Statistics and Probability Letters 105 (2015) 96–105

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Two-step semiparametric estimation of the Type-3 Tobit model

Xianbo Zhou, Zhewen Pan *

Lingnan (University) College, Sun Yat-sen University, Guangzhou, 510275, China

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 9 June 2015
Accepted 9 June 2015
Available online 16 June 2015

Keywords:
Type-3 Tobit model
Two-step
Semiparametric estimation

ABSTRACT

This paper proposes an alternative two-step semiparametric estimator for the Type-3 Tobit model and presents its asymptotic properties. Simulation shows that the estimator performs well in various designs. An empirical application is provided to verify the usefulness of the estimator.

1. Introduction

Selectivity bias is a frequently encountered problem in applied economic fields such as labor supply, migration, and health insurance models. Two types of sample selection models are commonly used in the empirical modeling, namely Type-2 and Type-3 Tobit models (cf. Amemiya, 1985: Chapter 10). The difference between the two is that the selection variable is a binary variable (e.g., participant or non-participant) in the former and a censored variable in the latter. Obviously the latter uses more information about the selection variable than the former does. This increment of information helps in both identifying the model parameters under weaker restrictions and obtaining more efficient estimations. For example, exclusion restrictions required for identification of the Type-2 Tobit model are no longer necessary for identification of the Type-3 Tobit model provided the selection variable is observable in a continuous range; see Lee (1994) for this discussion. Lee also provided some simulation evidences of efficiency gains from exploiting the additional information. As a result, estimations of the Type-3 Tobit model are of independent interest to econometrics literature, although it can be estimated by techniques developed for the Type-2 Tobit model.

The Type-3 Tobit model is typically represented as a two-equation model of the following form:

$$Y_1 = \max\left\{Y_1^*, 0\right\} = \max\left\{W'\alpha_0 + u_1, 0\right\}, \tag{1.1}$$

$$Y_2 = Y_2^* \cdot 1\left\{Y_1^* > 0\right\} = \left(X'\beta_0 + u_2\right) \cdot 1\left\{W'\alpha_0 + u_1 > 0\right\}, \tag{1.2}$$

where $1\{A\}$ is the usual indicator function of the event A, W and X are random vectors of regressors with dimensions k_1 and k_2 respectively, u_1 and u_2 are error terms that are independent of W and X, $\alpha_0 \in \Theta_1$ and $\beta_0 \in \Theta_2$ are conformable vectors consisting of unknown parameters. $Y_1^* = W'\alpha_0 + u_1$ and $Y_2^* = X'\beta_0 + u_2$ are latent dependent variables. We observe $Y_1 = \max\left\{Y_1^*, 0\right\}$, the selection variable, and $Y_2 = Y_2^* \cdot 1\left\{Y_1^* > 0\right\}$, the variable of main interest.

* Correspondence to: Room 304, Lam Woo Hall, Sun Yat-sen University, No.135 West Xingang Road, Guangzhou, China. Tel.: +86 13416167301.
E-mail address: panzhew@mail2.sysu.edu.cn (Z. Pan).

http://dx.doi.org/10.1016/j.spl.2015.06.012
0167-7152/

The Type-3 Tobit model is traditionally estimated by maximum likelihood method or Heckman's two-step method (Heckman, 1979). It is well known that, however, consistency of such likelihood-based parametric estimators relies heavily upon correct specification of error term's distribution function, in the presence of limited dependent variables. To lower the risk of misspecification, semiparametric estimation procedures relax parametric restrictions imposed on the functional form of error term's distribution. Some existing studies (Chen, 1997; Honoré et al., 1997; Lee, 1994; Li and Wooldridge, 2002) suggested various moment-based approaches to the semiparametric estimation of the Type-3 Tobit model (1.1)–(1.2). Min and Sheu (2003) presented a Monte Carlo comparison of these semiparametric estimators. In this paper, we exploit the entire jointly conditional distribution of the limited dependent variables Y_1 and Y_2, instead of the conditional moment of Y_2, to propose a new two-step semiparametric estimation method for the Type-3 Tobit model. In the spirit of Heckman's two-step procedure, the selection equation (1.1) that is a censored regression model is semiparametrically estimated first, and the obtained estimator is used to eliminate the selection effect on the second-step estimation of the outcome equation (1.2).

The remainder of the paper is organized as follows. Section 2 introduces the estimator. Section 3 presents the large sample properties of the estimator. Section 4 reports the simulation results. Section 5 gives a real data justification. Section 6 concludes.

2. The estimator

We first state the motivation of our semiparametric estimator for Model (1.1)–(1.2). Denote

$$S(a,b) = \Pr\{u_1 > a, u_2 > b\}.$$

Let Z be a random vector consisting of the distinct components in W and X, then for any $t \in (-\infty, \infty)$ and $z \in \Omega$, where Ω is the support of Z, we have

$$\begin{aligned} \Pr\{Y_1 > 0, Y_2 > t \,|Z = z\} &= \Pr\left\{u_1 > -W'\alpha_0,\ u_2 > t - X'\beta_0 \,|Z = z\right\} \\ &= S\left(-w'\alpha_0, t - x'\beta_0\right). \end{aligned} \tag{2.1}$$

On the other hand, denote $u_1(\alpha) \equiv Y_1 - W'\alpha$ and $u_2(\beta) \equiv Y_2 - X'\beta$, where $\alpha \in \Theta_1$ and $\beta \in \Theta_2$. For any $\tilde{z} \in \Omega$ satisfying $(\tilde{w} - w)'\alpha > 0$, we have

$$\begin{aligned} &\Pr\left\{u_1(\alpha) > -w'\alpha, u_2(\beta) > t - x'\beta \,\middle|Z = \tilde{z}\right\} \\ &= \Pr\left\{Y_1 > W'\alpha - w'\alpha, Y_2 > t + X'\beta - x'\beta \,\middle|Z = \tilde{z}\right\} \\ &= \Pr\left\{u_1 > -W'\alpha_0 + W'\alpha - w'\alpha, u_2 > t - X'\beta_0 + X'\beta - x'\beta \,\middle|Z = \tilde{z}\right\} \\ &= S\left(-\tilde{w}'\alpha_0 + \tilde{w}'\alpha - w'\alpha, t - \tilde{x}'\beta_0 + \tilde{x}'\beta - x'\beta\right), \end{aligned}$$

which is identical to (2.1) when $\alpha = \alpha_0$ and $\beta = \beta_0$. That is, for any fixed $z \in \Omega$,

$$E\left[1\{Y_1 > 0, Y_2 > t\} \,|Z = z\right] = \frac{E\left[1\left\{Y_1 > (W-w)'\alpha_0 > 0, Y_2 > t + (X-x)'\beta_0\right\}\right]}{E\left[1\left\{(W-w)'\alpha_0 > 0\right\}\right]}. \tag{2.2}$$

The key relationship (2.2) can be used to construct a least-square-type estimator, if the right hand side can be consistently estimated at a given pair of α and β. Assume that $\{(Y_{1i}, Y_{2i}, Z_i)\}_{i=1}^n$ is a random sample of (Y_1, Y_2, Z). We estimate the right hand side of (2.2) by its sample analog,

$$\tau_n(t, z, \alpha, \beta) = \frac{A_n(t, z, \alpha, \beta)}{B_n(z, \alpha)}, \tag{2.3}$$

where

$$A_n(t, z, \alpha, \beta) = \frac{1}{n}\sum_{j=1}^{n} 1\left\{Y_{1j} > \left(W_j - w\right)'\alpha, Y_{2j} > t + \left(X_j - x\right)'\beta\right\} m_1\left(\left(W_j - w\right)'\alpha\right)$$

and

$$B_n(z, \alpha) = \frac{1}{n}\sum_{j=1}^{n} m_1\left(\left(W_j - w\right)'\alpha\right).$$

Here we adopt $m_1(\cdot)$, a smoothed version of the indicator function $1\{\cdot > 0\}$, for technical reasons.

Suppose the preliminary estimator for α_0 in the selection equation (1.1) is given by $\hat{\alpha}$. For example, the pairwise differenced estimator proposed by Honoré and Powell (1994) for the censored regression model, which will be employed as the first-step estimator in our subsequent simulations, is defined by

$$\hat{\alpha} = \arg\min_{\alpha \in \Theta_1} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} s\left(Y_{1i}, Y_{1j}, \left(W_i - W_j\right)'\alpha\right) \tag{2.4}$$