

Applied Microeconometrics (L9): Corner solution model-Tobit

Nicholas Giannakopoulos

University of Patras
Department of Economics

ngias@upatras.gr

December 10, 2019

Overview

Introduction

Modeling

Tobit

Censored data

- ▶ Dependent variable is bounded but continuous within the bounds:
 - ▶ corner solution models - Tobit
 - ▶ duration data
 - ▶ truncated regression.
- ▶ A corner solution response variable is bounded such that where b_L denotes the lower bound (limit) and b_H the higher bound. Bounds refer to real economic constraints.

$$b_L \leq y_i \leq b_H \quad (1)$$

- ▶ Usually: $b_L = 0$ (lower limit) and $b_H \rightarrow \infty$ (upper limit).
- ▶ Example: micro data (household expenditure on education, health, alcohol, or investment in capital goods among small entrepreneurial firms).

Corner solution models: Modeling

- ▶ Let y be a variable that is equal to zero for some non-zero proportion of the population, and that is continuous and positive if it is not equal to zero. We model y as a function of a set of variables x_1, x_2, \dots, x_k :

$$X = [a, x_1, x_2, \dots, x_k] \quad (2)$$

OLS

- ▶ OLS is a useful starting point for modeling corner solution variables:

$$y_i = a + X_i\beta + u_i \quad (3)$$

- ▶ As in the case of binary models OLS may not be an ideal estimator for corner solution models
 1. negative predicted values
 2. non-linearity
 3. heteroskedastic residuals
 4. non-normality in y

Tobit model

- ▶ Latent variable model:

$$y_i^* = a + X_i\beta + u_i \quad (4)$$

- ▶ Latent variable y_i^* is unobserved.
- ▶ Instead we observe:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \quad (5)$$

- ▶ which can also be written:

$$y_i = \max(y_i^*, 0) \quad (6)$$

Tobit model

$$y_i = \max(y_i^*, 0) \quad (7)$$

- ▶ First, y_i^* satisfies the classical linear model assumptions, so had y_i^* been observed the obvious choice of estimator would have been OLS.
- ▶ Second, it is often helpful to think of y as a variable that is bounded below for economic reasons, and y_i^* as a variable that reflects the 'desired' value if there were no constraints. Actual household expenditure on health is one example - this is bounded below at zero. In such a case y_i^* could be interpreted as desired expenditure, in which case $y_i^* < 0$ would reflect a desire to sell off ones personal (or family's) health.

Tobit model

- ▶ We said above that a corner solution variable is a kind of hybrid: both discrete and continuous. The discrete part is due to the piling up of observations at zero.
- ▶ The probability that y is equal to zero can be written:

$$\text{Prob}(y_i = 0|X) = 1 - \Phi\left(\frac{X_i\beta}{\sigma_u}\right) \quad (8)$$

- ▶ In contrast, if $y_i > 0$ then it is continuous:

$$y_i = a + X_i\beta + u_i \quad (9)$$

- ▶ Estimation technique: MLE
- ▶ In Stata: tobit