Detection of a Tobit Model

Elouali Rahmani

University of Mohammed first (FS), Oujda, Morocco

Abdelali Kaouachi

University of Mohammed first (EST), Oujda, Morocco

Copyright © 2015 E. Rahmani and A. Kaouachi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this work we will focus to search the mathematical tools and statistics for detection a Tobit model from the observation of a sample (that is to say to prove that it is a censored linear model) we do remember, some properties of Tobit models, such as the estimate, focusing on least squares estimator and likelihood estimator. The purpose of this article is how to detect a Tobit model from a sample observations.

Keywords: Detection of a Tobit model, parametric estimation

1 Introduction

The founder of Tobit models is James Tobin [7]. In this paper, he analyzed household expenditure on durable goods using a regression model which specifically took account of the fact that the expenditure (the dependent variable of his regression model) cannot be negative. Tobin called his model the model of limited dependent variables. It and its various generalizations are known popularly among economists as Tobit models, a phrase coined by Goldberger [3] because of similarities to probit models. These models are also known as censored or truncated regression models. The model is called truncated if the observations outside a specified range are totally lost, and censored if one can at least observe the exogenous variables. Censored and truncated regression models have been developed in other disciplines (notably biometrics and engineering) more or less independently of their development in econometrics.
Biometricians use the model to analyze the survival time of a patient. Amemiya [1] identified five kinds of Tobit models, the Tobit model is simply called Type 1 Tobit model, his fundamental paper [7] on the censored normal regression model was the first systematic application in econometrics of the uniform law of large numbers and the central limit theorems required to establish consistency and asymptotic normality of nonlinear econometric models. The Tobit model is used when one is faced with many of observations for which the value of the endogenous variable is zero. This model implies that the observed value of the dependent variable is censored at zero.

In this paper, we do remember parametric estimators of Type 1 Tobit model that of the theoretical part was studied by Amemiya[2]. In particular, we examine the performance for finite samples of two different estimators of simple Tobit model: the LS estimator and the ML estimator. The purpose of this article is to determine properties of two estimators, namely bias and convergence.

The paper is organized as follows: In the second section, we discuss about parametric estimation of two different estimators simple Tobit model: the least squares estimator and the maximum likelihood estimator. In the third section, we will be interested in the detection of a Tobit model.

2 Parametric estimation

This section presents parametric estimators of Type 1 Tobit model: the least squares estimator and the maximum likelihood estimator.

A simple Tobit model (Type 1) is defined by the following equation:

\[
 y_i = \begin{cases} 
 y_i^* & \text{if } y_i^* > 0 \\
 0 & \text{if } y_i^* \leq 0 
\end{cases}
\]

(1)

where

\[
 y_i^* = x_i \beta + \varepsilon_i, \quad i = 1, ..., N,
\]

with \( y_i \) is the observed variable associated with the latent variable \( y_i^* \), \( x_i = (x_{i1}, ..., x_{iK}) \) is a vector of observable variables,

\[
 \beta' = (\beta_1, ..., \beta_K) \in \mathbb{R}^K; \text{ is the vector of unknown parameters and } (\varepsilon_i) \text{ is the errors which are independent and identically distributed according to a } N(0, \sigma^2).
\]
2.1 Least squares estimator

2.1.1 First situation: the exogenous variables \((x_i)\) are deterministic

The LS estimator, applied to all \(N\) pairs of observations \((y_i, x_i)\), is defined by:

\[
\hat{\beta}_{LS} = \left( \sum_{i=1}^{N} x_i x_i \right)^{-1} \sum_{i=1}^{N} x_i y_i.
\]

In this case, the LS estimator is inconsistent [2] and it is relatively difficult to give a general result on the form of bias of this estimator.

2.1.2 Second situation: \(x_i\) are random variables

Deal with this situation for which the exogenous variables \(x_i\) are random, Goldberger [4] studied this case by introducing a constant in the regression model. He wrote the model in the form:

\[
y_i = \begin{cases} 
y^*_i & \text{if } y^*_i > 0 \\
0 & \text{if } y^*_i \leq 0
\end{cases} \quad (2)
\]

with

\[
y^*_i = \alpha + x_i \beta + \varepsilon_i, \quad i = 1, ..., N,
\]

\(x_i = (x^1_i, ..., x^K_i)\) is a vector of random normal distribution \(N(0, \Omega)\), where \(\Omega\) is \((k \times k)\) defined positive matrix and

\[
\beta' = (\beta_1, ..., \beta_K) \in \mathbb{R}^K, \alpha \in \mathbb{R}
\]

are the unknown parameters.

**Fundamental assumption \((FH)\)**

We assume that the variables \((x_i)\) are distributed according to a normal \(N(0, \Omega)\), with

\[
cov(\varepsilon_i, x^k_i) = 0, \quad \text{for all } k = 1, ..., K.
\]

**Proposition 2.1** [Greene[5]] Under the assumption \((FH)\), the LS estimator, obtained on all observations \((x_i, y_i)\), satisfies \(\hat{\beta}_{LS} \to \beta \times \Phi \left( \frac{\alpha}{\sigma_y} \right)\), in probability, when \(N \to \infty\), where \(\sigma^2_y = \sigma^2 + \beta' \Omega \beta\).

Under the assumption \((FH)\), the estimator defined by \(\hat{\beta}_{LSC} = (N/N_1)\hat{\beta}_{LS}\), where \(N_1\) is the number of observations for which \(y^*_i > 0\), is a consistent estimator of \(\beta\):

\(\hat{\beta}_{LSC} \to \beta\), in probability, when \(N \to \infty\).
2.2 Maximum likelihood estimator

In 70 years, the estimation procedure in two steps of Heckman and other methods of estimation were preferred because they do not require too much calculation. Today with the development of computer that can treat quickly the optimization problems of functions, the maximum likelihood estimator is more favored. Using reparametrization of Olsen [6], the log-likelihood of a Tobit model is:

$$\sum_{i:y_i>0} \log L(y_i, \theta, h) = \sum_{i:y_i=0} \log(1 - \Phi(x_i\theta)) - 1/2 \sum_{i:y_i>0} (hy_i - x_i)^2 + N_1 \log(h)$$

where $N_1$ is the number of observations for which $y_i > 0$. The LM estimator $\hat{\gamma} = (\hat{\theta}', \hat{h}')$ of $\gamma$ is the solution of the problem:

$$\hat{\gamma} = \arg \max_{\gamma} \log L(y, \gamma) = \arg \max_{(\theta, h)} \log L(y, \theta, h).$$

Amemiya [1] proved that the maximum likelihood estimator $\hat{\beta} = \hat{\theta}/\hat{h}$ and $\hat{\sigma} = 1/\hat{h}$ are strongly consistent and asymptotically normal.

3 Detection of a model Tobit to from observations

In this section we give statistical tools for the detection of a model Tobit to from observations of a sample.

**Problem1** Let a sample $\xi$ of points $(x_i, y_i)$ for $i=1,\ldots,n$ We want to find if there exists a variable $y^*$ satisfying:

$$y^* = x_i\beta + \alpha + \epsilon_i$$

for $i=1,\ldots,n$ such as $y_i = \max(0, y^*)$ with $\alpha, \beta, \sigma^2$ known and $\epsilon_i$ is the errors which are independent and identically distributed according to a $N(0, \sigma^2)$.

**Proposition 3.1** The problem 1 admits a solution if the following conditions are satisfied:

1) For all $i \in \{1, \ldots, n\}$, such that $y_i > 0$ we have $\epsilon_i = y_i - x_i \star \beta - \alpha$

2) For every $i \in \{1, \ldots, n\}$, such that $y_i = 0$, there exists $\epsilon_i$ such that $y_i^* < 0$

3) $(y_i^* - x_i\beta - \alpha)/\sigma \sim N(0, 1)$ for $i=1,\ldots,n$
3.1 Application

Let a sample $\xi$ of points $(x_i, y_i) \in E$
with $E = \{(0.55,2.51); (-0.85,4.37); (0.35,2.44); (0.06,5.79); (-1.28,0.00); (-0.58,0.00); (-1.05,0.00); (-1.02,5.69); (-1.20,0.00); (-1.26,0.00); (1.27,10.71); (-0.34,2.73); (-1.63,0.00); (-0.04,1.19); (0.99,5.09); (0.42,7.25); (0.60,0.00); (0.75,0.00); (1.11,1.33); (-0.10,0.00); (-0.40,7.51); (0.15,8.46); (-0.42,6.31); (0.01,1.45); (-0.52,3.44); (-0.74,1.21); (1.27,5.79); (-0.77,0.00); (-1.66,0.00); (0.32,4.21); (0.41,0.97); (0.24,0.00); (-2.03,0.00); (-1.33,0.00); (0.73,0.00); (0.82,0.00); (-0.22,0.00); (-1.11,0.00); (0.71,0.07); (-1.88,0.00); (1.72,1.83); (0.83,0.00); (1.98,0.40); (-0.09,0.00); (-0.17,5.44); (-0.48,0.00); (0.56,2.84); (-1.14,0.00); (1.45,3.47); (1.40,9.54); (1.17,4.57); (-1.25,1.39); (-0.26,2.13); (0.61,0.00); (0.12,2.07); (1.68,11.09); (0.43,2.07); (-0.12,0.00); (0.10,5.82); (1.52,0.00); (0.61,0.00); (-0.74,1.08); (-0.80,0.00); (0.51,0.00); (-0.59,1.12); (0.37,5.37); (0.70,2.67); (1.05,5.65); (-0.01,4.12); (0.61,0.50); (-0.50,0.00); (-0.89,0.00); (1.59,3.23); (0.18,0.38); (0.22,0.00); (0.57,2.99); (-0.20,0.00); (-1.90,0.00); (-1.01,0.00); (1.47,4.95); (0.05,0.62); (0.35,0.00); (-0.82,0.00); (-1.34,0.00); (-0.21,5.05); (0.42,5.43); (0.52,5.53); (-0.52,6.58); (-0.28,0.22); (1.63,6.73); (0.69,0.00); (-1.46,0.00); (-0.48,0.00); (-0.14,2.93); (-0.96,0.00); (1.30,6.04); (0.77,10.60) \}

We search the $y_i^*$ satisfying the conditions of the proposition, with $(\alpha, \beta, \sigma) = (1,2,4)$

1) is verified by construction

2) for all $i = 1,...,N$ tel que $y_i = 0$ we generate $y_i^*$

$y_i^* \sim N(x_i \beta + \alpha, \sigma^2)$ and we chose $\epsilon_i$ way to have $y_i^* < 0$

Therefore we will have $N$ couples $(x_i, y_i^*) \in E'$ with $E' = \{(0.55,2.51); (-0.85,4.37); (0.35,2.44); (0.06,5.79); (-1.28,-4.93); (-0.58,-2.07); (-1.05,-4.56); (-1.02,5.69); (-1.20,-2.67); (-1.26,-11.76); (1.27,10.71); (-0.34,2.73); (-1.63,-2.10); (-0.04,1.19); (0.99,5.09); (0.42,7.25); (0.60,-1.14); (0.75,-4.22); (1.11,1.33); (-0.10,-6.69); (-0.40,7.51); (0.15,8.46); (-1.23,-4.34); (0.11,-1.49); (-0.42,6.31); (-0.93,0.34); (0.01,1.45); (-0.52,3.44); (-0.74,1.21); (1.27,5.79); (-0.77,-10.66); (-1.66,-8.85); (0.32,4.21); (0.41,0.97); (0.24,-7.04); (-2.03,4.95); (-1.33,-3.06); (0.73,-2.13); (0.82,-0.21); (-0.22,-2.04); (-1.11,-3.43); (0.71,0.07); (-1.88,-1.44); (1.72,1.83); (0.83,-1.41); (1.98,0.40); (-0.09,-4.12); (-0.17,5.44); (-0.48,-3.10); (0.56,2.84); (-1.14,-0.16); (1.45,3.47);
3) It remains to verify \( z_i = y_i^* - x_i \beta - \alpha / \sigma \sim N(0,1) \) for \( i = 1, \ldots, n \) by drawing a table of the \( z_i \) we shown by the test of the chi-squared that \( z_i \) can be adjusted by a standard normal distribution and reduced.

Applying test of Khi two of adjustment

\( H_0: z_i \sim N(0, \sigma^2) \).

\( H_1: z_i \) does not follow a standard normal distribution.

<table>
<thead>
<tr>
<th>Classe</th>
<th>( n_i )</th>
<th>( p_i )</th>
<th>( t_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( &lt; -1.5 )</td>
<td>10</td>
<td>0.0668</td>
<td>6.68</td>
</tr>
<tr>
<td>(-1.5, -1] )</td>
<td>09</td>
<td>0.0918</td>
<td>9.18</td>
</tr>
<tr>
<td>(-1, -0.5] )</td>
<td>14</td>
<td>0.1499</td>
<td>14.99</td>
</tr>
<tr>
<td>(-0.5, 0] )</td>
<td>17</td>
<td>0.1915</td>
<td>19.15</td>
</tr>
<tr>
<td>(0, 0.5] )</td>
<td>21</td>
<td>0.1915</td>
<td>19.15</td>
</tr>
<tr>
<td>(0.5, 1] )</td>
<td>14</td>
<td>0.1499</td>
<td>14.99</td>
</tr>
<tr>
<td>(1, 1.5] )</td>
<td>7</td>
<td>0.0918</td>
<td>9.18</td>
</tr>
<tr>
<td>( \geq 1.5 )</td>
<td>8</td>
<td>0.0668</td>
<td>6.68</td>
</tr>
</tbody>
</table>

\( n \) = Sample size
\( n_i \) is the observed effectif
\( p_i = p(a_i \leq x < b_i) = \text{pnorm}(b_i) - \text{pnorm}(a_i) \) with \( x \sim N(0, 1) \)
\( t_i \) theoretical effectif, \( t_i = n * p_i \) the test conditions are verified:
- \( t_i \geq 5 \) For all \( i = 0, 1, \ldots, 7 \)
- \( n \geq 50 \)

The number of degrees of freedom is \( ddl = 8 - 1 = 7 \)

\[ \chi^2_{\text{observed}} = \sum (n_i - t_i)^2 / t_i = 2.98 \]

and \( \chi^2_{\text{threshold}} = 14.07 \)

so, \( \chi^2_{\text{observed}} < \chi^2_{\text{threshold}} \)

As that we accept, the hypothesis \( H_0 \)
References


Received: February 17, 2015; Published: March 12, 2015