Statistics and Probability Letters 105 (2015) 149–156

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Threshold effect test in censored quantile regression

Yanlin Tang [a], Xinyuan Song [b,c], Zhongyi Zhu [d,*]

[a] Department of Mathematics, Tongji University, Shanghai, China
[b] Shenzhen Research Institute, The Chinese University of Hong Kong, Guangdong, China
[c] Department of Statistics, The Chinese University of Hong Kong, Hong Kong
[d] Department of Statistics, Fudan University, Shanghai, China

ARTICLE INFO

Article history:
Received 2 February 2015
Received in revised form 6 May 2015
Accepted 31 May 2015
Available online 12 June 2015

Keywords:
Covariate-threshold
Hypothesis testing
Inverse-censoring-probability-weighted quantile regression
Random censoring

ABSTRACT

We propose a new test for covariate-threshold caused change point in quantile regression with random censoring, based on partial subgradient. Critical values are obtained using wild bootstrap samples, where induced smoothing method is used to estimate the conditional density.

1. Introduction

Threshold regression models have wide applications in biostatistics, actuarial science, economics, etc. In threshold regression models, the coefficients are piecewise functions of a "threshold variable", with jumps occurring at some unknown change points. Threshold regression models are often used to address the problem when the regression functions have different forms in different regions of the covariates. See for example, Kerry et al. (2005) used threshold regression models to analyze relationships of the blood pressure and body mass index (BMI), since different relationships are detected for those below and above a BMI threshold; Ghosh and Vaida (2007) used threshold regression models to analyze the CD4 counts, since the counts increased rapidly in the first 2–4 weeks but decreased by degrees afterwards. See more examples in Huang (2009) and Pennell et al. (2010). The existing theoretical studies for threshold regression models mostly focused on mean regression models (Feder, 1975; Chappell, 1989; Tong, 1990; Hansen, 1999; Cho and White, 2007; David and Nicholas, 2014; Zou et al., 2014).

In this paper, we focus on quantile regressions (Koenker and Bassett, 1978; Koenker, 2005) with random censoring. Several authors have studied the testing and estimation of structural changes in quantile regression for time series, where the model structure changes after an unknown time point, see Su and Xiao (2008), Qu (2008) and Oka and Qu (2011) for examples. However, the problems discussed in these literature are different from threshold regression models, where the change point is caused by an unknown threshold in the continuous threshold covariate.

Recently, several authors discussed testing and estimation for threshold quantile regression models, see for examples, Lee et al. (2011), Li et al. (2011), Galvao et al. (2011), Galvao et al. (2013), and Zhang et al. (2014), etc. Lee et al. (2011) proposed

* Corresponding author.
E-mail address: zhuzy@fudan.edu.cn (Z. Zhu).

http://dx.doi.org/10.1016/j.spl.2015.05.019

a sup-likelihood-ratio-based method for testing the existence of threshold effects in regression models; Li et al. (2011) discussed estimation for a bent line quantile regression; Galvao et al. (2011) considered estimation in threshold quantile autoregressive models; Galvao et al. (2013) considered a uniform test of linearity against threshold effects in quantile regression model, where the focus is on the stationary time series processes; Zhang et al. (2014) proposed hypothesis testing for threshold quantile regression, based on CUSUM process of subgradient obtained by fitting the quantile regression model under the null hypothesis. All these work are for standard quantile regression, where no censoring is in presence in the responses.

In many applications, especially in survival analysis, the responses, T, cannot be completely observed due to possible censoring. In this paper, we focus on random right censoring. We propose a new hypothesis testing procedure for change points due to covariate-threshold, i.e., testing whether the regression functions is different in sub samples segmented by a threshold variable. To the best of our knowledge, this is the first attempt for quantile regression with random censoring. Our proposed test statistic is based on the partial subgradient of the inverse probability weighted quantile regression objective function, requiring fitting the model only under the null hypothesis. The critical values are obtained by wild bootstrap method, simulating the Gaussian process which characterizes the limiting distribution of the test statistic, and induced smoothing method is used to estimate the conditional density function of the regression errors. The test in Lee et al. (2011) was based on the difference between the maximum values of the objective functions under the null and alternative hypotheses, which was shown less powerful numerically than the score based test in Zhang et al. (2014). Under some regularity conditions, we present the limiting distribution of the test statistic under both null hypothesis and local alternative hypotheses. Simulation study suggests that the proposed test performs reasonably well.

The rest of the paper is organized as follows. In Section 2, we present the proposed hypothesis testing method and the generating of critical values. In Section 3, we present the theoretical properties of the testing statistic. In Section 4, we assess the finite sample performance of the proposed method through a simulation study. In Section 5, we apply the test to a real data. All the theoretical proofs are deferred to the Appendix A.

2. Proposed test statistic

At a given quantile level $0 < \tau < 1$, we consider the following threshold quantile regression model for the uncensored latent dependent variable of interest

$$T_i = \mathbf{x}_i^T \boldsymbol{\beta}_{0\tau} + \mathbf{z}_i^T \boldsymbol{\gamma}_{0\tau} I(U_i \geq u_0) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where U_i is the threshold variable, $\mathbf{x}_i$ is the p-dimensional design vector (may include U_i as one component), $\mathbf{z}_i$ is the q-dimensional design vector, u_0 is the unknown threshold parameter, and the conditional τth quantile of ϵ_i given $(\mathbf{x}_i, \mathbf{z}_i, U_i)$ is zero. Thus the conditional τth quantile of T_i given $(\mathbf{x}_i, \mathbf{z}_i, U_i)$ is

$$Q_{T_i}(\tau|\mathbf{x}_i, \mathbf{z}_i, U_i) = \mathbf{x}_i^T \boldsymbol{\beta}_{0\tau} + \mathbf{z}_i^T \boldsymbol{\gamma}_{0\tau} I(U_i \geq u_0), \quad i = 1, \ldots, n.$$

Without loss of generality, we assume that U_i and $\mathbf{z}_i$ are subsets of $\mathbf{x}_i$, and in the following we often omit the condition $\mathbf{z}_i$ since it is included in $\mathbf{x}_i$. We assume that the first element of $\mathbf{x}_i$ is 1 corresponding to the intercept. In addition, without loss of generality, we assume that U_i have a finite support on [0,1].

Let C_i denote the censoring variable. Due to censoring, we only observe $(\mathbf{x}_i, U_i, Y_i, \delta_i)$, where

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

are the observed (possibly censored) response variable and the censoring indicator, respectively. What we are interested in is testing the existence of a threshold effect at the quantile level τ, i.e., the null hypothesis and alternative hypotheses are

$$H_0 : \boldsymbol{\gamma}_{0\tau} = 0, \quad \text{for all } u_0 \in (0, 1) \quad \text{versus} \quad H_1 : \boldsymbol{\gamma}_{0\tau} \neq 0, \quad \text{for some } u_0 \in (0, 1). \tag{2}$$

Under H_0, assuming that C_i is independent from the covariates, Bang and Tsiatis (2002) proposed the inverse-censoring-probability-weighted (ICPW) estimator, which solves

$$U_n(\boldsymbol{\beta}, \widehat{G}) = n^{-1} \sum_{i=1}^{n} \frac{\mathbf{x}_i \delta_i}{\widehat{G}(Y_i)} \{\tau - I(Y_i - \mathbf{x}_i^T \boldsymbol{\beta} < 0)\} = 0, \tag{3}$$

where $\widehat{G}(t) = \widehat{P(C \geq t)}$ is the Kaplan–Meier estimator of $G_0(t)$, the survival function of the censoring time. More specifically, $\widehat{G}(t)$ is estimated by

$$\widehat{G}(t) = \prod_{k=1}^{n} \left\{ 1 - \frac{1}{\sum_{s=1}^{n} I(Y_s \geq Y_k)} \right\}^{I(Y_k \leq t, \delta_k = 0)}.$$

The estimator obtained by solving (3) is denoted as $\widehat{\boldsymbol{\beta}}_\tau$. Note that solving (3) is equivalent to minimize

$$L_n(\boldsymbol{\beta}, \widehat{G}) = n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{G}(Y_i)} \rho_\tau(Y_i - \mathbf{x}_i^T \boldsymbol{\beta}).$$