Journal of the Korean Statistical Society 43 (2014) 177–187

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Testing for the parametric component of partially linear EV models under random censorship

Linjun Tang *, Changchun Wu, Zhangong Zhou

Department of Statistics, Jiaxing University, Jiaxing, 314001, China

A R T I C L E I N F O

Article history:
Received 16 December 2012
Accepted 5 July 2013
Available online 6 August 2013

AMS 2000 subject classifications:
62N01
62N03

Keywords:
Partially linear model
Random censorship
Errors-in-variables
Hypothesis test

A B S T R A C T

This paper investigates the hypothesis test of the parametric component in partially linear errors-in-variables (EV) model with random censorship. We construct two test statistics based on the difference of the corrected residual sum of squares and empirical likelihood ratio under the null and alternative hypotheses. It is shown that the limiting distributions of the proposed test statistics are both weighted sum of independent standard chi-squared distribution with one degree of freedom under the null hypothesis. Based on the adjusted test statistics, we further develop two new types of test procedures. Finite sample performance of the proposed test procedures is evaluated by extensive simulation studies.

1. Introduction

Over the past decades, the semiparametric errors-in-variables (EV) model has frequently been applied in practice and has received much attention in the literature, see Carroll, Ruppert, Stefanski, and Crainiceanu (2006). For a long time, however, its research hotspots were mainly centered around parameter estimation. Based on the least squares approach, Liang, Hardle, and Carroll (1999), and Wang (1999) considered the estimation problem of the partially linear EV model. By using the empirical likelihood method, Liang, Wang, and Carroll (2007) provided the empirical likelihood confidence region for the parameter for the semiparametric EV model. Moreover, some researchers further considered the issue of statistical inference for partially linear varying coefficient EV models. For example, You and Chen (2006) introduced the modified profile least squares method to estimate the parametric component of partially linear varying coefficient EV models; Li and Greene (2008) proposed a semiparametric partially varying coefficient model and applied locally corrected score equations to estimate parameters and coefficient functions. Moreover, Hu, Wang, and Zhao (2009) and Wang, Li, and Lin (2011) also obtained the corresponding confidence region for partially linear varying coefficient EV models by the empirical likelihood method. Liang and Wang (2005) considered the statistical inference issue of the parametric component in the partially linear single-index EV model. The comprehensive reviews on the research and development of the semiparametric EV model can be also found in Fan and Huang (2005), and Liang, Su, and Thurston (2009).

In practice, especially in survival analysis, the response variable may not be usually completely observed due to random censoring. In this paper, we consider the following partially linear EV model with randomly right-censored data:

$$\begin{cases} Y = x^{\tau}\beta + g(T) + \varepsilon \\ X = x + u, \end{cases} \tag{1}$$

* Corresponding author. Tel.: +86 15888337096.
E-mail address: tljlqz@163.com (L. Tang).

1226-3192/$ – see front matter © 2013 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jkss.2013.07.001

where (X, T) is $R^p \times [0, 1]$ observable random vector, x is R^p unobservable random vector, β is a $p \times 1$ vector of unknown parameters, $g(\cdot)$ is an unknown function on $[0, 1]$, and ε is random error of regression model. In this model, the response Y is censored randomly on the right by a censoring variable C and hence cannot be completely observed. Further, we assume that model error ε, measurement error u, and random vector (x, T) are independent mutually and subject to

$$\mathrm{E}(\varepsilon) = \mathbf{0}, \qquad \mathrm{Var}(\varepsilon) = \sigma^2, \qquad \mathrm{E}(u) = \mathbf{0}, \qquad \mathrm{Cov}(u) = \Sigma_u > \mathbf{0},$$

where Σ_u is known, otherwise can be estimated by repeatedly measuring X. For model (1), Liu, Xue, and Chen (2009) constructed the confidence region of parametric component through empirical likelihood method. Moreover, Liu (2011) investigated the linear EV model with right random censoring, and proposed corrected least squares and maximum empirical likelihood confidence region.

In practical situation, some prior information about the unknown coefficient of regression model might be available from outside sample sources. Use of such information can greatly improve upon the efficiency of regression analysis. The aforementioned studies mainly focused on the estimation of parametric component rather than test, see Przystalski and Krajewski (2007), and Shalabh and Misra (2007). However, once these prior constraints have been imposed in regression model, the first task is to test whether they meet a prior constraint conditions. The problem of hypothesis test in model (1) is very important but under-studied. In this paper, our interest is to test the following linear hypothesis:

$$H_0 : A\beta = b \quad \text{v.s.} \quad H_1 : A\beta \neq b \tag{2}$$

where A is a known $k \times p$ constant matrix with $\mathrm{rank}(A) = k\,(k \leq p)$, and b is a known $k \times 1$ constant vector.

There have been a few studies on the statistical inference for restricted partially linear EV model without any censorship. For example, Wei (2012) considered restricted partially linear varying coefficient errors-in-variables models, and with Wang (2012) considered the statistical inference on restricted partially linear additive errors-in-variables models, respectively. As discussed in Wei (2012) and Wei and Wang (2012), the generalized likelihood ratio (GLR) test proposed by Fan and Huang (2005) cannot be directly applied to test the hypothesis of EV model. They developed an applicable test procedure based on corrected residual sum of squares method under the null and alternative hypotheses. Inspired by this, we propose two new test procedures for the hypothesis (2) in model (1) based on the corrected residual sum of squares and empirical likelihood ratio methods, and establish their asymptotic properties.

The rest of this article is organized as follows. In Section 2, two test statistics are proposed based on the corrected least squares and maximum empirical likelihood methods. Some simulations are conducted in Section 3 to illustrate the sample performance of the proposed procedures. The proofs of main theorems are presented in Section 4. Section 5 concludes the article with a discussion.

2. Proposed methods and main results

2.1. Model

Let $\{(Y_i, X_i, T_i)\}_{i=1}^n$ be a random sample generated from model (1), which is subject to

$$Y_i = x_i^\tau \beta + g(T_i) + \varepsilon_i, \qquad X_i = x_i + u_i. \tag{3}$$

If Y_i is randomly right-censored, it cannot be completely observed. Instead, one can only observe $\{Z_i, X_i, T_i, \delta_i\}$, where

$$Z_i = \min\big(Y_i, C_i\big), \qquad \delta_i = I(Y_i \leq C_i),$$

and $C_1, \ldots, C_n$ are random sample generated from distribution function $G(\cdot)$. When the response exists censoring, some classical statistical methods such as least squares and empirical likelihood ones, cannot be directly used to the inference of model (3). To overcome this problem, we introduce the following transformation for observable Z_i

$$Y_{iG} = \frac{Z_i \delta_i}{1 - G(Z_i-)}, \quad i = 1, 2, \ldots, n.$$

It is easily verified that $\mathrm{E}(Y_i|x_i, T_i) = \mathrm{E}(Y_{iG}|x_i, T_i) = x_i\beta + g(T_i)$. In practical situations, the distribution function $G(\cdot)$ is usually unknown. In this paper, we suggest to replace $G(\cdot)$ by its Kaplan and Meier (1958) estimator

$$G_n(z) = 1 - \prod_{i=1}^{n} \left[\frac{N^+(Z_i)}{1 + N^+(Z_i)}\right]^{I(Z_i \leq z,\, \delta_i = 0)}, \quad z > 0,$$

where $N^+(z) = \sum_{i=1}^n I\,(Z_i > z)$. Without measurement errors and random censorship, the partially linear model (3) can be written as

$$Y_i - \mathrm{E}(Y_i|T_i) = \big[x_i^\tau - \mathrm{E}(x_i^\tau|T_i)\big]\,\beta + \varepsilon_i. \tag{4}$$