

## Các biến phụ thuộc bị giới hạn

Các bộ dữ liệu cần thiết cho việc minh họa: **Lfpmw.wf1 & Ldv1.wf1**

### Dữ liệu về các tỉ lệ tính chất trong mẫu

Đôi lúc chúng ta có dữ liệu về tỉ lệ các cá nhân có một tính chất nào đó. Nghĩa là, chúng ta không có dữ liệu mang các giá trị 0 hoặc 1 về một tính chất nào đó về các cá nhân, nhưng chúng ta lại có tỉ lệ các cá nhân có tính chất  $Y = 1$  trong mẫu. Những tỉ lệ như vậy thường được gọi là  $\hat{p}_i$ .

Đối với các dữ liệu như vậy, tỉ lệ này là một ước lượng xác suất trực tiếp. Tuy nhiên, dạng mô hình thích hợp kết hợp  $p$  với một tập hợp các biến hồi qui vẫn cần thỏa mãn những ràng buộc của xác suất.

Nhớ rằng mô hình logit có dạng:

$$\Pr(Y_i = 1 | X_i \beta) = p_i = \frac{\exp(\beta_1 + \beta_2 X_{2i} + \dots + \beta_{K_i})}{1 + \exp(\beta_1 + \beta_2 X_{2i} + \dots + \beta_{K_i})}$$

Bằng động tác việc tuyến tính hoá dạng logarit nói trên chúng ta có:

$$\ln \left[ \frac{p_i}{1 - p_i} \right] = \beta_1 + \beta_2 X_{2i} + \dots + \beta_{K_i}$$

Biểu thức này trở thành một mô hình kinh tế lượng sẵn sàng cho việc ước lượng nếu chúng ta có dữ liệu về những xác suất này:

$$\ln \left[ \frac{\hat{p}_i}{1 - \hat{p}_i} \right] = \beta_1 + \beta_2 X_{2i} + \dots + \beta_{K_i} + \varepsilon_i$$

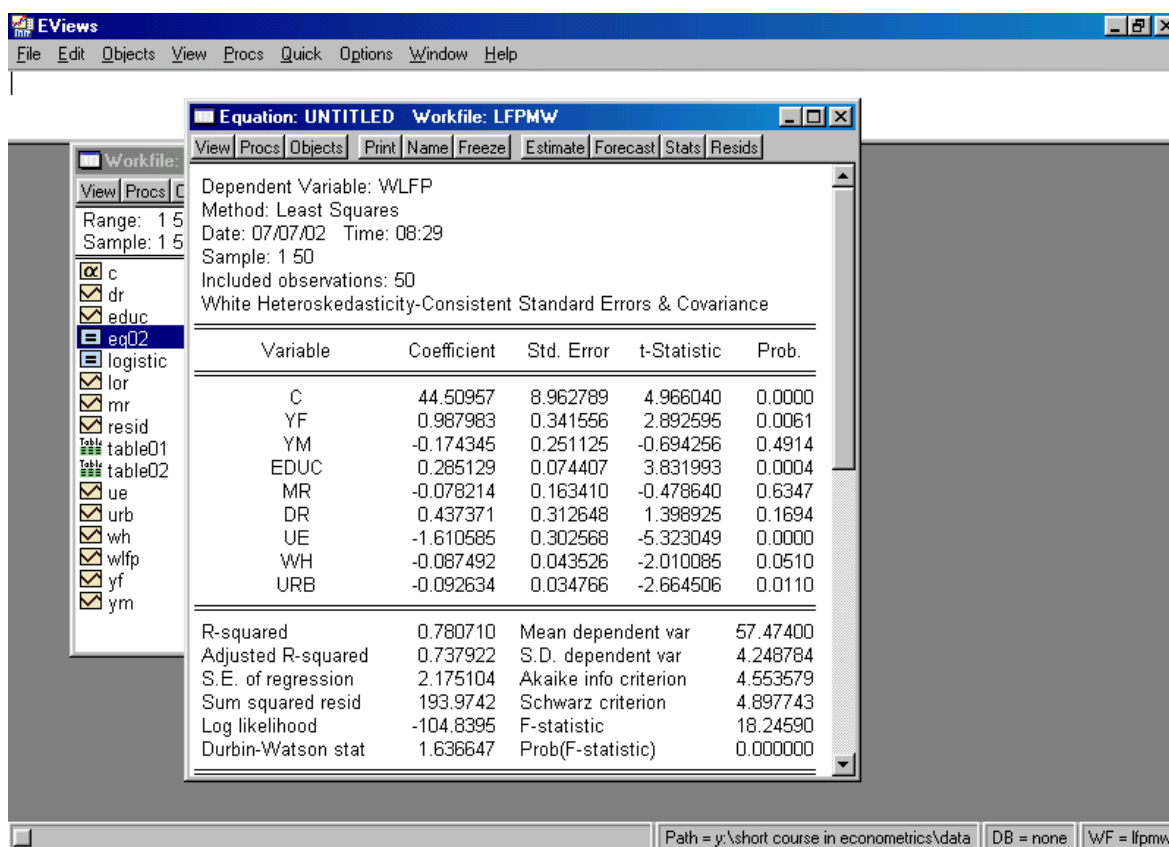
Chúng ta dùng ví dụ về sự tham gia lực lượng lao động của phụ nữ có gia đình có trong sách giáo khoa của Ramanathan :

Các biến sử dụng là các trung bình cho mỗi tiểu bang trong 50 tiểu bang của Mỹ và thành phố Hoa thịnh đốn.

- WLFP = Phần trăm phụ nữ tham gia vào lực lượng lao động.
- YF = Thu nhập trung vị của phụ nữ năm 1989 (1000 đô la).
- YM = Thu nhập trung vị của nam giới năm 1989 (1000 đô la).
- EDUC = Phần trăm của phụ nữ có bằng trung học hoặc cao hơn.
- UE = Tỷ lệ thất nghiệp .
- MR = Phần trăm phụ nữ có gia đình.
- DR = Phần trăm phụ nữ đã li hôn.
- URB = Phần trăm phụ nữ sống ở thành thị.
- WH = Phần trăm phụ nữ là người da trắng.

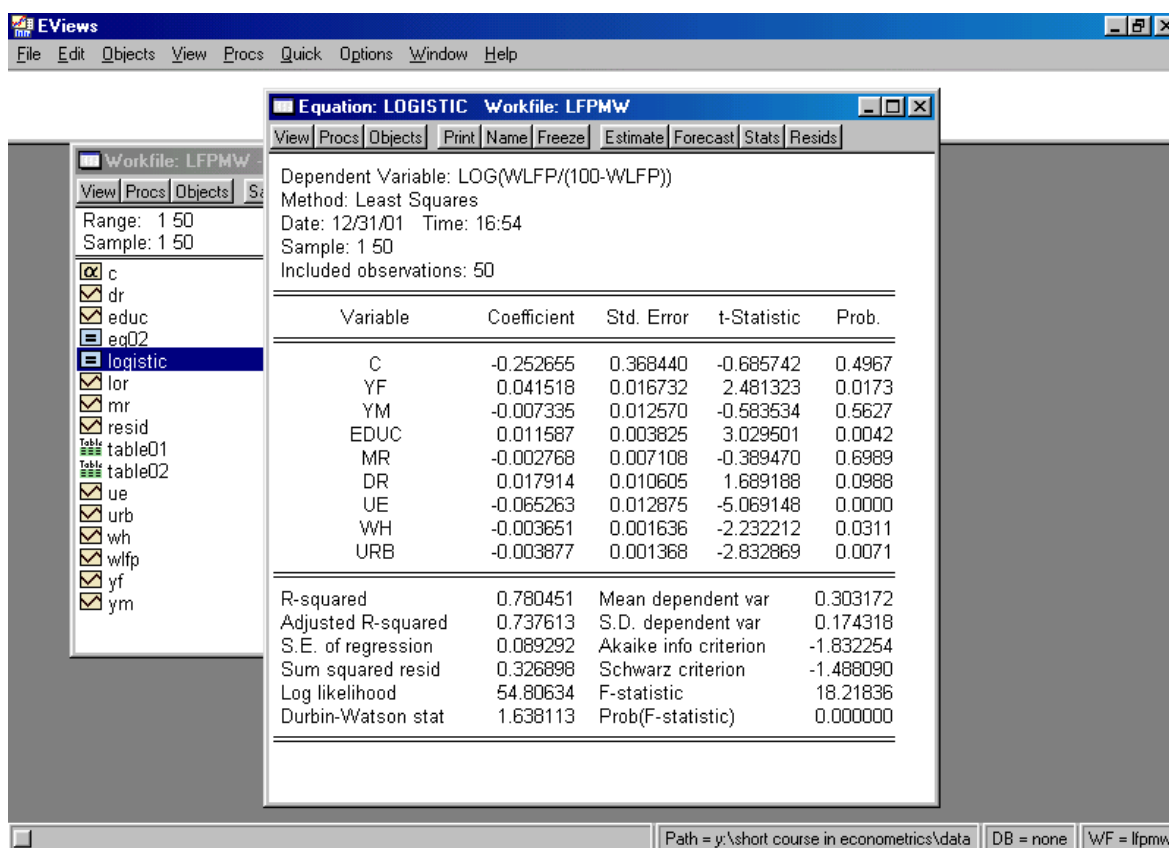
Chúng ta có thể ước lượng mô hình này bằng kỹ thuật hồi qui tuyến tính bội. Tuy nhiên, chúng ta biết là trong Bài giảng 17 thì thành phần sai số có khả năng sẽ có phương sai không đồng nhất, vì vậy chúng ta sử dụng phương pháp HCCM của White.

Đây là bảng kết quả của phương pháp HCCM của White:



Còn bây giờ hãy ước lượng bằng hàm logistic. Chúng ta có thể đưa trực tiếp việc biến đổi thành phương trình ước lượng này. Nhớ dùng cách ước lượng HCCM của White.

Xem kết quả ở trang sau:



Nhờ là không thể so sánh trực tiếp những kết quả của hai mô hình này. Để so sánh chúng một cách hợp lý, chúng ta sẽ sử dụng những kết quả hàm Logistic để tính các giá trị ước lượng của xác suất, các tác động biên v.v..., giống như cách thức mà chúng ta đã làm trước đây.

### Mô hình Tobit (các biến phụ thuộc đã kiểm qua kiểm tra)

Xét tình huống là Anh/Chị muốn xây dựng mô hình chi tiêu của hộ gia đình là một hàm số phụ thuộc vào các biến như giá cả hàng hóa, thu nhập của hộ gia đình, và các biến liên quan khác.

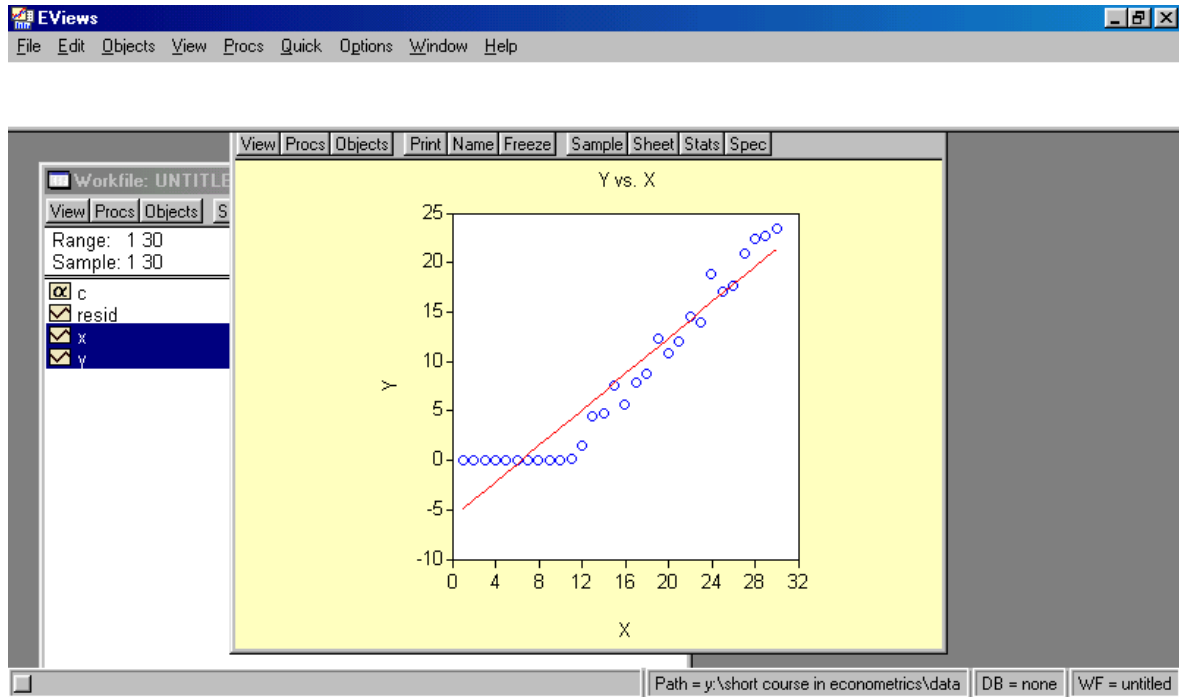
Dữ liệu của Anh/Chị bao gồm các loại quan sát sau:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + \varepsilon_i \quad \text{neu } Y_i > 0$$

$$Y_i = 0 \quad \text{neu không phải như trên}$$

Theo nghĩa nào đó, chúng ta có thể nghĩ tới việc ra quyết định chi tiêu sẽ tiến hành theo hai bước : quyết định có mua hàng này hay không, và quyết định chi bao nhiêu cho hàng hóa đó.

Xem biểu đồ phân tán kết hợp với đường hồi qui sau:



Rõ ràng là đường hồi qui không thích hợp tốt với dữ liệu. Chúng ta có thể đơn giản bỏ đi các quan sát có giá trị bằng không, nhưng các nhà kinh tế lượng lại không thích loại bỏ dữ liệu. Giải pháp lý tưởng sẽ là xây dựng mô hình kết hợp quyết định mua hàng và quyết định chi tiêu. Kỹ thuật Tobit giải quyết điều này.

Dữ liệu trên được hình thành như sau:

$$X = 1, 2, 3, \dots, 30$$

$$Y = -12 + 1,2 * X + @RNORM$$

khi đó Y được cho bằng không nếu giá trị tính ra có giá trị  $Y < 0$ .

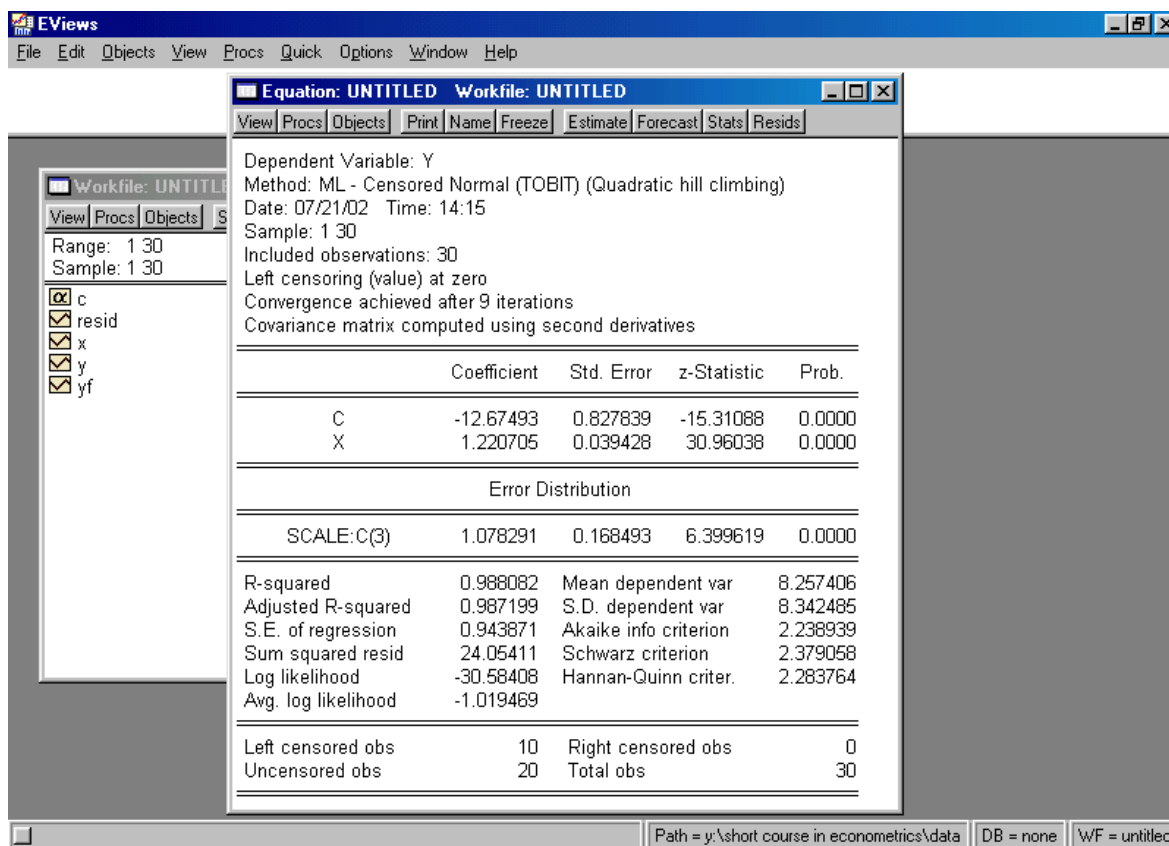
Để tạo mô hình Tobit hãy nhấp chuột theo cách dưới đây

Quick / Estimate Equation

Sau đó chúng ta vào mục Estimation Settings và nhấp chuột vào menu trái ra phía dưới có tên Method chúng ta cuộn xuống dưới và chọn

CENSORED --- Censored Data (Tobit)

Cửa sổ nhận dạng phương trình có một số lựa chọn cho công việc to mĩ hơn , nhưng đơn giản chúng ta chỉ đánh vào nhận dạng phương trình của chúng ta như thường lệ. Bảng kết quả sau khi chúng ta làm xong các thao tác được cho dưới đây:

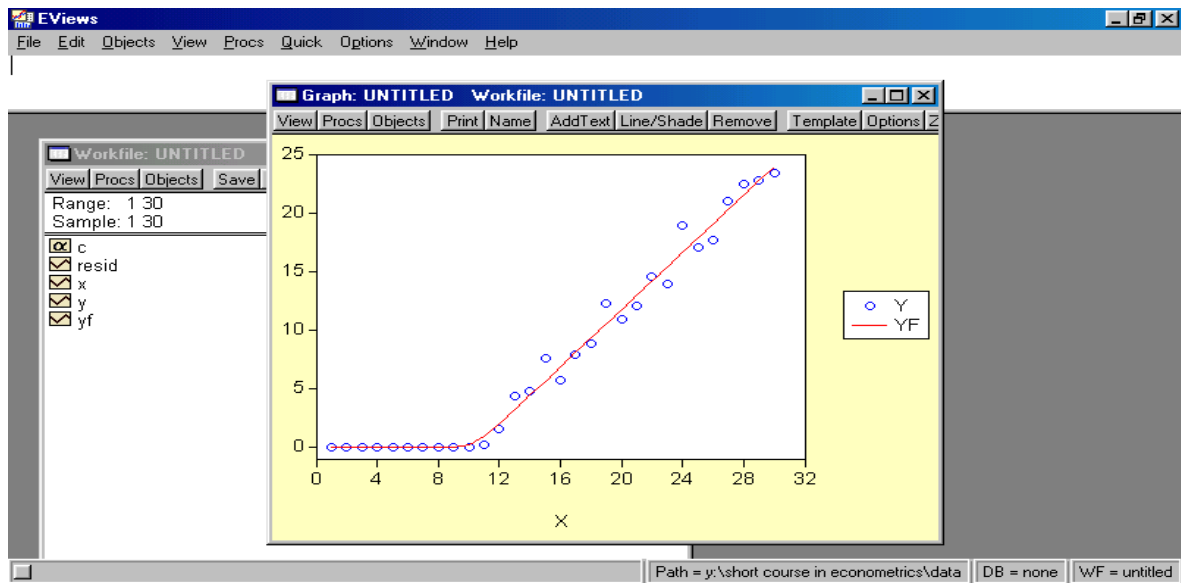


Điều thú vị là so sánh giá trị dự báo này với dữ liệu thực tế. Để làm điều này chúng ta nhấp chuột vào **Forecast** và để cho EViews đặt tên cho chuỗi giá trị dự báo YF.

Bây giờ hãy nhấp chuột theo trình tự sau:

**Quick / Graph / XY Line**

**X Y YF**



Bài tập 18 cho chúng ta một cơ hội sử dụng dữ liệu thực mà nó có đặc trưng đã nêu trên. VLSS có một biến về chi tiêu của hộ gia đình cho giáo dục và một biến chi tiêu của hộ gia đình cho thuốc lá.

Chi tiêu cho giáo dục được gán giá trị là 0 cho 1876 hộ gia đình .  
Các chi tiêu cho giáo dục được gán giá trị lớn hơn 0 cho 4123 hộ gia đình .

Các chi tiêu cho thuốc lá được gán giá trị là 0 cho 2113 hộ gia đình .  
Các chi tiêu cho thuốc lá được gán giá trị lớn hơn 0 cho 3886 hộ gia đình.