators is evaluated through simulation studies, and an application to a clinic study on chronic granulomatous disease (CGD) is illustrated.

1. Introduction

Recurrent events data are commonly encountered in medical and observational studies where each subject may experience a particular event repeatedly over time. Examples of such events include repeated hospitalization, multiple infection episodes, tumor recurrences, recurrent economic recessions, and repeated breakdowns of an automobile. In these studies, it is often of interest to assess the effects of covariates on certain features of the recurrent event times. The statistical analysis of such data is challenging due to the dependence of the recurrent event times within each individual and the presence of censoring such as the loss to follow-up. To analyze recurrent event data, the focus can be laid on two types of time scale: the time since enrollment and the time between two successive recurrent events (i.e., the gap time).

When the time since enrollment is used as time index, recurrent events of a subject are modeled as the realization of an underlying counting process (Cook and Lawless, 2007), and a variety of statistical methods has been proposed in the literature. For example, Prentice et al. (1981), Andersen and Gill (1982) and Zeng and Lin (2006) proposed some intensity-based methods. Nielsen et al. (1992), Murphy (1995) and Zeng and Lin (2007) developed some frailty model approaches. Lawless and Nadeau (1995), Lin et al. (2000), Schaubel et al. (2006) and Sun et al. (2011) considered some marginal means and rates models. Cook and Lawless (2007) provided an excellent review of statistical methods for the analysis of this type of data.

In many applications, however, the gap time is a natural outcome of interest (Gail et al., 1980). Some methods have been developed for the analysis of recurrent gap time data (Huang and Chen, 2003; Schaubel and Cai, 2004; Strawderman, 2005; Luo and Huang, 2011). For example, Huang and Chen (2003), Schaubel and Cai (2004) and Darlington and Dixon (2013)

* Corresponding author.
E-mail address: xingqiu.zhao@polyu.edu.hk (X. Zhao).

http://dx.doi.org/10.1016/j.csda.2014.10.005
0167-9473/

proposed the proportional hazards model for the gap times. Chang (2004) and Strawderman (2005) considered some accelerated gap time models. Sun et al. (2006) discussed the additive hazards model for the gap times. In addition, some authors suggested various nonparametric models for the gap time distribution (e.g., Lin et al., 1999; Wang and Chang, 1999; Peña et al., 2001 and Du, 2009). Luo and Huang (2011) demonstrated that many existing methods for recurrent gap time data can be viewed as weighted risk-set methods.

Recently, the semiparametric transformation models have been studied extensively in survival analysis (e.g., Chen et al., 2002; Zeng et al., 2005 and Zeng and Lin, 2007). However, there is a dearth of suitable transformation models for the analysis of recurrent gap time data. Lu (2005) studied the semiparametric linear transformation models for the gap times, which include the proportional hazards and proportional odds models as special case. Note that this class of models does not contain the additive hazards model as a special case. For classical survival data, Zeng et al. (2005) proposed a class of transformed hazards models which encompasses the proportional and additive hazards models and which accommodates time-varying covariates. In this paper, we consider this class of transformed hazards models for the analysis of recurrent gap time data, and propose an estimation procedure for the model parameters, which is easy to implement.

The remainder of the paper is organized as follows. In Section 2, we introduce data structure and the proposed models. Estimation procedures are presented for the model parameters, and the asymptotic properties of the proposed estimators are established. In Section 3, we develop a technique for checking the adequacy of the proposed model. Section 4 reports some results from simulation studies conducted for evaluating the proposed methods. An application to a clinic study on CGD is provided in Section 5, and some concluding remarks are given in Section 6. All proofs are given in the Appendix.

2. Model and estimation procedure

2.1. The model

Consider a longitudinal study that involves n independent subjects, each of which experiences recurrences of the same event (Huang and Chen, 2003; Luo and Huang, 2011). For subject i, let T_{ij} denote the time from the $(j-1)$th to the jth occurrence of the event. That is, $T_{i1}+\cdots+T_{ij}$ is the jth recurrent event time. Also let Z_i denote the p-dimensional vector of covariates associated with subject i, and C_i the follow-up or censoring time. Let $N_i=\{T_{ij}:j=1,2,\ldots\}$. Assume that $\{N_i,C_i,Z_i\}$ $(i=1,\ldots,n)$ are independent and identically distributed (i.i.d.), and N_i is independent of C_i given Z_i. Define M_i to be the index of observed gap times for subject i, which satisfies

$$\sum_{j=1}^{M_i-1} T_{ij}\le C_i \quad \text{and} \quad \sum_{j=1}^{M_i} T_{ij}>C_i,$$

where $\sum_{j=1}^{0}\cdot\equiv 0$. Then observed data are $\{T_{i1},\ldots,T_{i,M_i-1},C_i,Z_i\}$. That is, the first M_i-1 gap times are observed, but T_{i,M_i} is censored at $T^{+}_{i,M_i}=C_i-\sum_{j=1}^{M_i-1}T_{ij}$.

Following Huang and Chen (2003), we assume that each individual recurrent event process is a renewal process, which implies that for a given i, $\{T_{ij},\ j=1,2,\ldots\}$ are i.i.d., and that for given (C_i,M_i,T^{+}_{i,M_i}), the observed complete gap times $\{T_{ij},\ j=1,\ldots,M_i-1\}$ are identically distributed (Wang and Chang, 1999).

Let $\lambda_{ij}(t|Z_i)$ be the hazard function of T_{ij} given Z_i. The proposed transformed hazards models take the form

$$\lambda_{ij}(t|Z_i)=H\{\lambda_0(t)+\beta_0'Z_i\}, \tag{1}$$

where $\lambda_0(t)$ is an unknown function, β_0 is a $p\times 1$ vector of unknown regression parameters, and $H(\cdot)$ is pre-specified and assumed to be twice continuously differentiable and strictly increasing. Model (1) defines a very rich family of models through the link function $H(\cdot)$, which includes the proportional hazards model $(H(x)=\exp(x))$ and the additive hazards models $(H(x)=x)$. One example of $H(\cdot)$ is the Box–Cox transformation, in which $H(\cdot)$ is given by $H(x)=\{(1+x)^s-1\}/s$ for $s\ge 0$ with $s=0$ corresponding to $H(x)=\log(x+1)$. Another useful class is the logarithmic transformations, which are given by $H(x)=\log(1+\gamma x)/\gamma$ for $\gamma\ge 0$ with $\gamma=0$ corresponding to $H(x)=x$.

2.2. Inference procedure

Our inference procedure is based on the establishment of a connection between a subset of the observed gap times and clustered survival data. Let $\Delta_i=I(M_i>1)$, $S_i=\max(M_i-1,1)$, and

$$X_{ij}=\begin{cases} T_{ij} & \text{if } \Delta_i=1,\\ T^{+}_{ij} & \text{if } \Delta_i=0,\end{cases} \quad j=1,\ldots,S_i.$$

Then $\{X_{ij},\Delta_i,Z_i,\ j=1,\ldots,S_i\}$ $(i=1,\ldots,n)$ can be treated as clustered survival data. Since the cluster size is informative, the censored gap time needs to be removed for $M_i>1$ (Wang and Chang, 1999; Huang and Chen, 2003).

Define $N_{ij}(t)=\Delta_i I(X_{ij}\le t),\ Y_{ij}=I(X_{ij}\ge t)$, and

$$M_{ij}(t;\beta,\lambda)=N_{ij}(t)-\int_0^t Y_{ij}(u)H\{\lambda(u)+\beta'Z_i\}du.$$