Journal of Multivariate Analysis 105 (2012) 368–379

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Self-consistent estimation of censored quantile regression

Limin Peng

Department of Biostatistics and Bioinformatics, Emory University, 1518 Clifton Rd. NE, Atlanta, 30322, USA

ARTICLE INFO

Article history:
Received 11 June 2011
Available online 21 October 2011

AMS subject classifications:
62N01
62H99

Keywords:
Censoring
Martingale
Regression quantile
Self-consistency
Stochastic integral equation

ABSTRACT

The principle of self-consistency has been employed to estimate regression quantile with randomly censored response. The asymptotic studies for this type of approach was established only recently, partly due to the complex forms of the current self-consistent estimators of censored regression quantiles. Of interest, how the self-consistent estimation of censored regression quantiles is connected to the alternative martingale-based approach still remains uncovered. In this paper, we propose a new formulation of self-consistent censored regression quantiles based on stochastic integral equations. The proposed representation of censored regression quantiles entails a clearly defined estimation procedure. More importantly, it greatly simplifies the theoretical investigations. We establish the large sample equivalence between the proposed self-consistent estimators and the existing estimator derived from martingale-based estimating equations. The connection between the new self-consistent estimation approach and the available self-consistent algorithms is also elaborated.

1. Introduction

Quantile regression [12] has arisen into a useful regression technique for survival data (i.e. time-to-event data). Compared to traditional survival regression methods, including the Cox proportional hazards model and the accelerated failure time (AFT) model, quantile regression can accommodate a more general relationship between an event time of interest and covariates while providing straightforward interpretations.

Let T and C denote time to event and time to censoring, and let $\boldsymbol{Z}$ denote a $p \times 1$ covariate vector with the first component set as 1. Without loss of generality, the censored regression quantile model investigated in this paper takes the form

$$Q_T(\tau|\boldsymbol{Z}) = \exp\{\boldsymbol{Z}^T\boldsymbol{\beta}_0(\tau)\}, \quad \tau \in (0, 1), \tag{1}$$

where $Q_T(\tau|\boldsymbol{Z}) \equiv \inf\{t : \Pr(T \leq t|\boldsymbol{Z}) \geq \tau\}$ denotes the conditional quantile function of T given $\boldsymbol{Z}$ (with the same definition applied to any other random variable), and $\boldsymbol{\beta}_0(\tau)$ is a $p \times 1$ vector of unknown regression coefficients representing covariate effects on the τ-th quantile of $\log T$. Model (1) adopts the standard random censoring mechanism. That is, T and C are assumed to be independent conditional on $\boldsymbol{Z}$. Under this independent censoring assumption, the distribution of C is allowed to depend on $\boldsymbol{Z}$.

It is worth noting that much previous work on quantile regression with censored data cannot address the regression quantile problem defined above. For example, early efforts by Powell [20,21] require that all censoring times be known or fixed and so do the subsequent work by Fitzenberger [5], Buchinsky and Hahn [3], among others. Other methods, such as those by Ying et al. [25] and Honore et al. [7], demand unconditional independence between T and C, which is a stronger assumption than the standard random censorship, and thus cannot be applied here.

Portnoy [18] made the first attempt to tackle the censored regression quantile problem (1) by novelly employing the principle of self-consistency [4]. The principle of self-consistency here, in short, refers to an estimation scenario from

E-mail address: lpeng@sph.emory.edu.

0047-259X/$ – see front matter © 2012 Published by Elsevier Inc.
doi:10.1016/j.jmva.2011.10.005

equaling the estimator to an expression that contains the estimator itself. The estimator proposed in [18] reduces to the Kaplan–Meier estimator [9] in the one-sample case. The initial iterative algorithm was polished by Neocleous et al. [16] to a grid-based estimation procedure. We hereafter refer the estimator defined in [16] to as Portnoy's self-consistent estimator, denoted by $\widehat{\boldsymbol{\beta}}_{PSC}(\tau)$. Asymptotic studies on the self-consistent estimation of regression quantile were established only until recently by Portnoy and Lin [19], partly due to the complex representation of the existing estimators. Other subsequent work of [18] includes (but is not limited to) [15], which studied the monotonicity property associated with $\widehat{\boldsymbol{\beta}}_{PSC}(\tau)$, and [24], which relaxed the global linear assumption in model (1).

Peng and Huang [17] proposed an alternative approach to estimating $\boldsymbol{\beta}_0(\tau)$ in model (1) by utilizing the martingale structure of randomly censored data. A clearly defined grid-based algorithm was developed based on a set of monotone estimating equations. These estimating equations have appealing stochastic integral representations which greatly facilitate large-sample studies. Peng and Huang [17] derived the closed form of the limit process of their estimator, denoted by $\widehat{\boldsymbol{\beta}}_{PH}(\tau)$. It was also shown that $\widehat{\boldsymbol{\beta}}_{PH}(\tau)$ becomes a Nelson–Aalen type estimator [14,1] when there is no covariate. More recently, Huang [8] derived a grid-free estimation procedure based on a new concept of quantile calculus. The resulting estimator has the same distribution as Peng and Huang [17]'s estimator.

A brief introduction of the algorithms presented in [16,17] is provided in Section 2. Both of these approaches have been implemented for *R* in the contributed package *quantreg* [11]. Comprehensive numerical studies conducted by Koenker [10] demonstrated very similar empirical performance between $\widehat{\boldsymbol{\beta}}_{PSC}(\tau)$ and $\widehat{\boldsymbol{\beta}}_{PH}(\tau)$. Though the asymptotic equivalence between these two estimators can be established in the one-sample case given the fact that the Kaplan–Meier estimator and the Nelson–Aalen estimator are equivalent in the large sample sense, their connection remains unknown in general regression settings.

In this paper, we develop a new framework to study the self-consistent estimation of model (1) with T subject to standard random censorship. With the principle of self-consistency generally perceived as an estimation strategy that defines an estimator as some function of the observed data and the estimator itself, and then utilizes such a relationship to obtain a consistent well-defined estimator, throughout this paper, the self-consistent estimation of model (1) refers to as the estimation of $\boldsymbol{\beta}_0(\cdot)$ which involves a use of self-consistency principle. Compared to previous work, the proposed method preserves the computational-ease feature, while providing a more direct approach to the asymptotic theory. Of note, the current self-consistent estimators are defined under the assumption that no censoring would occur below $\boldsymbol{Z}^T\boldsymbol{\beta}_0(\epsilon_l)$ for all $\boldsymbol{Z}$, where ϵ_l is a prespecified constant in (0, 1). This assumption is not likely to incur serious concerns in real data analysis with ϵ_l selected small enough. Nevertheless, it is practically desirable to eliminate this restriction. In Section 3, we formulate censored regression quantiles based on stochastic integral equations derived from adopting self-consistency principle. The new representation of self-consistent censored regression quantiles entails clearly defined estimation and inference procedures, which avoid the artificial data constraint stated above. In Section 4, we show the asymptotic equivalence of the proposed self-consistent estimators to $\widehat{\boldsymbol{\beta}}_{PH}(\tau)$. Therefore, we establish the uniform consistency and weak convergence of the new estimators with the closed-forms for the limit distributions derived. Furthermore, we elaborate the connection between the proposed estimators and Portnoy's self-consistent estimator. The results aid in understanding the close proximity between $\widehat{\boldsymbol{\beta}}_{PH}(\tau)$ and $\widehat{\boldsymbol{\beta}}_{PSC}(\tau)$ observed in previous empirical studies. Monte-Carlo simulations reported in Section 5 confirm our theoretical findings.

It is important to note that the proposed framework for self-consistent estimation of censored quantile regression can be extended to survival settings with more complex censoring mechanisms. A few remarks are supplied in Section 6.

2. Two existing approaches for censored quantile regression

Define $\tilde{X} = T \wedge C$ and $\delta = I(T \le C)$, where $\wedge$ is the minimum operator and $I(\cdot)$ is the indicator function. The observed randomly censored data consist of n iid replicates of $(\tilde{X}, \delta, \boldsymbol{Z})$, denoted by $\{(\tilde{X}_i, \delta_i, \boldsymbol{Z}_i),\ i = 1, \ldots, n\}$. We define $X = \log \tilde{X}$ and accordingly $X_i = \log \tilde{X}_i$.

2.1. Portnoy's self-consistent approach

We here outline the grid-based algorithm presented in [16]. A grid of τ-values is defined as $0 < \tau_1 < \tau_2 < \cdots < \tau_M < 1$. Define $m(\boldsymbol{\beta}, i, k) = \max\{l : 1 \le l \le k-1, \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_l) < X_i \le \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_{l+1})\}$ if the set $\{l : 1 \le l \le k-1,\ \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_l) < X_i \le \boldsymbol{Z}_i^T\boldsymbol{\beta}(\tau_{l+1})\}$ is not empty, and $m(\boldsymbol{\beta}, i, k) = k+1$ otherwise. By this definition, $m(\boldsymbol{\beta}, i, 0) = 1$.

Step 1. Compute $\widehat{\boldsymbol{\beta}}_{PSC}(\tau_1)$ by fitting the uncensored quantile regression with data $\{X_i, \delta_i, \boldsymbol{Z}_i\}_{i=1}^n$. It is assumed that all censored X_i's are above the hyperplane determined by $\widehat{\boldsymbol{\beta}}_{PSC}(\tau_1)$. Set $k = 1$.

Step 2. Given $\widehat{\boldsymbol{\beta}}_{PSC}(\tau_l)$ $(l \le k)$, obtain $\widehat{\boldsymbol{\beta}}_{PSC}(\tau_{k+1})$ by minimizing the following weighted check function:

$$\sum_{\delta_i=1} \rho_\tau(X_i - \boldsymbol{Z}_i^T\boldsymbol{b}) + \sum_{\delta_i=0} \left\{ \hat{w}_{k+1,\ i}\rho_\tau(X_i - \boldsymbol{Z}_i^T\boldsymbol{b}) + (1 - \hat{w}_{k+1,\ i})\rho_\tau(X^* - \boldsymbol{Z}_i^T\boldsymbol{b}) \right\}, \tag{2}$$

where $\hat{w}_{k+1,i} = (\tau_{k+1} - \tau_{m(\widehat{\boldsymbol{\beta}}_{PSC},i,k)})/(1 - \tau_{m(\widehat{\boldsymbol{\beta}}_{PSC},i,k)})$ if $m(\widehat{\boldsymbol{\beta}}_{PSC}, i, k) < k+1$ and 1 otherwise, $\rho_\tau(u) = u\{\tau - I(u < 0)\}$, and X^* is an extremely large value.