Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 6488–6497
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

The Koul–Susarla–Van Ryzin and weighted least squares estimates for censored linear regression model: A comparative study

Yanchun Bao[a,*], Shuyuan He[b], Changlin Mei[c]

[a] School of Mathematics, Manchester University, Manchester, M13 9PL, UK
[b] School of Mathematical Sciences, Peking University, Beijing, People's Republic of China
[c] School of Science, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

Received 19 November 2005; received in revised form 25 February 2007; accepted 25 February 2007
Available online 5 March 2007

Abstract

The Koul–Susarla–Van Ryzin (KSV) and weighted least squares (WLS) methods are simple to use techniques for handling linear regression models with censored data. They do not require any iterations and standard computer routines can be employed for model fitting. Emphasis has been given to the consistency and asymptotic normality for both estimators, but the finite sample performance of the WLS estimator has not been thoroughly investigated. The finite sample performance of these two estimators is compared using an extensive simulation study as well as an analysis of the Stanford heart transplant data. The results demonstrate that the WLS approach performs much better than the KSV method and is reliable for use with censored data.
© 2007 Elsevier B.V. All rights reserved.

Keywords: Censored data; Linear regression model; Weighted least squares; Stanford heart transplant data

1. Introduction

Let Y be a response variable and $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_r)'$ be a random vector of covariates. Assume that Y and $\mathbf{Z}$ follow a linear regression model

$$Y = \mathbf{b}'\mathbf{Z} + \varepsilon, \tag{1}$$

where $\mathbf{b} = (b_1, b_2, \ldots, b_r)'$ is the vector of parameters and ε, uncorrelated with $\mathbf{Z}$, is the error term with mean zero and variance σ_ε^2. When the response Y is subject to random right censoring, we only observe (U_i, Z_i, δ_i), $i = 1, 2, \ldots, n$, which are n replications of (U, Z, δ), where $U = \min(Y, C)$, $\delta = I[Y \leqslant C]$ and C is the censoring random variable which is independent of the response Y. Note the censoring works both ways: if $\delta = 0$, then the response Y is censored by the censoring variable C; if $\delta = 1$, on the other hand, the censoring variable C is censored by the response Y. The censored regression problems focus on estimating the parameter vector $\mathbf{b}$ and investigating the related statistical properties of the estimators based on the observations (U_i, Z_i, δ_i), $i = 1, 2, \ldots, n$.

Many techniques have been proposed for handling the above regression problems. One methodology in this aspect is based on the *synthetic data* and uses the ordinary least squares procedure to obtain an estimator of $\mathbf{b}$. It includes

* Corresponding author.
E-mail addresses: ybao@maths.man.ac.uk (Y. Bao), syhe@maths.pku.edu.cn (S. He), clmei@mail.xjtu.edu.cn (C. Mei).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.02.025

Miller's estimator (Miller, 1976), Buckley and James' estimator (Buckley and James, 1979; James and Smith, 1984; Jin et al., 2006), Leurgans' estimator (Leurgans, 1987; Zhou, 1992a) and in this paper the so-called KSV estimator (Koul et al., 1981; Srinivasan and Zhou, 1994; Fygenson and Zhou, 1992, 1994; Lai et al., 1995). Another estimation method is the weighted least squares estimate (abbreviated as WLS in this paper) (Stute, 1993, 1996; He and Wong, 2003). In the former methodology, the KSV approach is the easiest to be carried out because no iterations are required and standard least squares computer routines can be used once the observations of the response are transformed by the censoring information. Similar advantages are also achieved by the WLS method. Although much work has been devoted to the study of the statistical properties such as consistency and asymptotic normality, respectively, for these two methods (see above references), it makes sense to compare their finite sample performance. This paper serves this purpose by means of an extensive simulation study and an analysis of the Stanford heart transplant data.

The remains of this paper are organized as follows. In Section 2, both the KSV and WLS procedures for linear regression models with censored data are briefly described and some theoretical properties of the estimators such as consistency and asymptotic normality are also summarized. Extensive simulations are conducted in Section 3 to compare the finite sample performance of these two estimators under various censoring patterns and underlying distributions of the covariates and error term. In Section 4, the famous Stanford heart transplant data are analyzed by these two methods and the results are compared with those obtained by Miller and Halpern (1982). The paper then concludes with a brief summary and discussion.

2. The WLS and KSV estimates for censored linear regression models

2.1. Some notations

For any distribution function $F(x)$, let $\bar{F}(x) = 1 - F(x)$ and (a_F, b_F) be the range of F defined by

$$a_F = \inf\{x : F(x) > 0\} \quad \text{and} \quad b_F = \sup\{x : F(x) < 1\}.$$

Let

$$F(x) = P(Y \leqslant x), \quad G(x) = P(C \leqslant x) \quad \text{and} \quad H(x, \mathbf{z}) = P(Y \leqslant x, \mathbf{Z} \leqslant \mathbf{z}) \tag{2}$$

be the distribution functions of the response Y, censoring variable C and the joint distribution function of Y and the covariates $\mathbf{Z}$, respectively, where $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_r)'$, $\mathbf{z} = (z_1, z_2, \ldots, z_r)'$ and $\mathbf{Z} \leqslant \mathbf{z}$ means that $Z_i \leqslant z_i$ for all $i = 1, 2, \ldots, r$. Furthermore, we assume throughout the paper that

$$\begin{aligned} &\text{(i) } P(Y \leqslant C | Y, \mathbf{Z}) = P(Y \leqslant C | Y), \\ &\text{(ii) } b_F \leqslant b_G. \end{aligned} \tag{3}$$

As pointed out by Stute (1996), the assumption (i), together with the independence of Y and C, will guarantee that the joint distribution of $(Y, \mathbf{Z})$ can be theoretically derived from that of $(U, \mathbf{Z}, \delta)$ and consistently estimated from a sample $(U_i, \mathbf{Z}_i, \delta_i)$, $i = 1, 2, \ldots, n$. (ii) is a commonly used assumption in the literature of regression with censored data.

2.2. The WLS estimation

The WLS estimator of $\mathbf{b}$ and its consistency and asymptotic normality were proposed by Stute (1993, 1996). From the moment estimation viewpoint, He and Wong (2003) obtained the WLS estimator of $\mathbf{b}$ as well as the estimator of σ_ε^2 and proved the asymptotic normality for both estimators. He and Wong (2003)'s method requires simpler conditions for the asymptotic normality of the estimators and provides a more concise form of limit covariance matrix in contrast to that derived by Stute, 1996. Here we briefly describe this method and the related results as follows.

Let

$$Z_0 = Y, \quad \mathbf{Z} = (Z_1, \ldots, Z_r)', \quad \boldsymbol{\Gamma} = \mathrm{E}(\mathbf{Z}\mathbf{Z}') = (\mu_{ij})_{i,j=1}^r \quad \text{and} \quad \mathbf{r} = (\mu_{01}, \ldots, \mu_{0r})', \tag{4}$$

where $\mu_{ij} = \mathrm{E}(Z_i Z_j)$, $1 \leqslant i, j \leqslant r$ and $\mu_{0j} = \mathrm{E}(Y Z_j)$, $1 \leqslant j \leqslant r$. Multiplying the two sides of Eq. (1) by $\mathbf{Z}'$ and taking expectations yields

$$\mathbf{r} = \boldsymbol{\Gamma}\mathbf{b}. \tag{5}$$

متن کامل مقاله

دریافت فوری

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات