Statistics and Probability Letters 100 (2015) 98–103

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Sequential change point detection in linear quantile regression models

Mi Zhou[a], Huixia Judy Wang[b,*], Yanlin Tang[c]

[a] Fannie Mae, 4000 Wisconsin Ave NW, Washington, DC, 20016, USA
[b] Department of Statistics, George Washington University, Washington, DC, 20052, USA
[c] Department of Mathematics, Tongji University, Shanghai, 200092, China

ARTICLE INFO

Article history:
Received 17 March 2014
Received in revised form 12 December 2014
Accepted 24 January 2015
Available online 14 February 2015

Keywords:
Change point detection
Linear regression
Quantile regression
Sequential testing
Structural change

ABSTRACT

We develop a method for sequential detection of structural changes in linear quantile regression models. We establish the asymptotic properties of the proposed test statistic, and demonstrate the advantages of the proposed method over existing tests through simulation.

1. Introduction

For applications where the relationship between the response and covariates has a structural change at a certain point, the change may occur at the tails of the response distribution but not at the center. The conventional mean regression method for change point detection cannot be used to identify such structural changes at tails or may be lack of power in distributions with heavy tails. To provide a more robust testing procedure and obtain a more comprehensive view of structural changes, we focus on linear quantile regression models.

Several researchers have studied change point detection and estimation for quantile regression models, for instance, Bai (1996), Su and Xiao (2008), Qu (2008), Oka and Qu (2011), Furno (2007) and Furno (2012), to name a few. These work for quantile regression all focused on detecting changes in observations within a fixed length in a retrospective way. These retrospective quantile methods cannot be applied to the sequential data where new data arrive steadily, because the replication of such tests yields a procedure that rejects a true null hypothesis of no change with probability approaching one; see Robbins (1970). To our knowledge there exists little work for sequential change point detection in quantile regression models. Among some related work, Koubková (2008) proposed a L_1-based monitoring procedure in linear regression models, and Chochola et al. (2013) discussed change point monitoring based on M-estimators. We develop a new procedure for sequentially monitoring structural changes in linear quantile regression models.

Let y be the response variable and $\mathbf{x}$ be the p-dimensional covariate vector with the first element 1. Denote $Q_y(\tau|\mathbf{x}) = \inf\{y : F_Y(y|\mathbf{x}) \geq \tau\}$ as the τth conditional quantile of y given $\mathbf{x}$, where $F_y(\cdot|\mathbf{x})$ is the conditional distribution of y given $\mathbf{x}$.

* Corresponding author.
E-mail address: judywang@gwu.edu (H.J. Wang).

http://dx.doi.org/10.1016/j.spl.2015.01.031
0167-7152/

Let $\{y_i, \mathbf{x}_i, i \geq 1\}$ denote a sample with i representing a time index or some other ordering. We consider the linear quantile regression model

$$Q_y(\tau|\mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}_{i,\tau}, \quad i \geq 1, \tag{1}$$

where $\boldsymbol{\beta}_{i,\tau}$ is the p-dimensional unknown quantile coefficient vector. We are interested in monitoring the changes of the effects of $\mathbf{x}$ on the quantiles of y over time, that is, monitoring the consistency of $\boldsymbol{\beta}_{i,\tau}$ over i.

In Section 2, we present the proposed methods for monitoring change points at a single quantile level or across quantiles. The performance of the proposed methods is assessed through a simulation study in Section 3. The technical proofs are provided in the Supplementary Material (see Appendix A).

2. Proposed method

2.1. Sequential change point detection at a single quantile

We assume that there exists a historical data of size m such that $\boldsymbol{\beta}_{1,\tau} = \cdots = \boldsymbol{\beta}_{m,\tau} = \boldsymbol{\beta}_\tau^0$. This assumption was called the "noncontamination" assumption in Chu et al. (1996). The historical data is used for obtaining an estimate for the pre-change regression coefficient $\boldsymbol{\beta}_\tau^0$. At a given quantile level $\tau \in (0, 1)$, we are interested in monitoring the future incoming observations sequentially for a change in the regression coefficient, that is, testing the null hypothesis

$$H_0 : \boldsymbol{\beta}_{i,\tau} = \boldsymbol{\beta}_\tau^0, \quad \text{for } i \geq m + 1,$$

against the alternative hypothesis

$$H_1 : \boldsymbol{\beta}_{i,\tau} = \begin{cases} \boldsymbol{\beta}_\tau^0, & \text{for } m + 1 \leq i < m + k^* \\ \boldsymbol{\beta}_\tau^1, & \text{for } i \geq m + k^*, \end{cases}$$

where $k^* \geq 1$ is the unknown change point, and $\boldsymbol{\beta}_\tau^0 \neq \boldsymbol{\beta}_\tau^1$ are the unknown pre- and post-change coefficients.

Let $\hat{\boldsymbol{\beta}}_\tau$ be the quantile coefficient estimator of $\boldsymbol{\beta}_\tau^0$ based on the historical data, that is, $\hat{\boldsymbol{\beta}}_\tau = \text{argmin}_{\boldsymbol{\beta} \in \mathbb{R}^p} \sum_{i=1}^m \rho_\tau(y_i - \mathbf{x}_i^T \boldsymbol{\beta})$, where $\rho_\tau(u) = u\{\tau - I(u < 0)\}$ is the quantile loss function; see Koenker and Bassett (1978). The building block of our monitoring process is the following subgradient-based CUSUM-type process

$$S(m, k) = m^{-1/2} J_m^{-1/2} \sum_{i=m+1}^{m+k} \mathbf{x}_i \psi_\tau(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_\tau), \quad k = 1, \ldots, T_m,$$

where $J_m = \tau(1 - \tau) D_m$ with $D_m = m^{-1} \sum_{i=1}^m \mathbf{x}_i \mathbf{x}_i^T$, $\psi_\tau(u) = \tau - \mathrm{I}(u < 0)$ and T_m is the monitoring horizon. Our proposed test statistic is defined as

$$Q_\tau = \sup_{1 \leq k \leq T_m} \Gamma(m, k, \gamma), \quad \text{where } \Gamma(m, k, \gamma) = \left\| \frac{S(m, k)}{g(m, k, \gamma)} \right\|_\infty,$$

$g(m, k, \gamma) = (1 + k/m)\{k/(m + k)\}^\gamma$ is the normalizing function and $0 \leq \gamma < 1/2$. The tuning parameter γ controls how soon the monitoring will be stopped. The procedure with a larger value of γ will stop sooner and thus is preferred if the structural change occurs shortly after m. Throughout, we call a procedure open-end if the monitoring is continued possibly to infinity if no alarm is raised (that is, the monitoring horizon $T_m = \infty$), and closed-end if the monitoring is stopped after a fixed number of observations even if no change is detected (specifically $T_m/m \to N$ with $N > 0$); see Husková and Kirch (2012) and Kirch and Kamgaing (2014) for similar definitions.

We propose to stop the monitoring process and reject H_0 at the stopping time defined by

$$ST(m) = \begin{cases} \inf\{k \geq 1 : \Gamma(m, k, \gamma) \geq c_\alpha\} \\ \infty, \quad \text{if } \Gamma(m, k, \gamma) < c_\alpha \text{ for all } k = 1, \ldots, T_m, \end{cases}$$

where c_α is the critical value chosen to control the false alarm rate at a given significance level $\alpha \in (0, 1)$, that is, $\lim_{m\to\infty} P\{ST(m) < \infty | H_0\} = \alpha$.

We make the following technical conditions.

Assumption A1. $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \ldots$ are independent random pairs.

Assumption A2. The conditional density function of y given $\mathbf{x}_i$, denoted as $f_y(\cdot|\mathbf{x}_i)$, is continuous, uniformly bounded away from zero and infinity and has a bounded first derivative in the neighborhood of $\mathbf{x}_i^T \boldsymbol{\beta}_{i,\tau}$.

Assumption A3. Let $\|\cdot\|$ denote the Euclidean norm. The sequence $\{\mathbf{x}_i, 1 \leq i < \infty\}$ is strictly stationary satisfying the following conditions: