COMSTA: 5881 Model 3Gsc pp. 1–13 (col. fig: NIL)

ARTICLE IN PRESS

Computational Statistics and Data Analysis xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Random weighting approximation for Tobit regression models with longitudinal data

Q1 L.Q. Xiao, B. Hou, Z.F. Wang *, Y.H. Wu

Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui 230026, China

ARTICLE INFO

Article history:
Received 23 June 2013
Received in revised form 6 March 2014
Accepted 30 May 2014
Available online xxxx

Keywords:
Longitudinal data
Random weighting
Tobit response
M-test statistic

ABSTRACT

Longitudinal data arise naturally in medical studies, psychology, sociology and so on. Due to some lower detection limits the responses are often left censored, which are called Tobit responses in econometrics. For Tobit response regression models with longitudinal data, quantile estimators of regression parameters and M-test statistics for linear hypotheses are constructed. In addition, distributions of the proposed estimators and test statistics are formed by random weighting method. The proposed methods do not need to estimate nuisance parameters involved in asymptotic distributions of the developed statistics. Extensive simulations and a real data example are presented to demonstrate the performance of the proposed methods.

1. Introduction

Longitudinal data, which arise naturally in medical studies, psychology, sociology and so on, represent such type of data that the outcomes and covariates are measured repeatedly over a study period. When there is a lower detection limit which may depend on the accuracy of measurement tools or mechanism, the observations are often left censored. Many methods have been introduced to handle left censored longitudinal data, such as multiple imputation (Paxton et al., 1997), likelihood-based methods (Hughes, 1999; Thiébaut and Jacqmin-Gadda, 2004), and Bayesian analysis (AlHamzawi, 2013; Kobayashi and Kozumi, 2012). It is noticed that the simple imputation methods lead to biased estimation and likelihood-based methods need to assume the distribution structure of error terms, at the cost of higher standard deviation. Based on the rank score test introduced by Gutenbrunner et al. (1993) for linear regression model, Wang and Fygenson (2009) proposed a quantile rank score test (QRS) to test hypotheses on parameters of quantile regression models for censored longitudinal data. Their approach does not need distributional assumption and the proposed quantile estimator of regression parameters is asymptotically unbiased.

There are, however, some common issues related with Wang and Fygenson's method. One issue is that the asymptotic distributions of parameter estimates and QRS statistics involve nuisance parameters such as density function and correlation structure of error terms among correlated observations. These nuisance parameters need to be estimated when conducting hypothesis testing. When the correlation structure is simple, such as when the error terms of any two measurements for one subject have same joint distribution (Wang and Fygenson, 2009), the correlation structure can be estimated. But it is well-known that complicated correlation structures or density functions may not be estimated well, especially when the sample size is small.

One approach to solve this problem is to obtain the distribution of corresponding estimates by random weighting method. For non-longitudinal data with Tobit response, i.e. responses that are left censored, Zhao and Fang (2004) applied random

* Corresponding author. Tel.: +86 551 63600560.
E-mail address: zfw@ustc.edu.cn (Z.F. Wang).

http://dx.doi.org/10.1016/j.csda.2014.05.020
0167-9473/

weighting method to approximate the distribution of the least absolute deviation (LAD) estimator; Wang et al. (2009) proposed a random weighting test statistic to test linear hypotheses. Recently, Liu et al. (2013) used random weighting method to obtain the distribution of a group variable shrinkage estimator of regression parameters for predefined groups of variables.

There is not much literature to apply the random weighting method to longitudinal data with left censored response. In this paper we consider the linear regression model with longitudinal data and left censored response. We call this Tobit regression model with longitudinal data. We use empirical process techniques to relax the correlation structure of the repeated measurements. We construct a quantile estimator to estimate the regression parameters. Linear hypotheses are tested by an M-test statistic. It is natural that the asymptotic distributions of the parameter estimates and the M-test statistic involve nuisance parameters such as some density function and correlation structure for repeated measurements. The within sample correlation also causes the asymptotic distribution of the M-test statistic to deviate from the usual chi-squared distribution. To overcome these problems we apply the random weighting method to construct the parameter estimates and the M-test statistics, which we call weighting estimator and weighting test statistic respectively. The distributions of the parameter estimates and testing statistics are estimated directly by those of the weighting estimator and weighting test statistic.

This paper is organized as follows. In Section 2, we introduce the weighting estimator and weighting M-test statistics for our model and provide their asymptotic properties. Simulation results are shown in Section 3. In Section 4, we apply our method to real data. Some conclusions are given in Section 5. All technical proofs are presented in the Appendix.

2. Proposed estimation and test

Suppose there are m subjects and for the ith subject, there are n_i measurements of the response and covariate variables. For the total of $n = \sum_{i=1}^{m} n_i$ observations, our regression model is

$$y_{ij}^{+} = \max(0, x_{ij}'\beta_0 + u_{ij}), \quad j = 1, \ldots, n_i, \; i = 1, \ldots, m, \tag{1}$$

where $y_{ij}^{+} = y_{ij} I(y_{ij} \geq 0)$ with $I(\cdot)$ being an indicator function, x_{ij} is the jth observation of the p covariate variables for subject i, β_0 is the $p \times 1$ vector of regression parameters and u_{ij} is an error term. Since the response variable has a non-negative constraint, we call model (1) the Tobit response regression model with longitudinal data. As usual, we assume that errors across subjects are independent and those within each subject may be dependent.

2.1. Estimation of regression parameters

Wang and Fygenson (2009) proposed a quantile estimate $\hat{\beta}_n$ of β_0

$$\hat{\beta}_n = \arg\min_{\beta} \sum_{ij} \rho_\tau \left\{ y_{ij}^{+} - \max(0, x_{ij}'\beta) \right\} \tag{2}$$

where $\rho_\tau(u) = u \cdot \{\tau - I(u < 0)\}$ is a quantile loss function. Under certain regularity conditions, they showed that $\hat{\beta}_n$ is consistent and asymptotically normal. The asymptotic covariance matrix of $\hat{\beta}_n$ depends on some unknown marginal density function and correlation structure (joint probabilities) of the error terms. In this paper, we use random weighting approach to estimate the distribution of $\hat{\beta}_n$.

Define the random weighting estimator β_n^* of β_0 as

$$\beta_n^* = \arg\min_{\beta} \sum_{i} \omega_i \sum_{j} \rho_\tau \left\{ y_{ij}^{+} - \max(0, x_{ij}'\beta) \right\}, \tag{3}$$

where $\omega_i, \; i = 1, \ldots, m$ are weight variables. Before studying the properties of β_n^*, we first list our assumptions as follows.

(C1) β_0 is an interior point of a compact parameter space $B \in R^p$.

(C2) For each $\epsilon > 0$ there is a finite $\nu > 0$ such that

$$n^{-1} \sum_{ij} \|x_{ij}\|^2 I(\|x_{ij}\| > \nu) < \epsilon \quad \text{for all } n \text{ large enough,}$$

where $\|\cdot\|$ is the Euclidean norm.

(C3) For each $\epsilon > 0$ there is κ such that

$$n^{-1} \sum_{ij} \|x_{ij}\|^2 I(|x_{ij}'\beta_0| \leq \kappa) < \epsilon \quad \text{for all } n \text{ large enough.}$$

(C4) $D_n = n^{-1} \sum_{ij} I(x_{ij}'\beta_0 > 0) x_{ij} x_{ij}' \to D$, as $n \to \infty$, where D is a positive definite matrix.

(C5) $\{u_{ij}\}$ have a common marginal distribution function F whose τth quantile is zero, and a continuous, strictly positive density $f(\cdot)$ near zero.

Please cite this article in press as: Xiao, L.Q., et al., Random weighting approximation for Tobit regression models with longitudinal data. Computational Statistics and Data Analysis (2014), http://dx.doi.org/10.1016/j.csda.2014.05.020