CHƯƠNG 2: HỒI QUI ĐƠN BIẾN

Ở bài trước, ta nêu lên ví dụ về mối quan hệ giữa khối lượng và trọng lượng của các mẫu nước. Dựa trên việc lấy các mẫu thử $\{x_n, y_n\}_{n=1}^{N}$, chúng ta có thể **ước lượng**, hay tìm lại mối quan hệ tuyến tính $Y = \alpha + \beta X$, mà nó thể hiện quy luật vật lý, hay tính xu thế, ổn định giữa hai đại lượng ngẫu nhiên là trọng lượng và khối lượng nước.

Trong chương này, chúng ta sẽ giới thiệu việc ước lượng các quy luật tự nhiên, kinh tế, hay xã hội kiểu như vậy thông qua phương pháp **hồi quy đơn** (**simple regression**). Chúng ta sẽ sử dụng học thuyết Keynes về tiêu dùng như là ví dụ điển hình cho việc giới thiệu phương pháp xây dựng và ước lượng mô hình hồi qui đơn biến.

2.1 Học thuyết Keynes về tiêu dùng

Chúng ta hãy trích định luật sau, nêu ra bởi Keynes (1936) trong *Lý thuyết tổng quát* (*general Theory*) của ông:

Chúng ta sẽ xác định quy luật mà ta gọi là khuynh hướng tiêu dùng theo thu nhập như là một mối quan hệ phụ thuộc f giữa X, được gọi là mức thu nhập khả dụng, và Y là chi tiêu cho tiêu dùng từ thu nhập đó, và vì vậy: $Y = f(X)$.

- Số tiền mà từng hộ gia đình chi tiêu cho tiêu dùng phụ thuộc (i) một phần vào thu nhập của hộ đó, (ii) vào những yếu tố khách quan khác của hoàn cảnh sống, và (iii) một phần vào đòi hỏi có tính thiết yếu, thói quen và những yếu tố tâm lý của các cá nhân trong hộ gia đình đó….

- Luật tâm sinh lý cơ bản mà chúng ta dựa vào một cách rất tin cậy, được kiểm chứng bởi tri thức của chúng ta về loài người, và bởi kinh nghiệm, rằng con người có xu hướng tăng tiêu dùng khi thu nhập của họ tăng, nhưng tăng không nhanh bằng thu nhập. Tức là $\frac{dY}{dX}$ là dương và nhỏ hơn 1.

- Về trung bình, nếu thu nhập tăng lên thì khoảng cách giữa tiêu dùng và thu nhập ngày càng mở rộng, nghĩa là có một tỉ lệ lớn hơn trong thu nhập được đưa vào tiết kiệm khi thu nhập tăng lên.

Lý thuyết của Keynes đã đặt một mối quan hệ ổn định giữa tiêu dùng và thu nhập $Y = f(X)$. Chúng ta muốn xác định cụ thể mối quan hệ này là như thế nào, tìm cách đo lường quan hệ đó, và kiểm định lại tính đích thức của học thuyết Keynes.

2.2 Cơ sở vi mô cho học thuyết Keynes về tiêu dùng

Gọi X là mức thu nhập dùng để chi cho tiêu dùng và tiết kiệm (nhằm tăng tiêu dùng cho tương lai). Gọi Y là mức tiêu dùng hiện tại; và S là tiêu dùng trong tương lai.

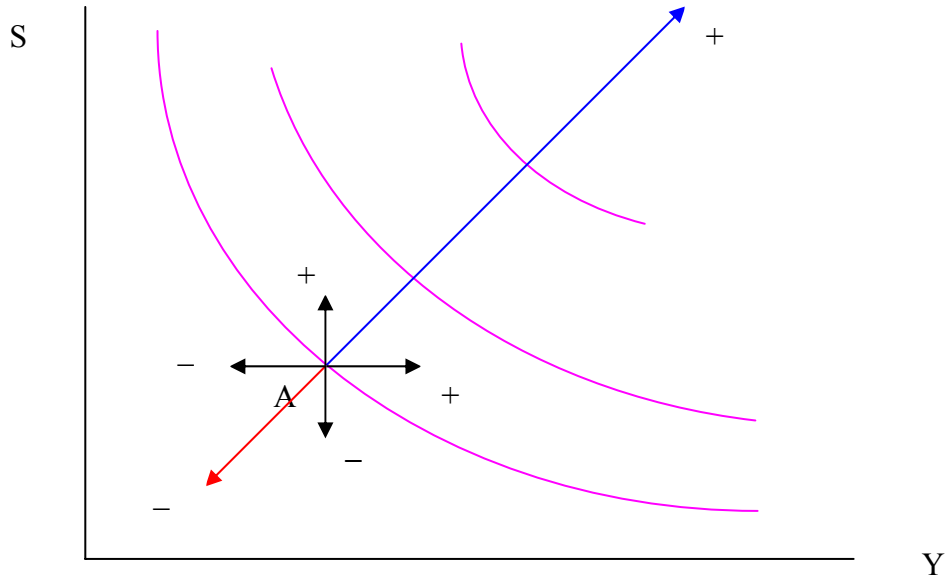
Khi đó, ta có **ràng buộc ngân sách** (budget constraint):

$$Y + \frac{1}{1+r} S = X \qquad (2.1)$$

Thành phần thứ hai trong vế trái $\frac{1}{1+r} S$ là khoản tiết kiệm. Nó thể hiện giá trị hiện tại (present value) của thu nhập cho tiêu dùng trong tương lai *S,* được chiết khấu bởi $\frac{1}{1+r}$. Trong đó, r là lãi suất tiền gửi tiết kiệm.

Về thực chất, 1 đồng tiền ngày hôm nay có thể sinh ra $(1+r)$ đồng thu nhập cho tiêu dùng ngày mai, nếu được gửi vào tiết kiệm. Vì vậy, 1 đồng tiền tiêu trong tương lai chỉ có giá bằng $\frac{1}{1+r}$ đồng tiền ngày hôm nay. Đó chính là khái niệm về **hệ số chiết khấu** (*discount rate*). Nó thể hiện rằng, nếu tiêu dùng bị trì hoãn đi tới một thời điểm trong tương lai, thì nó không thể có giá trị bằng việc được tiêu dùng ngay lập tức vào ngày hôm nay.

Tiếp theo, chúng ta hãy đo lường mức độ thỏa dụng của cá nhân với các lựa chọn khác nhau về tiêu dùng cho hiện tại và cho tương lai (Y, S).



Đồ thị 2.1: Đường bàng quan (indifference curve)

Trong đồ thị 2.1, điểm A thể hiện mức thỏa dụng hiện tại của cá nhân ứng với mức tiêu dùng tại điểm đó. Giả sử có một sự gia tăng về tiêu dùng hiện tại, trong khi tiêu dùng trong

tương lai vẫn giữ nguyên. Khi đó ta dịch chuyển từ điểm A sang bên phải và song song với trục hoành ($\rightarrow^+$). Dấu cộngthể hiện rằng độ thỏa dụng của cá nhân được nâng lên.

Ngược lại, giả sử ta giữ nguyên mức tiêu dùng hiện tại, nhưng tiêu dùng tương lai được tăng lên ($\uparrow^+$). Khi đó, sự cảm nhận về an toàn của cá nhân về cuộc sống tương lai cũng tăng, tức là độ thỏa dụng của cá nhân đó tăng.

Vì vậy, ¼ không gian, được xác định bởi sự gia tăng của tiêu dùng hiện tại ($\rightarrow^+$), hoặc tiêu dùng trong tương lai ($\uparrow^+$), hoặc sự gia tăng đồng thời của cả hai yếu tố đó, thể hiện độ thỏa dụng ngày càng tăng lên (+). Cá nhân cảm thấy giàu lên, sung sướng và an toàn hơn về vật chất.

Phân tích tương tự cho trường hợp ngược lại, khi độ thỏa dụng ngày càng giảm (-).

Trong ngắn hạn, mức thu nhập là không đổi. Do đó, sự gia tăng mức tiêu dùng hiện tại thường phải bị đánh đổi (hay trả giá) bằng việc giảm tiêu dùng trong tương lai. Tuy nhiên, cá nhân chỉ làm sự đánh đổi như vậy một khi độ thỏa dụng mới ít ra là không kém đi so với trạng thái đã có. Trong kinh tế học vi mô, người ta thể hiện các lựa chọn như vậy bằng đường bàng quan **(***indifference curve***)**. Nó có chiều dốc xuống mô tả sự đánh đổi. Nghĩa là, nếu muốn tăng mức tiêu dùng trong hiện tại thì phải giảm mức tiêu dùng trong tương lai, sao cho lợi ích hay độ thỏa dụng vẫn giữ nguyên.

Bây giờ, hãy đưa đường ràng buộc ngân sách vào đồ thị 2.1. Điểm tiếp xúc giữa đường ràng buộc ngân sách với đường bàng quan thể hiện sự lựa chọn tốt nhất của cá nhân về tiêu dùng ứng với mỗi mức thu nhập [xem đồ thị 2.2].

Ví dụ 2.1: Giả sử thu nhập (X) và tiêu dùng (Y) của 3 cá nhân có giá trị cụ thể như sau:

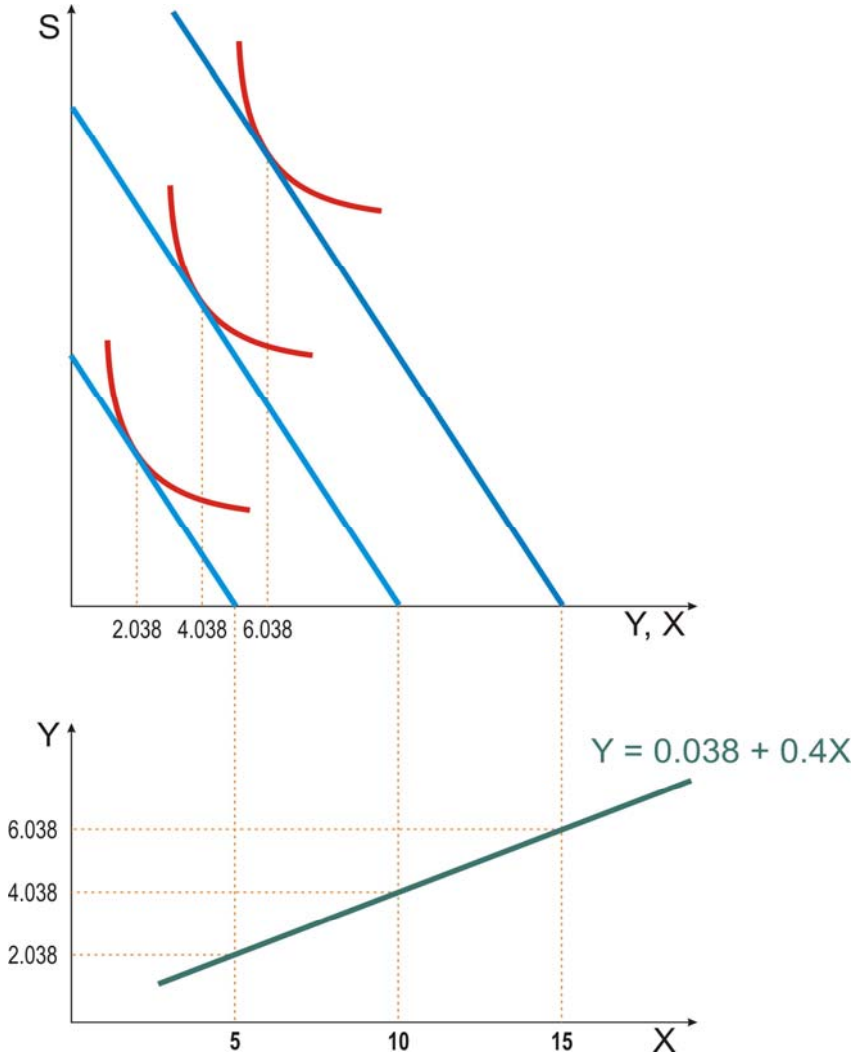
***X* [thu nhập]**	***Y* [tiêu dùng]**
5	*2.038*
10	*4.038*
15	*6.038*

Bảng 2.1: Quan hệ giữa thu nhập và tiêu dùng

Sử dụng phương pháp phân tích nêu trên, chúng ta có thể biểu diễn sự lựa chọn của mỗi cá nhân như sau:

Trong đồ thị 2.2, hình vẽ thứ nhất, ta thể hiện sự lựa chọn của cá nhân về tiêu dùng ứng với mỗi mức thu nhập khả dụng. Khi họ có 5 triệu đồng thu nhập, họ giành cho tiêu dùng hiện tại Y là 2.038 triệu. Phần còn lại được đưa vào tiêu dùng trong tương lai S. Tương tự cho các mức tiêu dùng 4.08 và 6.038 ứng với các mức thu nhập khác là 10 và 15 triệu.

Tiếp theo, trong hình vẽ thứ hai, ta chỉ ra mối quan hệ giữa tiêu dùng hiện tại Y với từng mức thu nhập khả dụng X. Đó chính là mối quan hệ cơ bản, mô tả bởi học thuyết Keynes về tiêu dùng.



Đồ thị 2.2: Sự lựa chọn tiêu dùng theo thu nhập của cá nhân.

Như chỉ ra trên hình vẽ thứ hai, quan hệ giữa tiêu dùng và thu nhập: $Y = f(X)$, là mối quan hệ tuyến tính. Trong ví dụ vừa nêu, quan hệ đó có dạng cụ thể là:

$$Y = 0.038 + 0.40\ X$$

Ý nghĩa của phương trình này như sau:

- Nếu X = 0 thì Y = 0.038, điều này có nghĩa rằng người không có thu nhập vẫn tiêu dùng ở mức tối thiểu là 0.038 triệu đồng một tháng.
- Hệ số 0.40 (hay khuynh hướng tiêu dùng theo thu nhập) cho biết, nếu thu nhập tăng lên 1 triệu thì tiêu dùng tăng lên 0.40 triệu. Tức là, mức tăng tiêu dùng không nhanh bằng mức tăng thu nhập.

- Về trung bình, khi thu nhập tăng thì tỉ lệ giữa thu nhập và tiêu dùng (X/Y) ngày càng giảm: $\frac{2.038}{5} > \frac{4.038}{10} > \frac{6.038}{15}$. Điều đó kiểm chứng lại điều mà Keynes nói là, có một tỷ lệ lớn hơn của thu nhập được đưa vào tiết kiệm khi người ta giàu lên.

Kết quả nghiên cứu trên phù hợp với những nhận định của Keynes về tiêu dùng.

Một cách tổng quát, dạng hàm mô tả tốt nhất khuynh hướng tiêu dùng theo thu nhập của Keynes có dạng tuyến tính:

$$Y = \alpha + \beta X \qquad (\alpha > 0, \beta \in (0,1)) \qquad (2.2)$$

Như đã chỉ ra qua ví dụ, dạng hàm này thỏa mãn mọi nhận định của Keynes về tiêu dùng.

Bây giờ, chúng ta hãy sử dụng các dữ liệu điều tra thực tế để nghiên cứu về nhu cầu tiêu dùng theo thu nhập thông qua lăng kính của học thuyết Keynes.

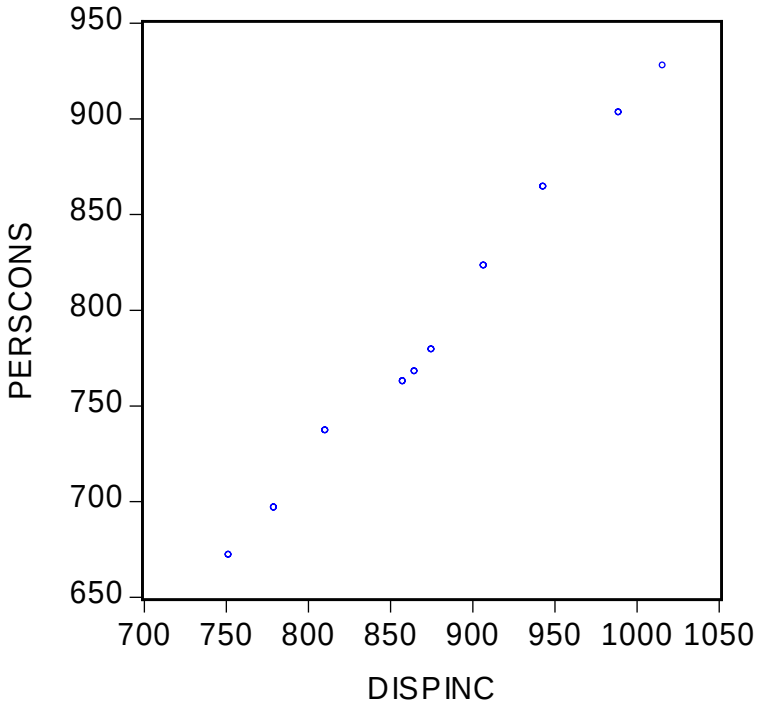
Ví dụ 2.2: Số liệu về tiêu dùng trung bình (PERCONS) và thu nhập khả dụng (DISPINC) theo giá cố định theo năm 1972 của nền kinh tế Mỹ trong 10 năm 1970 – 1979:

ĐVT: tỷ dollars

Năm	DISPINC	PERSCONS
1970	751.6	672.1
1971	779.2	696.8
1972	810.3	737.1
1973	864.7	767.9
1974	857.5	762.8
1975	874.9	779.4
1976	906.8	823.1
1977	942.9	864.3
1978	988.8	903.2
1979	1015.7	927.6

Bảng 2.2: Số liệu gộp về thu nhập và tiêu dùng tại Mỹ (1970-79)
(Nguồn: *Economic Report of the President*)

Đồ thị mô tả mối quan hệ giữa thu nhập và tiêu dùng của Mỹ được chỉ ra dưới đây:



Đồ thị 2.3: Mối quan hệ giữa thu nhập và tiêu dùng của nền kinh tế Mỹ từ 1970 đến 1979.

Mặc dù dữ liệu xem ra thể hiện khá tốt qui luật tuyến tính nêu ở trên nhưng rõ ràng mối quan hệ có tính xác định đó là **không đủ** để mô tả thực tiễn, vì còn rất nhiều yếu tố khác ảnh hưởng đến tiêu dùng (giới tính, tuổi tác, tâm lý,…).

Nói chung, chúng ta không có tham vọng đưa hết tất cả mọi yếu tố ảnh hưởng tới tiêu dùng vào mô hình, mà chỉ những yếu tố quan trọng, thiết yếu nhất.

Vì vậy, để có thể biểu diễn qui luật tiêu dùng trên thế giới thực, ta cần đưa thêm vào mô hình tuyến tính (2.2) một thành phần khác nữa, mang tính ngẫu nhiên, thể hiện sự tác động tổng gộp của các nhân tố nhỏ, không ổn định, tới tiêu dùng. Tức là, những yếu tố làm cho quan sát thật về tiêu dùng và thu nhập bị lệch khỏi xu thế ổn định, tuyến tính (2.2) nêu trên. Tức là, ta muốn biểu diễn mối quan hệ thực giữa các cặp dữ liệu quan sát được về thu nhập và tiêu dùng $\{x_n, y_n\}_{n=1}^{N}$ như sau:

$$y_n = \alpha + \beta x_n + \varepsilon_n, \qquad n = 1,2,3...,N. \tag{2.3}$$

Trong đó, $(X,Y) = (x_n, y_n)$: tiêu dùng và thu nhập thực tế của mẫu quan sát thứ *n*. Xét vế phải của phương trình (2.3), thành phần thứ nhất, $\alpha + \beta x_n$, là quy luật xác định [*deterministic part*], mà ta cần ước lượng; phần thứ hai, ε_n, là nhiễu [*random part*]. (Tức là, ε_n bao gồm sự tác động tổng hợp của mọi yếu tố khác của hoàn cảnh, có tính ngẫu nhiên, làm quan sát bị lệch khỏi khuynh hướng, hay qui luật ổn định). Cả hai phần này – tính xu thế, xác định; *và* yếu tố ngẫu nhiên - được gộp lại trong phương trình (2.3) để mô tả lý thuyết tiêu dùng của Keynes.

Do tác động của yếu tố ngẫu nhiên, trên đồ thị 2.3, chúng ta không quan sát thấy một đường thẳng thể hiện mối quan hệ tuyến tính $Y = \alpha + \beta X$ giữa tiêu dùng và thu nhập, như trên đồ thị 2.2 với số liệu giả định. Với dữ liệu điều tra thực tế, ta chỉ thấy một đám mây dữ liệu, dường như đang "bám" xung quanh một xu thế nào đó mà ta muốn ước lượng.

Ví dụ 2.3: Dữ liệu điều tra 44 nhân khẩu của nhóm gồm 5 sinh viên K04 khoa Kinh tế về thu nhập và tiêu dùng đầu người hộ gia đình tại TP HCM, Bình Dương, Thủ Dầu Một, Bà Rịa - Vũng Tàu, Mỹ Tho, và Nghệ An được ghi lại như sau[1]:

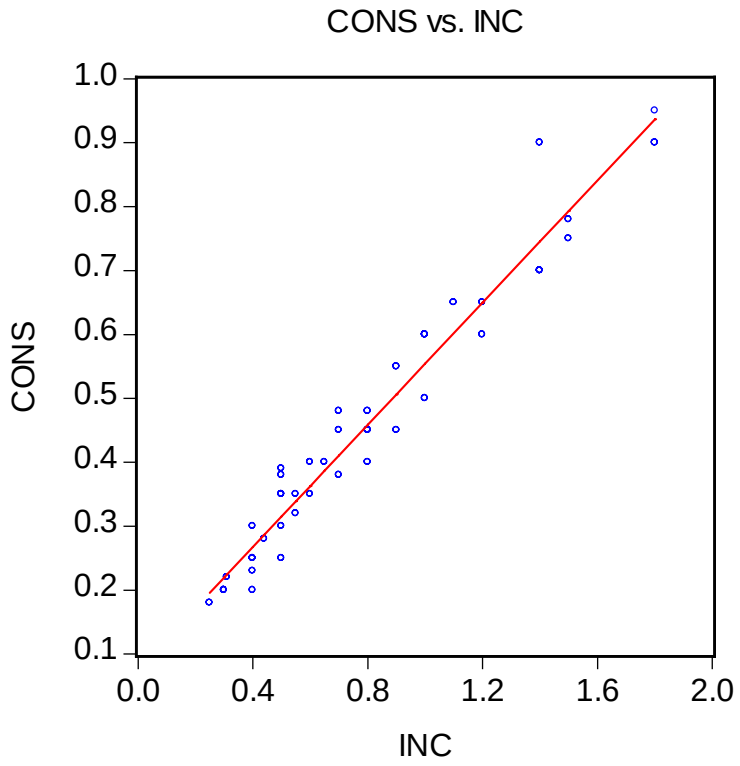
ĐVT: triệu đồng

Obs	INC	CONS	Obs	INC	CONS
1	1 .00	0.60	**23**	0.50	0.35
2	1.10	0.65	**24**	0.70	0.38
3	0.70	0.48	**25**	0.40	0.20
4	1.40	0.90	**26**	0.55	0.35
5	0.50	0.38	**27**	0.50	0.35
6	0.40	0.23	**28**	0.90	0.55
7	0.55	0.32	**29**	0.40	0.30
8	0.80	0.48	**30**	0.31	0.22
9	0.70	0.45	**31**	1 .20	0.65
10	0.25	0.18	**32**	0.60	0.40
11	0.65	0.40	**33**	0.30	0.20
12	0.40	0.25	**34**	0.80	0.40
13	1.80	0.95	**35**	0.44	0.28
14	0.40	0.25	**36**	0.50	0.39
15	0.50	0.30	**37**	1.00	0.60
16	0.30	0.20	**38**	1.80	0.90
17	1.00	0.50	**39**	1.40	0.70
18	0.50	0.25	**40**	1.50	0.75
19	0.80	0.45	**41**	1 .20	0.60
20	1.40	0.70	**42**	0.80	0.45
21	0.80	0.45	**43**	0.90	0.45
22	0.60	0.35	**44**	1.50	0.78

Bảng 2.3: Điều tra về thu nhập và tiêu dùng đầu người hộ gia đình tại một số tỉnh Việt nam

(*Ghi chú*: INC và CONS là thu nhập và tiêu dùng đầu người, đơn vị triệu đồng, tính tại thời điểm tháng 6, 2006.)

[1] Trưởng nhóm nghiên cứu này có mã số sinh viên là K 04 406 0975



Đồ thị 2.4: Thu nhập và tiêu dùng đầu người hộ gia đình tại một số tỉnh ở Việt Nam, năm 2006.

Như chỉ ra trên đồ thị, dữ liệu điều tra về tiêu dùng và thu nhập đầu người của hộ gia đình Việt nam tại một số tỉnh được điều tra cho thấy học thuyết tiêu dùng của Keynes phản ánh khá đúng về quy luật tiêu dùng của hộ gia đình tại các địa phương này.

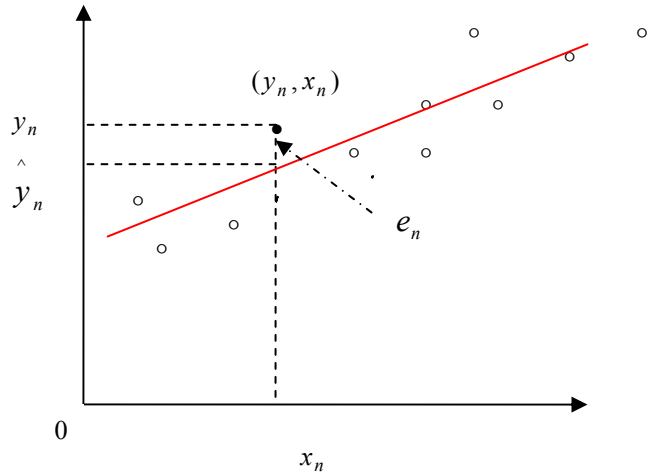
Bước tiếp sau là chúng ta hãy sử dụng những dữ liệu quan sát được này để xác định trở lại các tham số α, β trong mô hình hồi quy tuyến tính (2.2) và (2.3).

2.3. Ước lượng qui luật tiêu dùng:

Ta hãy vẽ các cặp quan sát về thu nhập và tiêu dùng $\{x_n, y_n\}_{n=1}^{N}$ lên đồ thị. Giả sử vạch đỏ trên đồ thị 2.5 dưới đây mô tả đường ước lượng quy luật tiêu dùng theo thu nhập. Nói khác đi, ta muốn ước lượng xu thế tiêu dùng bằng qui luật **tuyến tính**:

$$\hat{y}_n = \hat{\alpha} + \hat{\beta} x_n \qquad (2.4)$$

Trong đó, $\hat{y}_n$ là *ước lượng* về tiêu dùng, khi cho trước quan sát thu nhập x_n. Tương ứng, $\hat{\alpha}, \hat{\beta}$: các tham số ước lượng của các tham số tổng thể, chưa biết α, β.



Đồ thị 2.5: Ước lượng quy luật tiêu dùng qua các quan sát $(x_n, y_n), n = \overline{1, N}$

Mức độ tốt của việc ước lượng có thể được đo lường qua số dư (*residual*):

$$e_n = y_n - \hat{y} \tag{2.5}$$

Như đã nói, y_n là giá trị quan sát thực tế về tiêu dùng ứng với thu nhập x_n. Và $\hat{y}_n$: giá trị ước lượng về tiêu dùng.

Về mặt toán học, ta có thể viết tổng bình phương của sai số ước lượng (2.5) như sau:

$$\sum_n e_n^2 = \sum_n (y_n - \hat{y}_n)^2 \tag{2.6}$$

Sử dụng quan hệ (2.4), ta viết lại tổng bình phương sai số [*error sum of squares*], ký hiệu là ESS, ghi trong (2.6) như sau:

$$\sum_n e_n^2 = \sum_n (y_n - \hat{\alpha} - \hat{\beta} x_n)^2 \tag{2.7}$$

Một cách tự nhiên, chúng ta muốn rằng tổng bình phương sai số phần dư là nhỏ nhất**.** Vì vậy phương pháp có tên gọi là **bình phương cực tiểu** [*Least Squares*]:

$$S(\hat{\alpha},\hat{\beta}) = \sum_n (y_n - \hat{\alpha} - \hat{\beta} x_n)^2 \to \min \qquad (2.8)$$

Lưu ý rằng ở bài toán (2.8), chúng ta muốn chọn các tham số ước lượng $\hat{\alpha},\hat{\beta}$ sao cho tổng bình phương các sai số ước lượng, ESS, là nhỏ nhất.

Sử dụng điều kiện tìm điểm cực trị, (first order condition, FOC), chúng ta thấy rằng:

$$\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial \hat{\alpha}} = 2\sum_n (y_n - \hat{\alpha} - \hat{\beta}\cdot x_n)(-1) = 0 \qquad (2.10)$$

$$\frac{\partial S(\hat{\alpha},\hat{\beta})}{\partial \hat{\beta}} = 2\sum_n (y_n - \hat{\alpha} - \hat{\beta} x_n)(-x_n) = 0 \qquad (2.11)$$

Từ (2.10) ta có:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\cdot\bar{x} \Rightarrow \bar{y} = \hat{\alpha} + \hat{\beta}\cdot\bar{x} \qquad (2.12)$$

Nói khác đi, điểm $(\bar{x},\bar{y})$ nằm trên đường hồi qui $\hat{y}_n = \hat{\alpha} + \hat{\beta} x_n$.

Tiếp theo, từ phương trình (2.11), ta cũng có:

$$\sum_n y_n x_n = \hat{\alpha}\sum_n x_n + \hat{\beta}\sum_n x_n^2$$

Thay thế $\hat{\alpha} = \bar{y} - \hat{\beta}\cdot\bar{x}$ trong (2.12) vào biểu thức trên, sắp xếp lại các vế, ta tìm ra:

$$\sum_n (y_n - \bar{y}) x_n = \hat{\beta}\sum_n (x_n^2 - n\cdot\bar{x}^2)$$

Hay cũng vậy,

$$\hat{\beta} = \frac{\sum_n (x_n - \bar{x})(y_n - \bar{y})}{\sum_n (x_n - \bar{x})^2} \qquad (2.13)$$

Tóm lại, kết quả ước lượng $\hat{\alpha}, \hat{\beta}$ theo phương pháp bình phương cực tiểu như sau:

$$\hat{\alpha} = \bar{y} - \hat{\beta} \cdot \bar{x}$$
$$\hat{\beta} = \frac{\sum_n (x_n - \bar{x}) \cdot (y_n - \bar{y})}{\sum_n (x_n - \bar{x})^2} = \frac{S_{XY}}{S_{XX}} \qquad (2.14)$$

Trong đó, S_{XY} là Covariance mẫu, và S_{XX} : Variance mẫu của X.

2.5 Đo lường mức độ phù hợp của Ước lượng

Công thức (2.14) thể hiện hai điều: Thứ nhất, đường hồi quy đi qua điểm trung bình $(\bar{x}, \bar{y})$. Thứ hai, hệ số góc $\hat{\beta}$ là covariance mẫu của X và Y, cho phép đánh giá những biến động trong thu nhập X có tác động thế nào tới biến động trong tiêu dùng Y. Nếu mô hình phân tích và dự báo là tốt, thì một sự tăng (giảm) mạnh của thu nhập so với trung bình sẽ dẫn tới một sự tăng (giảm) mạnh tương của tiêu dùng so với trung bình.

Câu hỏi đặt ra là: liệu ta có thể sử dụng mô hình ước lượng để dự báo không? Liệu sự giao động của thu nhập so với trung bình $(x - \bar{x})$ có phải là dự đoán tốt cho sự giao động của tiêu dùng so với trung bình $(y - \bar{y})$ hay không?

Hãy lấy một quan sát cụ thể về tiêu dùng và thu nhập (x_n, y_n). Khi đó, sự khác biệt của thu nhập cá nhân thứ *n* so với trung bình $(y_n - \bar{y})$ có thể được viết lại như sau:

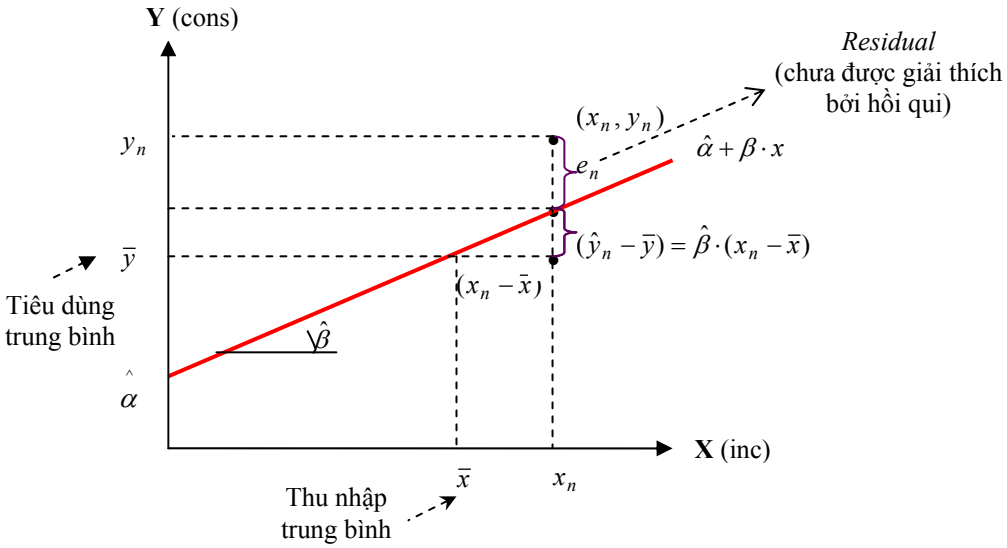
$$(y_n - \bar{y}) = \hat{y}_n - \bar{y} + e_n \qquad (2.15)$$

Hay cũng vậy,

$$(y_n - \bar{y}) = \hat{\beta}(x_n - \bar{x}) + e_n \qquad (2.16)$$

Vế trái là giao động của tiêu dùng so với mức trung bình; thành phần thứ nhất của vế phải là phần mà giao động đó đã được giải thích bởi mô hình hồi quy; và phần cuối cùng là sai

số ước lượng, thể hiện những giao động trong tiêu dùng chưa được giải thích bởi mô hình. Nói khác đi, đó là sai số dự báo từ mô hình.



Đồ thị 2.6: Phân tách các thành phần của $(y_n - \bar{y})$

Sử dụng các điều kiện tìm cực trị FOC (2.10) và (2.11), quan hệ (2.15) có thể viết lại như sau:

$$\sum_n (y_n - \bar{y})^2 = \sum_n (\hat{y}_n - \bar{y})^2 + \sum_n e_n^2 \qquad (2.17)$$

Vế trái là tổng bình phương các giao động trong tiêu dùng, ký hiệu là TSS (*total sum of squares*). Vế phải phân ra thành tổng bình phương phần đã được giải thích bởi mô hình hồi quy RSS (*regression sum of squares*), cộng với tổng sai số ước lượng ESS (*error sum of squares*).

Nói khác đi, ta có:

$$\text{TSS} = \text{RSS} + \text{ESS} \qquad (2.18)$$

Vì vậy,

$$1 = \frac{RSS}{TSS} + \frac{ESS}{TSS}$$

Hay cũng thế,

$$\frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} \qquad (2.19)$$

Vế phải của (2.19) được ký hiệu là $R^2 = 1 - \frac{ESS}{TSS}$. Ta thấy $0 \le R^2 \le 1$.

Ví dụ 2.4: Ước lượng khuynh hướng tiêu dùng cho một số tỉnh thành ở Việt Nam, sử dụng dữ liệu điều tra trong Ví dụ 2.3.

Kết quả ước lượng theo phương pháp bình phương cực tiểu được ghi lại dưới đây [các tham số $\hat{\alpha}, \hat{\beta}$ được tính theo công thức (2.14), và hệ số đo lường mức phù hợp R^2 theo (2.19)]:

Dependent Variable: CONS
Method: Least Squares
Date: 06/24/06 Time: 21:39
Sample: 1 44
Included observations: 44
Weighting series: INC

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.038254	0.004082	9.370437	0.0000
INC/HHSIZE	0.401771	0.014340	28.01749	0.0000

Weighted Statistics

R-squared	0.989838	Mean dependent var	0.137000
Adjusted R-squared	0.989596	S.D. dependent var	0.083538
S.E. of regression	0.008521	Akaike info criterion	-6.648174
Sum squared resid	0.003050	Schwarz criterion	-6.567075
Log likelihood	148.2598	F-statistic	784.9795
Durbin-Watson stat	2.221397	Prob(F-statistic)	0.000000

Bảng 2.4: Kết quả hồi quy mô hình tiêu dùng với dữ liệu điều tra tại Việt nam

Để có một hình dung rõ ràng về độ tốt của mô hình, ta dùng 40 quan sát đầu tiên để ước lượng mô hình. Sau đó dùng 4 quan sát cuối để kiểm tra độ tốt của dự báo(*ex post forecasting*). Kết quả dự báo cho 4 mẫu quan sát cuối cùng trong dữ liệu điều tra là như sau:

Obs	CONS	CONSF
41	0.600000	0.638548
42	0.450000	0.438385
43	0.450000	0.478911
44	0.780000	0.798185

Bảng 2.5: Kết quả dự báo

Ở đây, CONS là dữ liệu thu thập được về tiêu dùng của mẫu quan sát, [tương ứng với ký hiệu $y_n, n = 41,..,44.$]; và CONSF là kết quả dự báo từ mô hình, [tương ứng với ký hiệu $\hat{y}_n, n = 41,..,44.$] Như đã thấy, kết quả dự báo là khá phù hợp với dữ liệu thực có được từ điều tra.

Trong mục tiếp sau, chúng ta sẽ đánh giá độ tốt của ước lượng theo các tiêu chuẩn thống kê.