Journal of the Korean Statistical Society 44 (2015) 619–631

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Empirical likelihood for the class of single index hazard regression models[☆]

Jianbo Li[a,*], Minglong Guo[a], Changxin Du[a], Riquan Zhang[b], Heng Lian[c]

[a] *School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, 221116, PR China*
[b] *School of Finance and Statistics, East China Normal University, Shanghai, 200241, PR China*
[c] *School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore*

ARTICLE INFO

Article history:
Received 22 October 2014
Accepted 2 April 2015
Available online 27 April 2015

AMS 2000 subject classifications:
primary 62M10
secondary 62H12

Keywords:
Single index hazard models
Empirical likelihood
B spline
Partial likelihood

ABSTRACT

Based on the B spline approximation technique and right censored data, we consider the empirical likelihood inference for the index parameters and its partial components in a class of single index hazard regression models. Under some regular conditions, we show that our proposed empirical likelihood ratio statistics follow the standard χ^2 distribution. Some numerical studies are given to illustrate our proposed methodology.

1. Introduction

Proportional hazard Cox's regression model (Cox, 1972) is one of most popular and important tools in describing the relationship between the failure time and its corresponding covariates. It is classically formulated as

$$\lambda(t|X) = \lambda_0(t)\exp\{X'\beta\}, \tag{1}$$

where t is a realization of the failure time $T \in R_+, X \in R^p$ is the corresponding covariate vector, $\lambda(t|X)$ is the conditional hazard function given the covariate X, $\lambda_0(t)$ is the baseline hazard function, $\beta \in R^p$ is the regression parameter vector and X' stands for the transpose of X. Cox (1975) proposed a corresponding partial likelihood estimation method; Andersen and Gill (1982), Andersen, Borgan, Gill, and Keiding (1993), Cox and Oakes (1984), Fan and Gijbels (1996) and others proposed some parametric and nonparametric algorithms for the estimation of β.

Because of the linear effect of X on the logarithm of hazard function $\lambda(t|X)$, which may be unrealistic in many applications, many relaxed models have been proposed, for example, the completely nonparametric hazard model (Fan, Gijbels, & King, 1997; Gentleman & Crowley, 1991; Gu, 1996; O'Sullivan, 1993; Tibshirani & Hastie, 1987), additive proportional hazard model (Sleeper & Harrington, 1990), varying coefficient hazard model (Cai & Sun, 2003; Gray, 1992; Marzec & Marzec, 1997;

☆ This work is partially supported by Qing Lan Project in Jiangsu Province, PRC, Humanities and Social Fund of Ministry of Education in China (12YJC910004), National Natural Science Foundation of China (11201190, 11171112) and the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

* Correspondence to: NO. 101, Shanghai Road, Xuzhou, Jiangsu, 221116, PR China.
E-mail address: ljianb66@gmail.com (J. Li).

http://dx.doi.org/10.1016/j.jkss.2015.04.002
1226-3192/

Murphy, 1993; Murphy & Sen, 1991; Sasieni & Winnett, 2003; Tian, Zucker, & Wei, 2005; Verweij & van Houwelingen, 1995; Zucker & Karr, 1990), functional ANOVA hazard model (Huang, Kooperberg, Stone, & Truong, 2000), partially linear additive hazard model (GØrgens, 2004; Huang, 1999), and single index hazard model (Huang & Liu, 2006; Lu, Chen, Song, & Singh, 2006; Wang, 2004). By extending the following single index hazard model (SIHM) in Huang and Liu (2006)

$$\lambda(t|X) = \lambda_0(t)\exp\{\psi(X'\beta)\}. \tag{2}$$

Li and Zhang (2011) and Sun, Kopciuk, and Lu (2008) respectively proposed a partially linear single index hazard model (PLSIHM) and a partially varying coefficient single index hazard model (PVCSIHM) as follows:

$$\lambda(t|X,W) = \lambda_0(t)\exp\{\psi(X'\beta) + W'\alpha\} \tag{3}$$

and

$$\lambda(t|X,W,U) = \lambda_0(t)\exp\{\psi(X'\beta) + W'\alpha(U)\}, \tag{4}$$

where $\psi(\cdot)$ is a completely unknown univariate function with $\psi(0) = 0$ for model identification, classically termed by index function, $W \in R^q$ is the covariate vector in linear or varying coefficient part, $\alpha \in R^q$ is the regression parameter vector in linear part, $\alpha(u) = (\alpha_1(u), \alpha_2(u), \ldots, \alpha_q(u))'$ is the functional coefficient vector in the varying coefficient part, $U \in R$ is a univariate random variable, which can be any variable such as anyone entry of X or W, and $\|\beta\| = 1$ with last nonzero element being positive for model identification. Based on the B spline approximation of unknown functions in models (2)–(4), they proposed a two-step estimation algorithm by incorporating maximum partial likelihood approach. Shang et al. (2013) considered the nested case-control studies through the partially linear single index hazard model (3). The class of single index hazard models are very flexible in modeling failure time data and include many popular semiparametric hazard regression models mentioned above as its special cases. More detailed discussion can be seen in Huang and Liu (2006), Li and Zhang (2011), Sun et al. (2008) and the others.

In the statistical analysis of models (2)–(4), the authors mainly focused on the model estimation and less on the construction of confidence region and significance testing. Classically, we can deal with such problems using the asymptotic normality theory. However, the asymptotic covariance–variance matrix usually has no closed form and must be approximated using some numerical methods. So the confidence region and significance testing methods based on the asymptotic normality may suffer from low efficiency. Owen (1988, 1990) developed an empirical likelihood method for constructing the joint confidence region. Based on estimating equation, Hjort, McKeague, and Van Keilegom (2009) extended the scope of empirical likelihood in Owen (1990) and proposed a general plug-in empirical likelihood inference frame for a large range of models. Chen and Van Keilegom (2009) further reviewed the empirical likelihood inference over a wide range of parametric, semiparametric and nonparametric regression models. One advantage of the empirical likelihood methodology is that it is free of covariance–variance matrix estimation and thus significantly improve coverage accuracy. The empirical likelihood approach has been studied in the context of proportional hazard Cox's model and its extensions. For instance, Qin and Jing (2001) considered the empirical likelihood inference for the proportional hazard Cox's regression model; Sun, Sundaram, and Zhao (2009) studied the empirical likelihood pointwise confidence regions for the time-dependent regression coefficients via local partial likelihood. Ren and Zhou (2011) proposed a profile empirical likelihood inference for Cox's model based on the full likelihood; Zhao and Jinnah (2012) studied a profile empirical likelihood inference for the regression parameters in Cox's model via an adjusted methods. Tang, Li, and Lian (2013) considered the empirical likelihood inference for partially linear proportional hazards models with growing dimensions.

To the best of our knowledge, we did not find the study on the empirical likelihood inference for the single index hazard models (2)–(4). In this paper, based on the partial likelihood function and B spline approximation technique in Huang and Liu (2006), we mainly consider the empirical likelihood inference for the index parameters and its partial components in models (2). Under some regular conditions, we show that the proposed empirical likelihood ratio statistics asymptotically follow the standard χ^2 distribution. We also extend the proposed empirical likelihood methodology to the context of partially linear single index hazard model (3) and partially varying coefficient single index hazard model (4). Some numerical studies will be given to illustrate our proposed approach.

The rest of this paper will be organized as follows. In Section 2, we describe the empirical likelihood inference procedure for the full index parameters in model (2), including B spline approximation, approximated partial likelihood function, empirical likelihood ratio statistic and the corresponding asymptotic properties. We further study the empirical likelihood for the partial components of index parameters and its limit distribution in Section 3. We extend the proposed methodology to the context of models (3)–(4) in Section 4; In Section 5, we present some numerical studies to illustrate our proposed empirical likelihood approach. We make some discussion in Section 6; Some technique proof will be given in the Appendix.

2. EL for full parameters

In this section, we present the empirical likelihood inference methodology for the full index parameters in the single index hazard model (2) with reparameterization technique.