

# Mô hình Tobit với Biến Phụ thuộc bị chặn

Lê Việt Phú  
Chương trình Giảng dạy Kinh tế Fulbright

Ngày 26 tháng 12 năm 2015

# Table of contents

Khái niệm biến phụ thuộc bị chặn

Hồi quy OLS với biến phụ thuộc bị chặn

Mô hình Tobit với biến phụ thuộc bị chặn

Ước lượng mô hình Tobit

Thực hành trên STATA

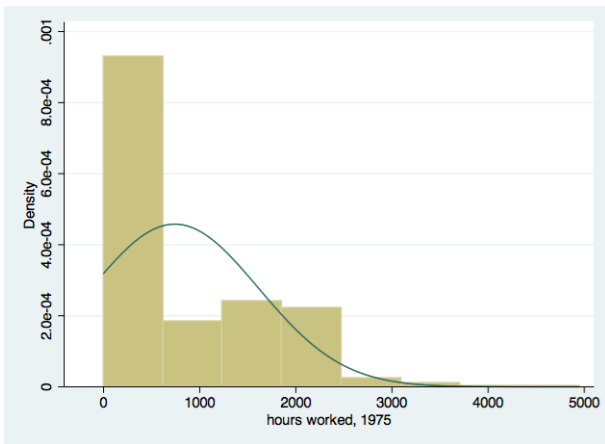
# Khái niệm biến phụ thuộc bị chặn

- ▶ Biến phụ thuộc nhận giá trị 0 đối với một số lượng lớn quan sát.
- ▶ Phần còn lại nhận giá trị dương. Ví dụ:
  - ▶ Số giờ lao động của phụ nữ đã có gia đình.
  - ▶ Số tiền làm từ thiện của một người trong một năm.
  - ▶ Số lít rượu bia một người uống trong một năm.
  - ▶ Chi tiêu cho hàng hoá xa xỉ của hộ gia đình trong dịp lễ tết.
  - ▶ Thời gian thất nghiệp của một người lao động.

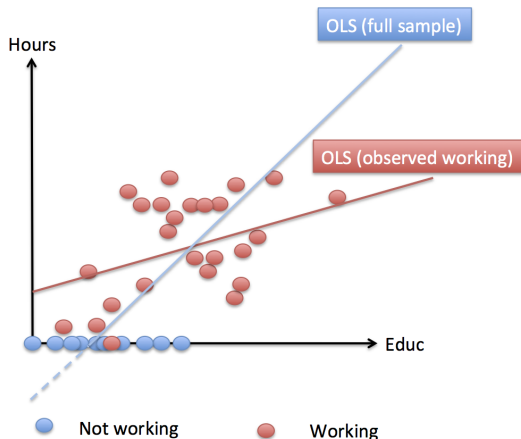
# Biểu đồ phân phối số giờ làm việc của phụ nữ có gia đình

Sử dụng bộ dữ liệu MROZ.dta

- hist hours, bin(8) normal



# Hồi quy OLS với biến phụ thuộc bị chặn



- ▶ Ước lượng OLS với nhóm phụ nữ làm việc bị thiên lệch giảm (downward bias) do bỏ qua nhóm không làm việc.
- ▶ Ước lượng OLS với toàn bộ dữ liệu gặp phải vấn đề dự báo số giờ lao động âm tương tự như mô hình xác suất tuyến tính LPM.

# Các cách xử lý biến phụ thuộc bị chặn

- ▶ Cách 1: ước lượng mô hình Logit/Probit với biến phụ thuộc là có làm việc hay không. Tuy nhiên cách làm này chỉ ước lượng được xác suất có làm việc hay không, nhưng không ước lượng được tác động của biến giải thích lên số giờ làm việc của những người đi làm như thế nào.
- ▶ Cách 2: mô hình Tobit xử lý được cả hai vấn đề trên.

# Mô hình Tobit với biến phụ thuộc bị chặn

Mô hình Tobit được diễn giải dưới dạng hàm ẩn:

$$y^* = X * \beta + u \quad (1)$$

với  $u \sim N(0, \sigma^2)$ .  $y^*$  không quan sát được. Ví dụ  $y^*$  có thể là độ thoả dụng của việc đi làm.

Chúng ta chỉ quan sát được  $y$  với điều kiện:

$$y = \begin{cases} y^* & \text{if } y^* > 0 \\ 0 & \text{if } y^* \leq 0 \end{cases}$$

# Mô hình Tobit với biến phụ thuộc bị chặn

Mô hình Tobit gồm có 2 bước:

- ▶ Bước 1: Ước lượng xác suất quan sát được một người có tham gia lao động hay không.
- ▶ Bước 2: Ước lượng các nhân tố ảnh hưởng đến số giờ lao động.

# Ước lượng mô hình Tobit - Bước 1

Xác định xác suất một quan sát có tham gia lao động hay không:

$$P(y > 0|x) = P(X * \beta + u > 0|X) = P(u > -X * \beta) \quad (2)$$

sau khi biến đổi:

$$P(y > 0|x) = P\left(\frac{u}{\sigma} > \frac{-X * \beta}{\sigma}\right) = 1 - \Phi\left(\frac{-X * \beta}{\sigma}\right) = \Phi\left(\frac{X * \beta}{\sigma}\right) \quad (3)$$

với  $\Phi(.)$  là hàm phân phối tích lũy chuẩn.

## Ước lượng mô hình Tobit - Bước 2

Tính tác động của các biến giải thích lên biến phụ thuộc:

$$E[y|x] = P(y > 0|x) * E[y|y > 0, x] + \underbrace{P(y = 0|x) * E[y = 0|x]}_{=0} \quad (4)$$

$$\Rightarrow E[y|x] = P(y > 0|x) * E[y|y > 0, x] = \Phi\left(\frac{X * \beta}{\sigma}\right) * E[y|y > 0, x] \quad (5)$$

với  $E[y|x]$  là kỳ vọng không điều kiện, và  $E[y|y > 0, x]$  là kỳ vọng với điều kiện biến phụ thuộc dương.

## Ước lượng mô hình Tobit - Bước 2

Bỏ qua một số biến đổi trung gian,

$$E[y|y > 0, x] = X * \beta + \sigma \lambda\left(\frac{X * \beta}{\sigma}\right) \quad (6)$$

và

$$E[y|x] = \Phi\left(\frac{X * \beta}{\sigma}\right) * \left[X * \beta + \sigma \lambda\left(\frac{X * \beta}{\sigma}\right)\right] \quad (7)$$

với  $\lambda(c) = \frac{\phi(c)}{\Phi(c)}$ , còn được gọi là tỷ số Mills nghịch đảo (inverse Mills ratio).

# Tác động biên trong mô hình Tobit

Từ công thức trên ta có thể tính được tác động biên của biến giải thích lên biến phụ thuộc bằng đạo hàm bậc nhất. Đối với kỳ vọng có điều kiện biến phụ thuộc dương:

$$\frac{\partial E[y|y > 0, x]}{\partial x_j} = \beta_j * \left\{ 1 - \lambda(c) * [c + \lambda(c)] \right\} \quad (8)$$

và đối với kỳ vọng không điều kiện:

$$\frac{\partial E[y|x]}{\partial x_j} = \beta_j * \Phi(c) \quad (9)$$

với  $c = \frac{X^*\beta}{\sigma}$  và  $\lambda(c) = \frac{\phi(c)}{\Phi(c)}$ .

## Tác động biên trong mô hình Tobit (2)

Ví dụ từ công thức (9):

$$\frac{\partial E[y|x]}{\partial x_j} = \beta_j * \Phi\left(\frac{X * \beta}{\sigma}\right)$$

- ▶ Tác động biên của mô hình Tobit được tính gián tiếp bằng  $\beta$  có điều chỉnh giảm theo hệ số  $P(y > 0|x) = \Phi\left(\frac{X*\beta}{\sigma}\right) < 1$ .
- ▶ Nếu  $P(y > 0|x) = \Phi\left(\frac{X*\beta}{\sigma}\right) = 1$  thì biến phụ thuộc nhận giá trị dương cho toàn bộ mẫu quan sát (tất cả phụ nữ đều làm việc). Khi đó OLS và Tobit là đồng nhất.
- ▶ Tương tự như phương pháp MLE,  $\Phi\left(\frac{X*\beta}{\sigma}\right)$  được tính tại các giá trị đặc trưng như trung bình, các tứ phân vị... của các biến giải thích.

# Thực hành trên STATA

Sử dụng bộ dữ liệu MROZ.dta và ước lượng hàm cung lao động của phụ nữ đã có gia đình. Trong bài trước chúng ta đã ước lượng các nhân tố ảnh hưởng đến xác suất tham gia lao động của phụ nữ bằng mô hình LPM, Logit và Probit.

```
tab inlf
```

```
reg hours nwifeinc educ exper expersq age kidslt6 kidsge6
```

```
eststo ols
```

```
tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll(0)
```

```
eststo tobit
```

```
esttab ols tobit
```

# Thực hành trên STATA

|                                  | OLS                  | TOBIT                |
|----------------------------------|----------------------|----------------------|
| nwifeinc                         | -3.447<br>(-1.35)    | -8.814*<br>(-1.98)   |
| educ                             | 28.76*<br>-2.22      | 80.65***<br>-3.74    |
| exper                            | 65.67***<br>-6.59    | 131.6***<br>-7.61    |
| expersq                          | -0.700*<br>(-2.16)   | -1.864***<br>(-3.47) |
| age                              | -30.51***<br>(-6.99) | -54.41***<br>(-7.33) |
| kidslt6                          | -442.1***<br>(-7.51) | -894.0***<br>(-7.99) |
| kidsge6                          | -32.78<br>(-1.41)    | -16.22<br>(-0.42)    |
| _cons                            | 1330.5***<br>-4.91   | 965.3*<br>-2.16      |
| sigma                            |                      |                      |
| _cons                            |                      | 1122.0***<br>-26.99  |
| N                                | 753                  | 753                  |
| t statistics in parentheses      |                      |                      |
| * p<0.05, ** p<0.01, *** p<0.001 |                      |                      |
|                                  |                      |                      |

## So sánh ước lượng OLS và Tobit thế nào?

Dựa vào công thức (8) và (9). Ví dụ đối với kỳ vọng không điều kiện (áp dụng cho toàn bộ các quan sát gồm cả những người đang đi làm và không đi làm), dựa vào công thức (9) để tính tác động biên trong mô hình Tobit:

$$\frac{\partial E[y|x]}{\partial x_j} = \beta_j * \Phi\left(\frac{X * \beta}{\sigma}\right)$$

Cần tính  $\Phi\left(\frac{X * \beta}{\sigma}\right)$  tại giá trị trung bình của các biến giải thích.

## So sánh ước lượng OLS và Tobit thế nào?

- ▶  $\Phi(20.12 * -8.81 + 12.28 * 80.65 + 10.63 * 131.6 + 10.63^2 * -1.86 + 42.53 * -54.41 + .24 * -894.0 + 1.35 * -16.22 + 965.3)/1122.02) = \Phi(.3727) = .645$ . (Dùng lệnh "di normal(.3727)" trong STATA)
- ▶ Tác động biên của việc học thêm một năm lên số giờ lao động của phụ nữ, tại giá trị trung bình của các biến giải thích, tính cho toàn bộ mẫu gồm cả những người đang tham gia lao động và không lao động theo công thức (9), là  $80.65 * .645 = 52$  giờ. Như vậy ước lượng Tobit cao hơn ước lượng OLS (28.76 giờ).
- ▶ Ước lượng nào hợp lý hơn, OLS hay Tobit?

# Tự thực hành

- ▶ Học viên tự thực hành tính tác động biên của việc học thêm một năm lên số giờ lao động của phụ nữ đối với những người hiện nay đã đi làm theo công thức (8).
- ▶ Chỉ giới hạn vào 428 phụ nữ đang tham gia lao động, so sánh ước lượng OLS và Tobit.

## Tài liệu tham khảo

Wooldridge, Jeffrey M. 2002. Introductory Econometrics: A Modern Approach, South-Western (2nd edition, Chapter 17).