

**SỬ DỤNG MÔ HÌNH TOBIT TRONG PHÂN TÍCH NHU CẦU  
TIÊU THỤ HÀNG HOÁ CỦA CÁC HỘ GIA ĐÌNH**

*Nguyễn Khắc Hoàn  
Trường Đại học Kinh tế, Đại học Huế*

**TÓM TẮT**

*Nhu cầu tiêu thụ một loại hàng hoá nào đó phụ thuộc vào nhiều nhân tố, chẳng hạn như thu nhập, giá cả, thị hiếu sở thích của người tiêu dùng... Thông thường, các dữ liệu quan sát về mức tiêu thụ hàng hoá của hộ gia đình cho thấy có một tỷ lệ đáng kể các quan sát về mức tiêu thụ hàng hoá của hộ gia đình trong một thời điểm nhất định là không. Điều này có nghĩa là biến phụ thuộc bị kiểm duyệt (censored). Phương pháp hồi qui thông thường sẽ thất bại trong việc xác định sự khác nhau về chất lượng giữa các quan sát giới hạn và các quan sát không giới hạn (liên tục) vì sự không đồng nhất và thiên lệch của nó. Do vậy, phương pháp hồi qui kiểm duyệt (TOBIT) được sử dụng trong nghiên cứu này là hữu ích.*

**1. Mở đầu**

Nhu cầu tiêu dùng hàng hoá của các hộ gia đình ngày càng đa dạng theo hướng nâng cao chất lượng và độ an toàn.

Nhu cầu tiêu thụ một loại sản phẩm hàng hoá nào đó phụ thuộc vào nhiều nhân tố khác nhau như: giá cả của chính hàng hoá đó, giá cả của hàng hoá thay thế và hàng hoá bổ trợ, thu nhập, thị hiếu, sở thích, số nhân khẩu trong một gia đình... Mỗi quan hệ này có thể được xác định bằng một quan hệ hàm số, đó là hàm số cầu, trong đó, biến phụ thuộc là lượng cầu về một hàng hoá nào đó và biến độc lập là các yếu tố về giá cả, thu nhập, thị hiếu, nhân khẩu...

Các số liệu điều tra về mức tiêu thụ của các hộ gia đình về một loại hàng hoá nào đó thường gặp khó khăn là: Số liệu được cung cấp thường chỉ trong một khoảng thời gian ngắn (khoảng một tuần trước thời điểm điều tra). Do vậy, sẽ có nhiều hộ gia đình không có số liệu về mức tiêu thụ hàng hoá đó, điều đó có nghĩa là một phần lớn giá trị của biến phụ thuộc sẽ nhận giá trị bằng 0.

Số liệu như thế sẽ không phản ánh đúng nhu cầu tiêu thụ của các hộ gia đình và dạng số liệu như thế thường thấy nhiều trong nghiên cứu nhu cầu tiêu thụ hàng hoá. Trong phân tích kinh tế lượng, dạng dữ liệu như vậy được gọi là số liệu bị kiểm lọc (censored data)

Phân tích nhu cầu tiêu thụ hàng hoá trong trường hợp này, người ta thường sử dụng một mô hình hàm cầu phản ánh mối quan hệ giữa số lượng hàng hoá tiêu thụ và các nhân tố ảnh hưởng đến nhu cầu.

Phương pháp Bình phương bé nhất (OLS) sẽ thất bại trong việc ước lượng mô hình dạng số liệu bị kiểm lọc (censored data). Mô hình TOBIT sẽ rất hữu hiệu trong trường hợp này với việc sử dụng phương pháp Maximum Likelihood (ML), cùng với sự trợ giúp của các phần mềm Shazam, Rats hoặc Limdep.

Đã có nhiều nghiên cứu về vấn đề này sử dụng mô hình TOBIT, chẳng hạn như:

- House purchase of durable goods (Tobin, 1958)
- The number of extramarital affairs (Fair, 1978)
- The number of hours worked by a women in laborforces (Greene, 1982)
- The number of arrests after release from prison (Witte, 1980)
- Household of expenditure on various commodity group (Jarque, 1987)

## 2. Thảo luận

Theo tác giả Greene.H (1982), Lý thuyết phân phối cho trường hợp biến bị kiểm lọc là giống với lý thuyết phân phối cho các biến bị chặt (Truncated variable).

Khi số liệu bị kiểm lọc thì phân phối của nó là sự trộn lẫn của phân phối rời rạc và phân phối liên tục [2,4].

Để phân tích phân phối này, ta xác định một biến ngẫu nhiên mới  $y$  được chuyển đổi từ một biến gốc  $y^*$  như sau:

$$y = 0 \text{ nếu } y^* \leq 0$$

$$y = y^* \text{ nếu } y^* > 0$$

Nếu  $y^* \sim N(\mu, \sigma^2)$  thì phân phối được áp dụng là:

$$\text{Prob}(y = 0) = \text{prob}(y^* \leq 0) = \Phi\left(-\frac{\mu}{\sigma}\right) = 1 - \Phi\left(-\frac{\mu}{\sigma}\right)$$

và nếu  $y^* > 0$  thì  $y$  có mật độ  $y^*$

Mô hình hồi qui dựa vào lý thuyết trên được coi như mô hình hồi qui được kiểm lọc (Censored Regression) hay mô hình Tobit. Dạng tổng quát được viết như sau [1,5]:

$$y_i^* = \beta'X_i + \varepsilon_i$$

$$y_i = 0 \text{ nếu } y_i^* \leq 0$$

$$y_i = y_i^* \text{ nếu } y_i^* > 0$$

Hàm trung bình có điều kiện của mô hình là:

$$E(y_i/x_i) = \phi\left(\frac{\beta X_i}{\sigma}\right)(\beta X_i + \sigma \lambda_i)$$

$$\text{Trong đó: } \lambda_i = \frac{\phi(\beta X_i / \sigma)}{\Phi(\beta X_i / \sigma)}$$

$\phi$  là hàm mật độ xác suất (PDF)

$\Phi$  là hàm mật độ lũy tích (CDF)

Ảnh hưởng biên (marginal effects) từ biến  $y^*$  là {6}

$$\frac{\partial E[y^* / x_i]}{\partial x_i} = \beta$$

Còn ảnh hưởng biên từ biến  $y$  là:

$$\frac{\partial E[y^* / x_i]}{\partial x_i} = \beta \cdot \Phi\left(\frac{\beta' X_i}{\sigma}\right)$$

Theo Berndt E.R (1980) và Greene H (1933), ảnh hưởng biên được cấu thành từ hai ảnh hưởng sau {3}:

$$\frac{\partial E(y_i / X_i)}{\partial x_i} = \text{Prob}(y^* > 0) \cdot \frac{\partial E(y^* / y^* > 0)}{\partial x_i} + E(y^* > 0) \cdot \frac{\partial \text{Prob}(y^* > 0)}{\partial x_i}$$

Một sự thay đổi trong  $X$  sẽ tạo ra hai ảnh hưởng đối với hàm trung bình có điều kiện của  $Y^* > 0$  và ảnh hưởng đến xác suất cho  $Y^* > 0$ .

Theo Maddala (1977), nếu chúng ta làm một vài giả thiết cụ thể về phân phối của  $U$  (Disturbance), chúng ta có thể sử dụng phương pháp MLE để ước lượng các thông số của mô hình. Chẳng hạn như chúng ta giả định rằng  $U$  có phân phối chuẩn với trung bình bằng 0 và varian thì phân phối xác suất đồng thời (Joint distribution) từ các quan sát là {8}:

$$\prod_1 \frac{1}{\delta \sqrt{2\pi}} \exp\left[-\frac{1}{2\delta^2}(y_i - \beta x_i)^2\right] \prod_2 \left\{ \int_{-\infty}^0 \frac{1}{\delta \sqrt{2\pi}} \exp\left[-\frac{1}{2\delta^2}(y_j - \beta_j)^2\right] \right\} dy_j$$

Trong đó, phần đầu của biểu thức đúng với các quan sát của  $y_i > 0$  và phần thứ hai đúng với các quan sát cho  $y_i \leq 0$ .

Một cách ngắn gọn biểu thức trên được viết như sau:

$$L = \prod_1 \frac{1}{\delta} f\left(\frac{y_i - \beta x_i}{\delta}\right) \prod_2 F\left(\frac{-\beta x_j}{\delta}\right)$$

Trong đó,  $f(\cdot)$  là hàm mật độ chuẩn và  $F(\cdot)$  là hàm mật độ chuẩn tích lũy

(Commulative normal density funsction)

Ước lượng ML với việc tối đa hóa L đối với  $\beta$  và  $\delta$ . Ở đây gặp phải vấn đề phi tuyến (non linear) và có thể giải bằng việc sử dụng hành trình tối đa hóa chẳng hạn như:

Cauchy's method, The Newton - Raphson method, Quadratic hill - Climbing method, Davidon - Fletche - Powell method. [5; 8]

Để ước lượng mô hình hồi qui bị kiểm lọc hay mô hình TOBIT được phát triển bởi James Tobin (1958), chúng ta không thể sử dụng phương pháp bình thường bé nhất (OLS) bởi vì không đáp ứng được điều kiện  $E(u) = 0$ . Ước lượng OLS sẽ Bias và Inconsistent đối với các thông số.

Một cách tốt nhất để ước lượng mô hình hồi qui TOBIT là phương pháp maximum likelihood (MLE) với sự trợ giúp của các phần mềm máy tính như SHAZAM, RATS và LIMDEP (Gujarati. D (1955)).

Mô hình ước lượng nói trên được sử dụng phương pháp MLE (the maximum likelihood estimation method) ở dạng logarit hóa như sau [6, 8]:

$$\ln L = \sum_{Q_i > 0} -\frac{1}{2} \left[ \ln(2\pi n + \ln \delta^2 + \frac{(Q_i - \beta b_i X_i)^2}{\delta^2}) \right] + \sum_{Q_i = 0} \ln \left[ 1 - \Phi \left( \frac{b_i X_i}{\delta} \right) \right]$$

Trong đó  $Q_i$  là nhu cầu hàng hóa tiêu thụ của hộ gia đình;  $b_i$  là hệ số hồi qui và  $X_i$  là các biến độc lập.

Phần thứ nhất phía bên phải của phương trình thể hiện phân phối cho các quan sát liên tục (non limit) và phần thứ hai phía bên trái của phương trình là xác suất cho các quan sát không liên tục (discrete).

Khi phân tích nhu cầu tiêu thụ một hàng hóa nào đó ta có thể thiết lập một mô hình hồi qui dạng mô hình TOBIT như sau [9]:

$$Q_{da}^* = a + b_1 P_1 + b_2 P_2 + b_3 P_3 + b_4 P_4 + \dots b_5 I + b_6 M + dD$$

$$Q_{da} = Q_{da}^* \text{ khi } Q_{da}^* > 0$$

$$Q_{da} = 0 \text{ khi } Q_{da}^* \leq 0$$

Trong đó:

$Q_{da}^*$  : Số lượng một hàng hóa mà gia đình tiêu thụ trong một tháng (kg/tháng)

$Q_{da}$  : Số lượng một hàng hóa mà gia đình thực sự tiêu thụ trong một tháng (kg/tháng)

$a$  : Hằng số

$b_i$  : ( $i = 1 \rightarrow 6$ ) : Hệ số hồi qui

$d$  : Hệ số của biến dummy:

P1: Giá của chính hàng hóa được nghiên cứu (1.000đ/kg)

P2 - P4: Giá của các hàng hóa khác có liên quan (1.000đ/kg)

I : Thu nhập của hộ gia đình (1.000đ/hộ/tháng)

M: Số thành viên trong gia đình (người/hộ)

D: Biến giả theo vùng D = 1 thành phố; D = 0 vùng nông thôn

Với việc sử dụng mô hình TOBIT được giải bằng phương pháp MLE sẽ cho các kết quả ước lượng tốt hơn nhiều so với phương pháp bình phương bé nhất OLS do đặc điểm của số liệu trong trường hợp bị kiểm lọc.

### 3. Kết luận

Kết quả ước lượng từ phương pháp MLE là tốt hơn kết quả ước lượng từ phương pháp OLS trong trường hợp số liệu điều tra có phân phối liên tục và rời rạc, biến phụ thuộc nhận một số giá trị 0.

Việc giải mô hình TOBIT bằng phương pháp MLE và sử dụng phần mềm LIMDEP cho kết quả rất đáng tin cậy.

### TÀI LIỆU THAM KHẢO

1. Bemdt, ER. *The Practice of Econometrics Classic and Contemporary*, Mc. Graw. Hill Book Company, 1980.
2. Chow. G.C. *Econometrics*, Mc. Graw Hill Book Company, 1983.
3. Draper N.R and Smith H.. *Applied Regression Analysis*, Second Edition by John Willy and Sons Inc, 1981.
4. Greene.W.H. *Econometric Analysis*, Macmilan Publishing Company, New York, 1993.
5. Gujarati. D. *Basic Econometrics*, Mc. Graw Inc, 1995.
6. Intriligator. MD. *Econometric Models, Techniques and Applications*, Prentice Hall, Inc, 1981.
7. Koutsoyiannis A.. *Theory of Econometric Introducing Exposition of Econometric Methods*. Happers Ron Publisher, Inc, 1976.
8. Madala G.S. *Econometrics*. Mc. Graw Hill Kogakusha LTD, 1977.
9. Nguyen Khac Hoan. *Analysis of chicken demand in Yogyakarta province, Indonesia by using TOBIT model*. A master thesis in economics, Gadjaja Mada University, 1999.
10. Pindyck R.S and Daniel L. *Econometric Models and Economic Forecast*. Third Edition, John Wiley & Sons, Inc, 1984.

# USE MODEL TOBIT FOR ANALYSIS OF COMMODITY CONSUMPTION DEMAND OF HOUSEHOLDS

Nguyen Khac Hoan  
College of Economics, Hue University

## SUMMARY

*The demand of consumption for a given commodity depends on many factors such as income, price, consumer's taste and so on. Data collected from observations of household consumption in a commodity actually showed a significant fraction of the observed households is zero in consumption. This means a dependent variable is censored. The conventional regression method may fail to account for the qualitative difference between limit (zero) observations and non limit (continuous) observations because of its inconsistent and biased. Therefore, the censored regression method (TOBIT) is used for this study.*