Computational Statistics and Data Analysis 112 (2017) 154–169

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

The generalized moment estimation of the additive–multiplicative hazard model with auxiliary survival information

Wenpeng Shang [a], Xiao Wang [b,*]

[a] School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan 430074, China

[b] School of Mathematics and Statistics, Henan University, Henan 475004, China

ARTICLE INFO

Article history:
Received 18 June 2016
Received in revised form 21 February 2017
Accepted 9 March 2017
Available online 18 March 2017

Keywords:
Additive–multiplicative hazard
Generalized moment
Loss function
Parametric estimation
Subgroup information
Survival analysis

ABSTRACT

Additive–multiplicative hazard model is a natural extension of the proportional hazard model and the additive hazard model in survival analysis. It is classical for applying the martingale estimating functions to estimate the regression parameters. However, the generalized moment method is employed to estimate the coefficients via synthesizing the auxiliary subgroup survival information. The estimators are established to be consistent and asymptotically normal. Furthermore, the method is more efficient than the famous martingale approach. In particular, these asymptotic variance–covariances are identical as the number of subgroups is equal to one. The large sample property of the Breslow estimator for the baseline cumulative hazard function is also investigated. Some extensive simulation studies are conducted to evaluate the finite-sample performances of the proposed method. A real data study is analyzed to show its practical utility.

1. Introduction

There are a number of models in biomedical studies. The semiparametric regression models focused on the hazard function are usually used. In particular, the multiplicative risk model and the additive hazard model are mainly two types of risk models. Cox (1972) proposed the most famous proportional hazard model

$$\lambda(t|\boldsymbol{X}) = \lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}},$$

where $\lambda_0(t)$ is an unknown baseline risk function, $\boldsymbol{\beta}_0$ is a p-vector of unknown regression parameters and $\boldsymbol{X}$ is a p-vector of covariates. Let $\boldsymbol{\beta}_0 = (\boldsymbol{\beta}_{01}, \ldots, \boldsymbol{\beta}_{0p})^{\mathrm{T}}$ and $\boldsymbol{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$, then $\boldsymbol{\beta}_{0j}$ can be interpreted as the logarithm of the ratio of hazard for one unit change in X_j while holding the other components in $\boldsymbol{X}$ the same. The partial likelihood method has been deduced by Cox (1975) to estimate the parameters. However, when the true relationship of the baseline risk function and covariate effects is additive, a plausible alternative would be the additive hazard model,

$$\lambda(t|\boldsymbol{Z}) = \lambda_0(t) + \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z},$$

where $\boldsymbol{\theta}_0$ is a q-vector of unknown regression coefficients and $\boldsymbol{Z}$ is a q-vector of covariates. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}, \ldots, \boldsymbol{\theta}_{0q})^{\mathrm{T}}$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_q)^{\mathrm{T}}$, then $\boldsymbol{\theta}_{0k}$ can be interpreted as the direct risk difference for one unit change in Z_k while holding the

* Corresponding author.
E-mail addresses: swp_07tj@163.com (W. Shang), xiaoheda06@163.com (X. Wang).

http://dx.doi.org/10.1016/j.csda.2017.03.013

other components in $\boldsymbol{Z}$ the same. The related models were considered by Aalen (1980), Breslow and Day (1980), Cox and Oakes (1984), Buckley (1984), Thomas (1986), Breslow and Day (1987), Aalen (1989), Huffer and Mckeague (1991), Andersen et al. (1993), McKeague and Sasieni (1994) and so on. Lin and Ying (1994) subsequently presented a celebrated martingale based inference method for the regression parameters through the techniques analogous to partial likelihood method. However, the additive and multiplicative effects could exist for the different covariates. Lin and Ying (1995) expanded the Cox model and the additive risk model to the general additive–multiplicative hazard model. In this article, we consider the additive–multiplicative hazard model

$$\lambda(t|\boldsymbol{X},\boldsymbol{Z}) = \lambda_0(t)e^{\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{X}} + \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z},$$

which is a particular case of the general additive–multiplicative hazard model with the completely different covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$. The model allows certain covariate effects to be additive while the others to be multiplicative. When $\boldsymbol{\beta}_0$ is equal to zero, it degenerates to the additive hazard model. When $\boldsymbol{\theta}_0$ is equal to zero, it is the Cox model. Thus, the additive–multiplicative hazard model is a very rich model which contains the proportional risk model and additive hazard model as its special cases. Lin and Ying (1995) simultaneously proposed a series of martingale methods to estimate the regression parameters and studied the asymptotic properties of their estimators.

In the survival data, some useful auxiliary survival information could be included in the subgroup data associated with the division of covariates. If we directly adopt the approach of Lin and Ying (1995) to estimate regression coefficients without combining the auxiliary information from the subgroups, the efficiency of these estimators is decreased and some crucial information would be lost. There are some references for synthesizing the auxiliary covariate information to improve the efficiency of estimation. For example, Wu and Sitter (2001) applied the empirical likelihood approach to combine the auxiliary covariate information to improve the efficiency of estimation. Huang et al. (2015) proposed the double empirical likelihood method to synthesize the auxiliary information to estimate the regression parameters in the proportional hazard model. The auxiliary survival information could correspond to some estimating functions in the paper. When the number of estimating equations is greater than $p + q$, the generalized moment method is a very useful tool for estimating parameters. Since it has many desirable statistical properties, many authors carried out much research. Hansen (1982) considered the large sample properties of the generalized moment estimation. The finite sample properties were studied by Hansen et al. (1996). For a comprehensive coverage of the generalized method of moments, see Hall (2005). In vector autoregressive models, Park et al. (2011) used the generalized method of moments to estimate cointegration and Hayakawa (2016) improved the estimators of generalized moment method. Li and Yin (2009) employed the generalized method of moments in the accelerated failure time model with correlated survival data.

In the article, we synthesize the auxiliary subgroup survival information and use the generalized moment method to estimate the regression coefficients of the additive–multiplicative risk model. This approach can automatically combine the vitally auxiliary subgroup information. The consistency and asymptotic normality of the proposed estimates are shown. Moreover, the asymptotic variance–covariances of our estimators could be directly estimated. We also establish that our method is more effective than the approach of Lin and Ying (1995). When the number of subgroups is equal to one, these efficiencies are identical. In order to demonstrate the usefulness of utilizing the auxiliary information, the generalized moment method without auxiliary subgroup information is also proposed and its asymptotic properties are established. It is the same as Lin and Ying (1995) for the efficiency of the generalized moment method without auxiliary information. For the baseline cumulative hazard function, we use the elegant Breslow estimator, which was suggested by Breslow (1972, 1974). The weak convergence of the Breslow estimator is also listed.

The article structure is organized as follows. Some estimating functions and the Breslow estimator for the baseline cumulative hazard function are demonstrated in Section 2. We introduce the generalized moment method with and without auxiliary survival information to estimate the regression parameters in Section 3. In Section 4, the consistency and asymptotic normality of these parametric estimators are illustrated with all technical proofs deferred to the Appendix. In addition, we describe the weak convergence of the Breslow estimator. Some simulation studies are conducted in Section 5 to evaluate the finite sample performance of our approach. In Section 6, the proposed method is applied to a real data study. Some discussions are given in Section 7.

2. Model setup

In this section, we describe some estimating functions and the Breslow estimator for the proposed model, beginning with some notation.

2.1. Notation

Let T denote the failure time. Assume that T is generated from the additive–multiplicative hazard model. Due to the study end, the observed time is frequently subject to right censoring. Let C indicate the censoring time and $\tilde{T}$ denote the observed time. Thus $\tilde{T} = T \wedge C$, where $a \wedge b$ means the minimum of a and b. Assume that the censoring mechanism is random, that is, the survival time T and the censoring time C are conditionally independent given the covariates $\boldsymbol{X}$ and $\boldsymbol{Z}$. Let $\Delta = I(T \leq C)$