KINH TẾ LƯỢNG
(Econometrics)
Cu nhan: Nguyen Thanh Hai

Tel: 0918.738.043

TÀI LIỆU THAM KHẢO

1. Giáo trình KINH TẾ LƯỢNG, Ths. Hoàng Ngọc Nhậm (Chủ biên), NXB Lao động – Xã hội, 2008
2. Kinh tế lượng ứng dụng, Ths. Phạm Trí Cao – Ths. Vũ Minh Châu, NXB Thống kê, TP. HCM, 2009
3. Bài tập Kinh tế lượng, Ths. Hoàng Ngọc Nhậm (Chủ biên),

Chương 1

KHÁI QUÁT VỀ KINH TẾ LƯỢNG

TRANG 1

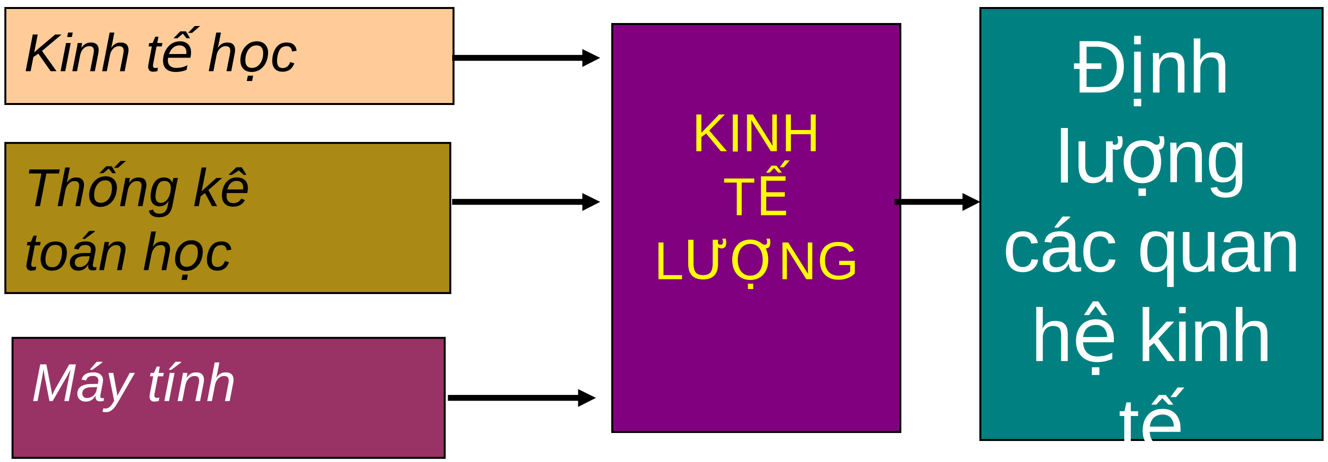
I.TỔNG QUAN

■ Năm 1936, Tinbergen trình bày trước Hội đồng kinh tế Hà lan một mô hình toán đầu tiên để phân tích khả năng cân gằng ngoại thương của Hà Lan trước tình hình Đại suy thoái kinh thế giới. Đây là lần đầu một loại mô hình mới được giới thiệu, bao gồm các phương trình và đẳng thức với các tham số được ước lượng .

■ Năm 1939 Tinbergen xây dựng một mô hình kinh tế lượng cho nước Mỹ. Sau đó kinh tế lượng phát triển lan ra khắp thế giới. Năm 1950, nhà kinh tế Mỹ được giải thưởng Nobel, Laurance Klein đưa ra mô hình Klein. Ông là chủ tịch danh dự của LINK PROJECT là dự án dự báo kinh tế thế giới thường niên của LHQ, với 2 Trung tâm xử lý dữ liệu và chạy mô hình với quy mô thế giới - Trung tâm Pensynvania (Mỹ) và trung tâm Toronto (Canada).»

1. Kinh tế lượng là gì?

KTL phát triển dựa trên kiến thức của 3 lĩnh vực: Kinh tế học, Thống kê toán học và Máy tính.



Dữ liệu trong KTL là dữ liệu thực tế trong sản xuất kinh doanh, trong quản lý kinh tế, khác với dữ liệu trong thống kê là do thí nghiệm

KTL có rất nhiều phần mềm chuyên dụng. Do tiện dụng và hiệu quả cao nên sẽ thực hành trên phần mềm EVIEWS 5.0. EVIEWS 5.0 chạy trong môi trường Windows nên có thể trao đổi dữ liệu và kết xuất kết quả dễ dàng sang các khuôn dạng khác như EXCEL, Word.

Mục đích KTL?

Kinh tế lượng nhằm:

- *Định lượng các quan hệ KT*
- *Dự báo kinh tế*
- *Phân tích chính sách*

(1) Từ số liệu kinh tế ước lượng các tham số mô hình, định lượng các quan hệ kinh tế

(2) Từ mô hình dự báo cho thời gian tiếp theo

(3) Từ mô hình mô phỏng phản ứng của các chính sách

«

2. Mô hình kinh tế và mô hình kinh tế lượng

So sánh:

$$Q = c_0 - c_1 P \qquad (1)$$

$$Q = c_0 - c_1 P + \varepsilon \qquad (2)$$

Mô hinh (1) mô tả quy luật nhu cầu. Nhu cầu số lượng hàng hóa Q phụ thuộc vào giá hàng hóa P. Giá P tăng, Q giảm.Quan hệ giữa Q và P là **chính xác hoàn toàn**

Mô hình (2) cũng phản ảnh quy luật nhu cầu nhưng quan hệ giữa Q và P **không chính xác hoàn toàn** mà có sai số ε phụ thuộc vào giá trị P và Q cụ thể quan sát được.

Mô hình (1) là mô hình kinh tế nói chung, mô hình (2) là mô hình kinh tế lượng. Mô hình KTL **ước lượng** từ các số liệu lấy **mẫu** từ thực tế nên luôn có sai số ngẫu nhiên, còn mô hình kinh tế chỉ cho biết quy luật chung

CAÙC BÖÔÙC XAÂY DÖÏNG VAØ AÙP DUÏNG KINH TEÁ LÖÔÏNG

1. Neâu vaán ñeà lyù thuyeát caàn phaân tích vaø caùc giaû thuyeát
2. Thieát laäp MH toaùn hoïc

3 **Thu thaäp soá lieäu**

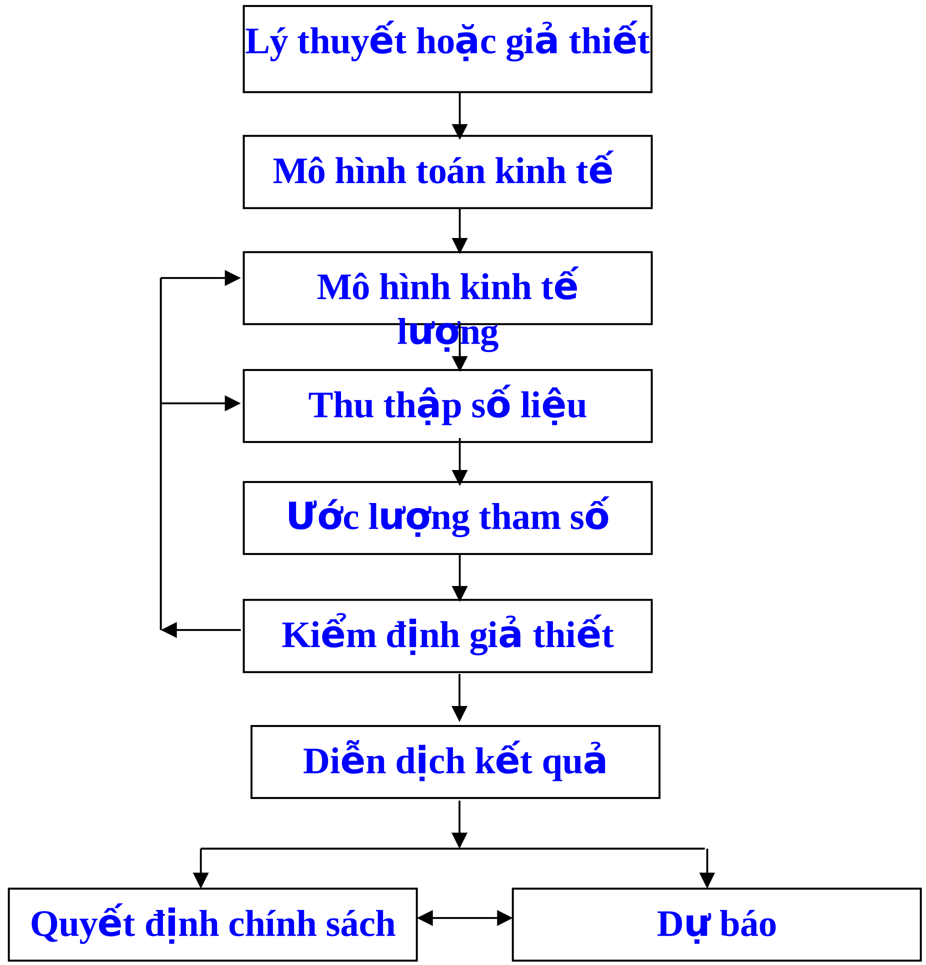
4 **Öôùc löôïng caùc tham soá**

5 **Phaân tích keát quaû**

6 **Döï baùo**

7 **Ra quyeát ñònh**

SƠ ĐỒ CÁC BƯỚC ÁP DỤNG



VÍ DỤ

PHÂN TÍCH TÁC ĐỘNG CỦA THU NHẬP LÊN TIÊU DÙNG TẠI CÁC QUỐC GIA VÙNG ĐÔNG Á – THÁI BÌNH DƯƠNG NĂM 1998

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 1: PHÁT BIỂU LÝ THUYẾT

Keynes cho rằng:

Theo Qui luật tâm lý cơ sở, con người thường sẽ tăng tiêu dùng khi thu nhập của họ tăng lên, nhưng không nhiều như là gia tăng của thu nhập. (2)

Vậy Keynes cho rằng xu hướng tiêu dùng biên (marginal propensity to consume-MPC), tức tiêu dùng tăng lên khi thu nhập tăng 1 đơn vị tiền tệ, lớn hơn 0 nhưng nhỏ hơn 1, tức là 0 < MPC < 1

(2) *John Maynard Keynes, 1936, theo D.N.Gujarati, Basic Economics, 3rd, 1995, trang 3.*

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 2: MÔ HÌNH TOÁN

Dạng hàm đơn giản nhất thể hiện mối quan hệ giữa tiêu dùng và thu nhập, theo Keynes, là dạng hàm tuyến tính.

$$\mathbf{TD = \beta_1 + \beta_2 TN}$$

Trong đó β_1, β_2 là các tham số và $0 < \beta_2 < 1$.

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 3:

XÂY DỰNG MÔ HÌNH KINH TẾ LƯỢNG

Quan hệ đúng giữa TD và TN như sau

$$\mathbf{TD = \beta_1 + \beta_2 TN + u_i}$$

Trong đó u_i là sai số

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 4: THU THẬP SỐ LIỆU

ĐVT: tỷ USD

Quốc gia	Tiêu dùng	Thu nhập	Quốc gia	Tiêu dùng	Thu nhập
Australia	289.35	372.72	Macao	3.3185	6.4474
Cambodia	2.7132	2.8709	Malaysia	37.344	72.488
China	560.53	946.31	Mongolia	0.76041	1.0417
Fiji	1.3677	1.5774	New Zealand	42.507	52.944
Hong Kong	113.88	162.94	Papua New Guinea	2.9644	3.8208
Indonesia	62.779	98.827	Philippines	57.088	65.535
Japan	2715.3	3808.1	Singapore	40.911	82.773
Korea, Rep.	208.48	317.08	Thailand	73.261	112.09
Lao PDR	0.94699	1.2609	Vietnam	21.443	27.184

Nguồn: World Development Indicators 2001, WB.

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 5: ƯỚC LƯỢNG HỆ SỐ

Để ước lượng các hệ số hồi quy, chúng ta sử dụng phương pháp bình phương tối thiểu thông thường (Ordinary Least Squares) và thu được kết quả hồi quy như sau:

$$TD = -6,27 + 0,709TN + u_i$$

$$t \quad [-0,859] \quad [90,58]$$

$$R^2 = 0,999$$

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 6: KIỂM ĐỊNH GIẢ THIẾT

Với kết quả hồi quy như trên:

Hãy kiểm định lý thuyết tiêu dùng biên của Keynes:

$$0 < \beta_2 < 1.$$

THU NHẬP VÀ TIÊU DÙNG

BƯỚC 7: DIỄN GIẢI KẾT QUẢ

Với kết quả hồi quy như sau:

$$TD = -6{,}27 + 0{,}709TN + u_i$$

$$t \quad [-0{,}859] \quad [90{,}58]$$

Tiêu dùng tự định của các quốc gia này là -6,27 tỷ USD?

Hệ số tiêu dùng biên của các quốc gia trong khu vực này là 0,709, tức là tiêu dùng tăng 0,709 tỷ USD nếu thu nhập tăng 1 tỷ USD.

THU NHẬP VÀ TIÊU DÙNG

<u>BƯỚC 8:</u> DỰ BÁO VÀ PHÂN TÍCH CHÍNH SÁCH

Dự báo: Giả sử với mức thu nhập là 100 tỷ USD, thì dự báo về chi tiêu như thế nào?

TD = -6,27 + 0,709*(100) = 64,63 (tỷ USD)

Phân tích chính sách: Giả sử chính phủ một quốc gia tính được mức chi tiêu trung bình ứng với một tỷ lệ thất nghiệp thích hợp. Tìm mức thu nhập cần thiết?

*** Dữ liệu cho nghiên cứu kinh tế lượng**

- **Dữ liệu chéo**: bao gồm quan sát cho nhiều đơn vị kinh tế ở một thời điểm cho trước.
- **Dữ liệu chuỗi thời gian**: bao gồm các quan sát trên một đơn vị kinh tế cho trước tại nhiều thời điểm.
- **Dữ liệu bảng:** là sự kết hợp giữa dữ liệu chéo và dữ liệu chuỗi thời gian.

* **Lượng biến rời rạc hay liên tục**

- **Lượng biến rời rạc** là một lượng biến có tập hợp các kết quả có thể đếm được, chiếm 1 vị trí trên trục số.
- **Lượng biến liên tục** là một lượng biến nhận kết quả một số vô hạn các kết quả, chiếm 1 khoảng trên trục số.

Chương 2

HỒI QUY 2 BIẾN

2.1. Giới thiệu

2.1.1. Khái niệm về hồi quy

Phân tích hồi qui là nghiên cứu sự phụ thuộc của 1 biến (biến phụ thuộc) vào 1 hay nhiều biến khác (biến độc lập), nhằm mục đích ước lượng (hay dự đoán) giá trị trung bình của biến phụ thuộc trên cơ sở các giá trị biết trước của các biến độc lập.

2.1.2. Sự khác nhau giữa các dạng quan hệ

- **Quan hệ thống kê và quan hệ hàm số:**

 Y = aX + b

 Năng suất lúa = f(nhiệt độ, lượng nắng, mưa, phân bón…)

- **Hồi quy và quan hệ nhân quả:**

Phân tích hồi quy không đòi hỏi giữa biến phụ thuộc và các biến độc lập phải có mối quan hệ nhân quả.

- **Hồi quy và tương quan:**

  - Phân tích tương quan là đo mức độ tuyến tính giữa hai biến; không có sự phân biệt giữa các biến; các biến có tính chất đối xứng.

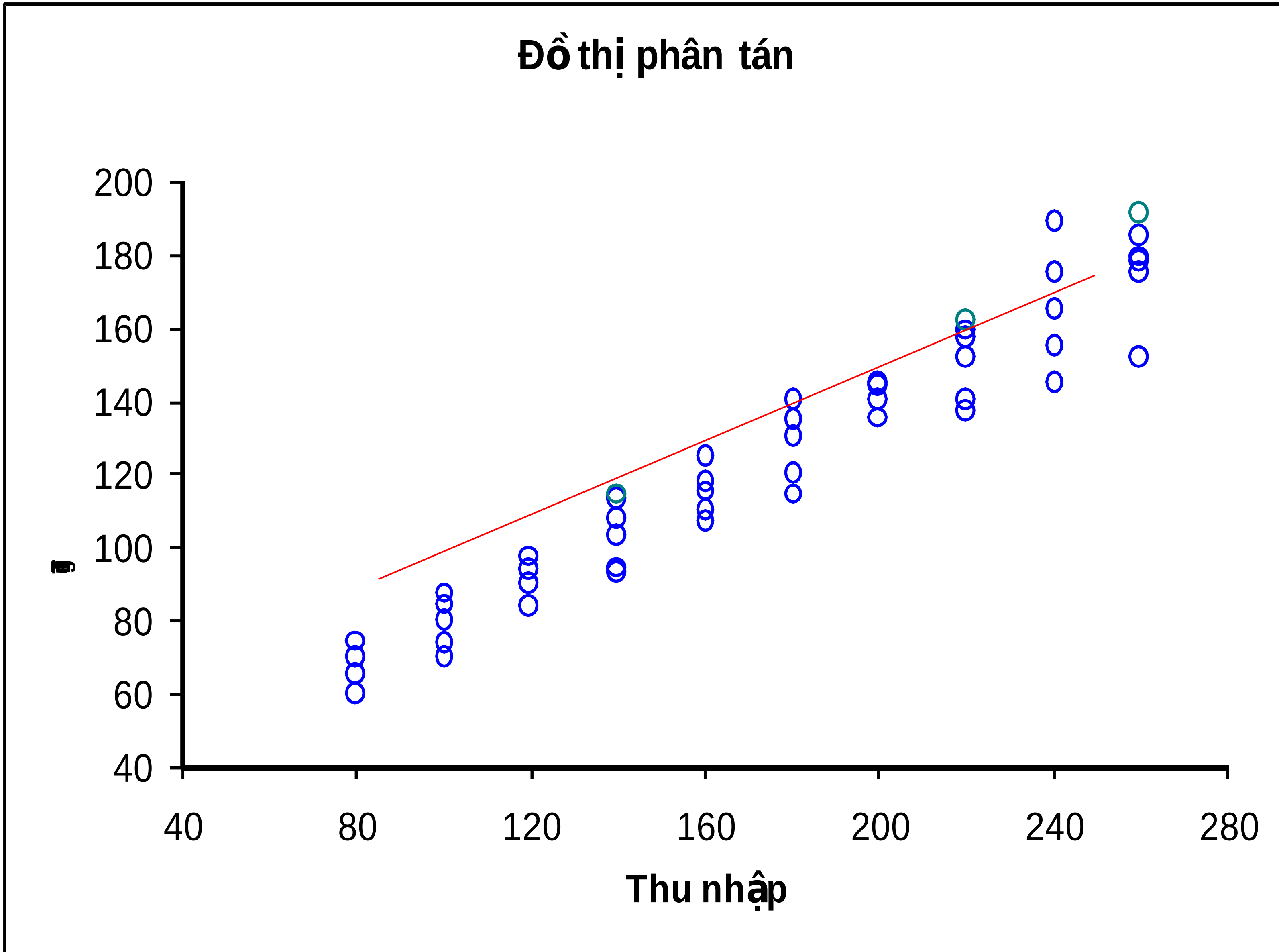
  - Phân tích hồi quy ước lượng hoặc dự báo một biến trên cơ sở giá trị đã cho của các biến khác.

2.2.Mô hình hồi quy tổng thể và hồi quy mẫu

2.2.1. Mô hình hồi quy tổng thể (PRF)

Ví dụ 2.1. Hồi quy tiêu dùng Y theo thu nhập X. Xét sự phụ thuộc chi tiêu của một gia đình vào thu nhập ở một địa phương có tổng cộng 40 hộ gia đình. Ta được số liệu cho ở bảng sau:

Y \X	80	100	120	140	160	180	200
	55	65	79	80	102	105	120
	60	70	84	93	107	110	136
	65	74	90	95	110	110	140
	70	80	94	103	116	115	144
	75	85	98	108	118	120	145
		88		113	125	130	
				115			
Σ	325	462	445	707	678	690	685
E(Y/Xi)	65	77	89	101	113	115	137



Đồ thị phân tán
200
180
160
140
120
100
80
60
40
40
80
120
160
200
240
280
Thu nhập

$E(Y/X_i) = f(X_i) = \beta_1 + \beta_2 X_i$

β_1 : là hệ số chặn – tung độ gốc

β_2 : hệ số góc - hệ số đo độ dốc đường hồi quy

Ví dụ ở hộ gia đình có mức chi tiêu 130 ta có:

$130 = \beta_1 + \beta_2.180 + 15$

115

Mô hình hồi quy tổng thể ngẫu nhiên:

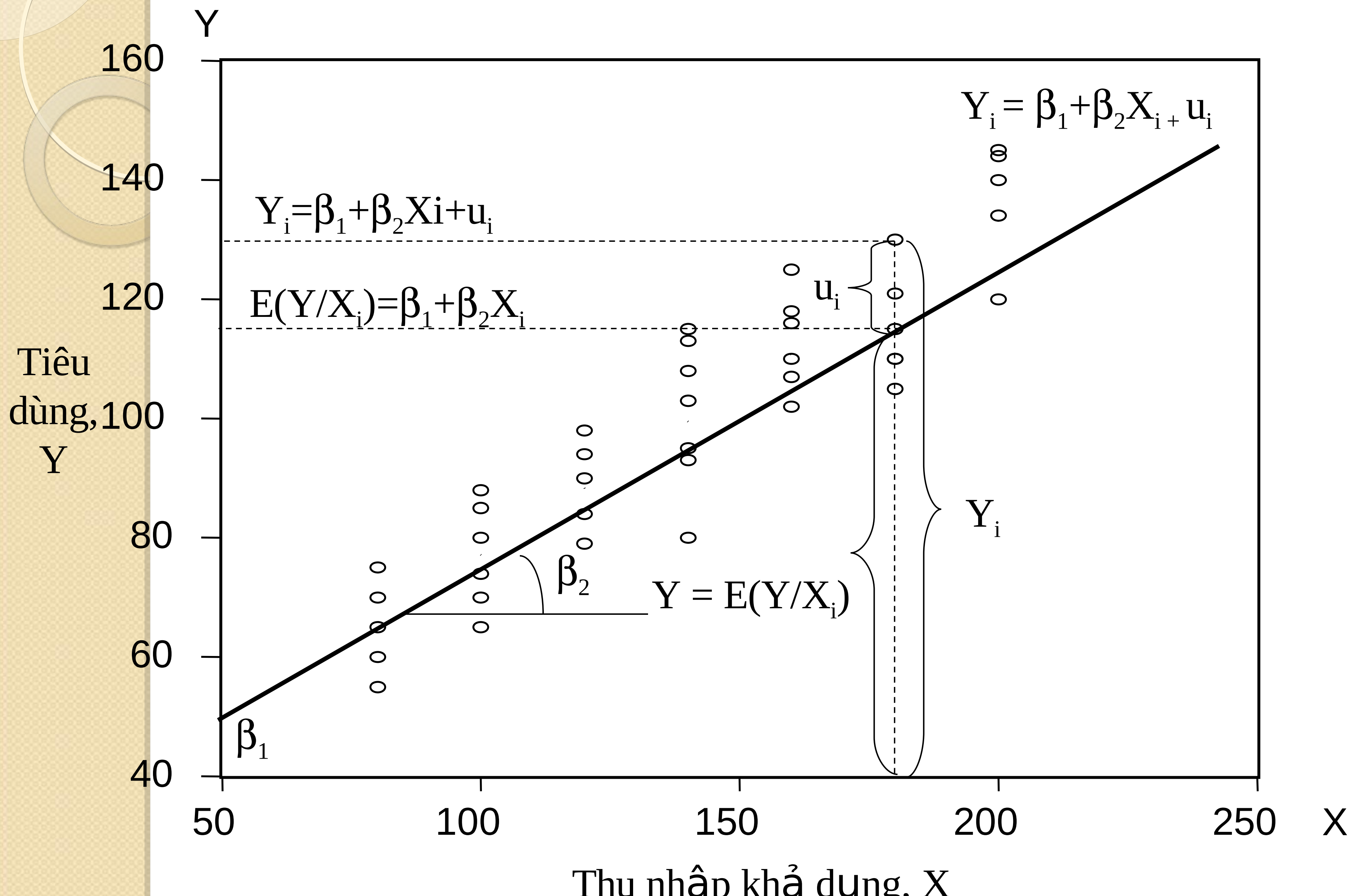
$Y_i = \beta_1 + \beta_2 X_i + u_i$

u_i:sai số ngẫu nhiên của tổng thể ứng với quan sát thứ i

u_i: đại diện những nhân tố còn lại ảnh hưởng đến chi tiêu

Sai số ngẫu nhiên hình thành từ nhiều nguyên nhân:

- Bỏ sót biến giải thích.
- Sai số khi đo lường biến phụ thuộc.
- Dạng mô hình hồi quy không phù hợp.
- Các tác động không tiên đoán được.


Y
$Y_i = \beta_1 + \beta_2 X_{i} + u_i$
$Y_i = \beta_1 + \beta_2 Xi + u_i$
$E(Y/X_i) = \beta_1 + \beta_2 X_i$
u_i
Y_i
β_2
$Y = E(Y/X_i)$
β_1
Tiêu dùng, Y
160
140
120
100
80
60
40
50
100
150
200
250
X
Thu nhập khả dụng, X

2.2.2. Mô hình hồi quy mẫu (SRF)

Mô hình hồi quy mẫu:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

Trong đó

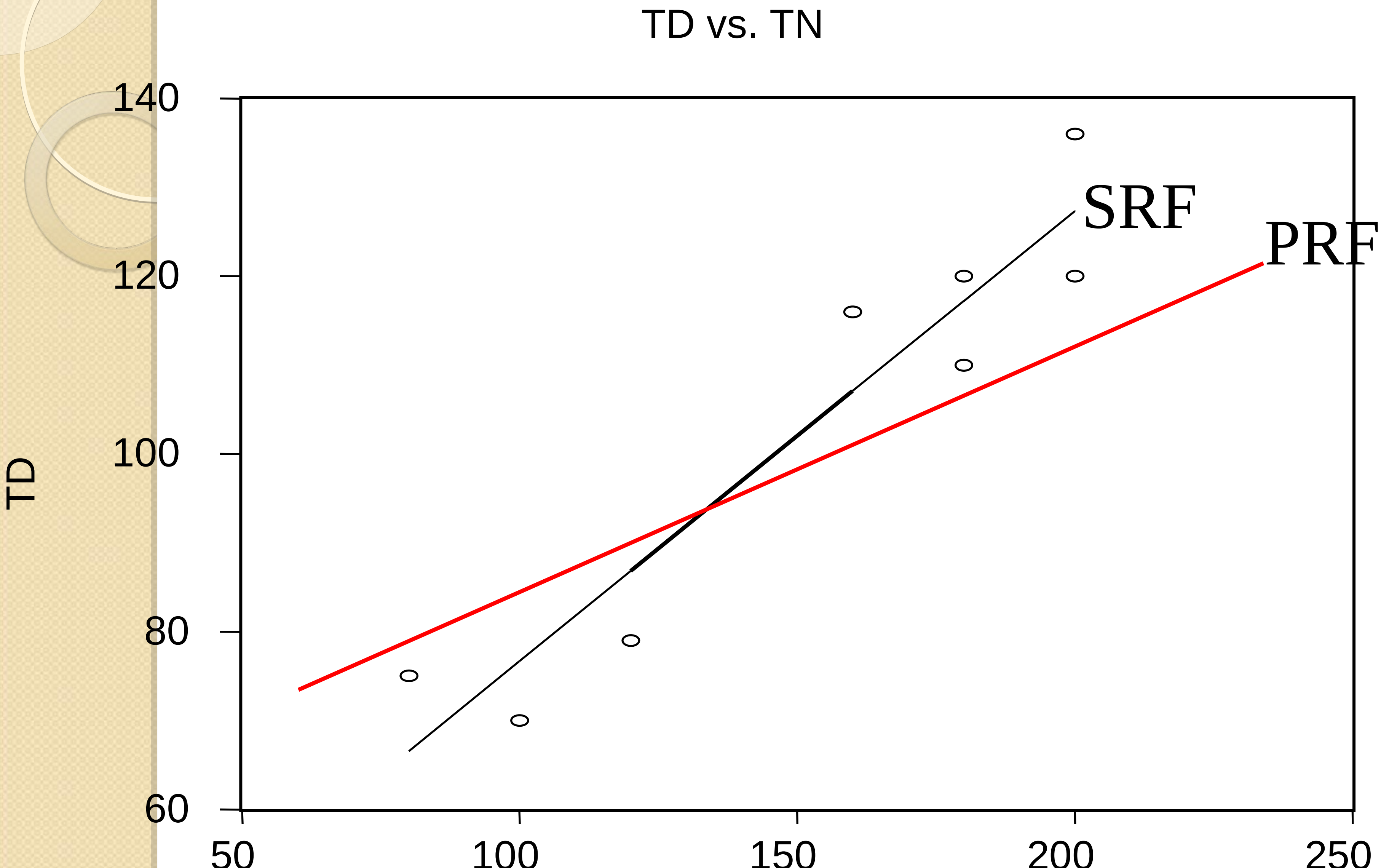
$\hat{\beta}_1$: ước lượng cho β_1.

$\hat{\beta}_2$: Ước lượng cho β_2.

$\hat{Y}_i$: Ước lượng cho E(Y/Xi)

Mô hình hồi quy mẫu ngẫu nhiên

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$



Hình 2.1. Mô hình hồi quy tổng thể và mẫu tuyến tính

2.2.3. Mô hình hồi quy tuyến tính (LRF)

Hồi quy tuyến tính chỉ yêu cầu tuyến tính trong các tham số, không yêu cầu tuyến tính trong biến số.

* Mô hình

$$Y = \beta_1 + \beta_2 \frac{1}{X} + u_i$$

là mô hình tuyến tính trong các tham số nhưng phi tuyến theo biến số.

* Mô hình

$$Y = \beta_1 + (1 - \beta_2)^2 X + u_i$$

là mô hình phi tuyến trong các tham số nhưng tuyến tính trong biến số.

Hồi quy tuyến tính theo OLS chỉ chấp nhận dạng mô hình tuyến tính trong tham số.

2.3. Ước lượng các hệ số của mô hình hồi quy theo phương pháp bình phương tối thiểu-OLS

2.3.1.Các giả định của mô hình hồi quy tuyến tính cổ điển

Giả thiết 1:Các biến giải thích là phi ngẫu nhiên tức là các giá trị của chúng được cho trước hoặc được xác định.

Giả thiết 2: Kỳ vọng của yếu tố ngẫu nhiên u_i bằng 0, tức là: $$E[u_i|X_i]=0$$

Giả thiết 3: Các u_i có phương sai bằng nhau (phương sai thuần nhất)

$$\operatorname{var}[u_i|X_i]=\operatorname{var}[u_j|X_i]=\sigma^2 \quad \forall i\neq j$$

Giả thiết 4: Không có tự tương quan giữa các u_i:

$$\operatorname{cov}[u_i, u_j] = 0 \qquad \forall i \neq j$$

Giả thiết 5: Không tự tương quan giữa u_i với X_i:

$$Cov\ (u_i, X_i) = 0$$

Định lý Gauss-Markov

Với các giả định của mô hình hồi quy tuyến tính cổ điển, mô hình hồi quy tuyến tính theo phương pháp bình phương tối thiểu là ước lượng tuyến tính không thiên lệch tốt nhất

2.3.2. Nội dung của phương pháp

Cho n quan sát của 2 đại lượng (Y_i, X_i) $i = \overline{1, n}$

Mô hình hồi quy mẫu ngẫu nhiên

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i$$

$$e_i = Y_i - \hat{Y}_i$$

$$e_1 = Y_1 - \hat{Y}_1 = Y_1 - (\hat{\beta}_1 + \hat{\beta}_2 . X_1) \Rightarrow \min \Leftrightarrow 0$$

$$e_2 = Y_2 - \hat{Y}_2 = Y_2 - (\hat{\beta}_1 + \hat{\beta}_2 . X_2) \Rightarrow 0$$

$$e_3 = Y_3 - \hat{Y}_3 = Y_3 - (\hat{\beta}_1 + \hat{\beta}_2 . X_3) \Rightarrow 0$$

=> tìm $\sum e_i^2$ => 0: Phương pháp bình phương bé nhất

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \right)^2$$

Điều kiện để phương trình trên đạt cực trị là:

$$\frac{\partial \left(\sum_{i=1}^{n} e_i^2 \right)}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \right) = -2 \sum_{i=1}^{n} e_i = 0$$

$$\frac{\partial \left(\sum_{i=1}^{n} e_i^2 \right)}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^{n} \left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \right) X_i = -2 \sum_{i=1}^{n} e_i X_i = 0$$

Giải hệ phương trình trên được:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} Y_i X_i - n.\bar{X}.\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n.(\bar{X})^2}$$

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\bar{Y} = \frac{\sum Y_i}{n}$$

đặt $x_i = X_i - \bar{X}$

$$y_i = Y_i - \bar{Y}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

Thí dụ: Bảng sau đây cho số liệu về mức chi tiêu tiêu dùng (Y – đô la/tuần) và thu nhập hàng tuần (X - $/tuần) của một mẫu gồm 10 hộ gia đình. Giả sử X và Y quan hệ tương quan tuyến tính. Hãy ước lượng hàm hồi quy của Y theo X.

Yi	70	65	90	95	110	115	120	140	155	150
Xi	80	100	120	140	160	180	200	220	240	260

Phân tích:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} Y_i X_i - n.\bar{X}.\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n.(\bar{X})^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Giải:

$$\sum Y_i = 1110; \quad \sum X_i = 1700; \quad \sum X_i^2 = 322000; \quad \sum X_i Y_i = 205500;$$

$$\overline{X} = \frac{\sum X_i}{n} = \frac{1700}{10} = 170; \qquad \overline{Y} = \frac{1110}{10} = 111$$

$$\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} Y_i X_i - n.\overline{X}.\overline{Y} = 205500 - 10 \times 170 \times 111 = 16800$$

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} X_i^2 - n.(\overline{X})^2 = 322000 - 10.\left(170\right)^2 = 33000$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{16800}{33000} = 0,5091 \qquad \hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X} = 111 - 0,5091\left(170\right) = 24,453$$

$$\hat{Y}_i = 24,453 + 0,5091 X_i$$

Giá trị $\hat{\beta}_1 = 24,453$ là tung độ gốc, chỉ mức tiêu dùng trung bình hàng tuần khi mà thu nhập hàng tuần bằng 0.

Giá trị $\hat{\beta}_2 = 0,5091$ chỉ ra rằng, xét các giá trị của X nằm trong khoảng (80;260), khi thu nhập tăng 1$/tuần thì chi tiêu tiêu dùng của một gia đình tăng trung bình khoảng 0,51 $/tuần.

2.4. Phương sai, sai số chuẩn của các ước lượng, hệ số xác định R^2, hệ số tương quan r

2.4.1. Phương sai và sai số chuẩn của các ước lượng

Phương sai

Sai số chuẩn

$$Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{X_i^2}{n \quad x_i^2}\sigma^2$$

$$se(\hat{\beta}_1) = \sigma_{\hat{\beta}_1} = \sqrt{\sigma^2_{\hat{\beta}_1}}$$

$$Var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{1}{x_i^2}\sigma^2$$

$$se(\hat{\beta}_2) = \sigma_{\hat{\beta}_2} = \sqrt{\sigma^2_{\hat{\beta}_2}}$$

Trong đó : σ^2 = var (U_i). Do σ^2 chưa biết nên dùng ước lượng của nó là

$$\hat{\sigma}^2 = \frac{e_i^2}{n-2}$$

TSS (Total Sum of Squares): Tổng bình phương tất cả các sai lệch giữa giá trị thực tế của Y với giá trị trung bình của nó.

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum Y_i^2 - n.\left(\bar{Y}\right)^2$$

ESS (Explained Sum of Squares): Tổng bình phương tất cả các sai lệch giữa giá trị của Y được tính theo mô hình với giá trị trung bình của nó.

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \left(\hat{\beta}_2\right)^2 \left(\sum X_i^2 - n.\left(\bar{X}\right)^2\right)$$

RSS (Residual Sum of Squares): Tổng bình phương tất cả các sai lệch giữa giá trị thực tế với giá trị lý thuyết theo mô hình của Y.

$$RSS = \sum e_i^2 = \sum \left(Y_i - \hat{Y}_i\right)^2$$

$$TSS = ESS + RSS$$

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2 = \left(\hat{\beta}_2\right)^2 \sum x_i^2 = (0,5091)^2 \; 33000 = 8553,0327$$

$$TSS = \sum Y_i^2 - n.\left(\bar{Y}\right)^2 = 132100 - 10\left(111\right)^2 = 8890$$

$$RSS = TSS - ESS = 8890 - 8553,0327 = 336,9678$$

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-2} = \frac{RSS}{n-2}$$

$$\hat{\sigma}^2 = \frac{336,9678}{10-2} = 42,1210$$

$$Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{X_i^2}{n \quad x_i^2}\sigma^2$$

$$Var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{1}{x_i^2}\sigma^2$$

$$Var(\hat{\beta}_1) = \frac{322000}{10 \quad 33000} 42,1210 = 41,1$$

$$Var(\hat{\beta}_2) = \frac{42,1210}{33000} = 0,001276$$

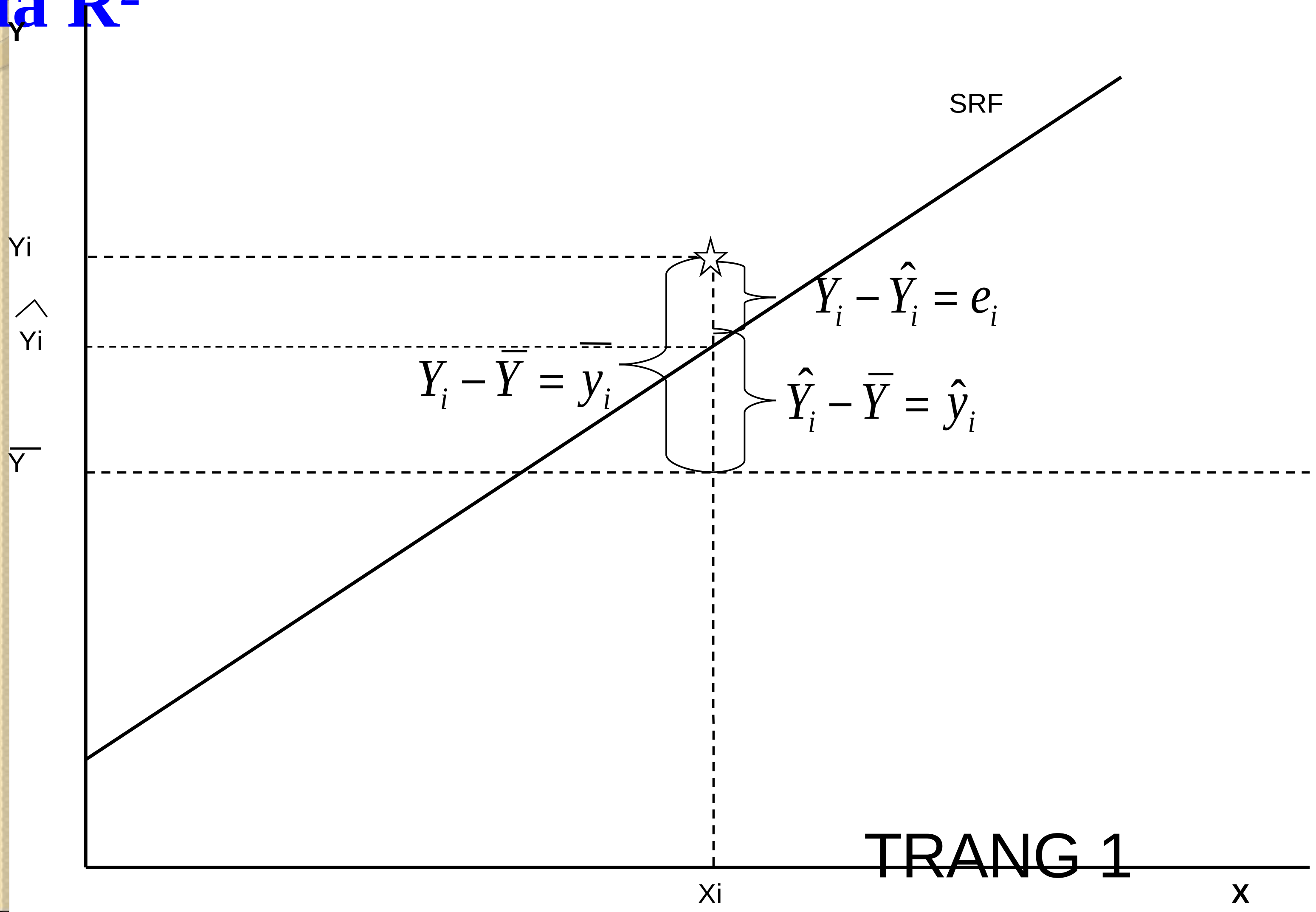
$$Var(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \frac{X_i^2}{n \quad x_i^2}\sigma^2$$

$$Var(\hat{\beta}_2) = \sigma^2_{\hat{\beta}_2} = \frac{1}{x_i^2}\sigma^2$$

$$se(\hat{\beta}_1) = \sigma_{\hat{\beta}_1} = \sqrt{\sigma^2_{\hat{\beta}_1}} = \sqrt{41,1}$$

$$se(\hat{\beta}_2) = \sigma_{\hat{\beta}_2} = \sqrt{\sigma^2_{\hat{\beta}_2}} = \sqrt{0,001276}$$

2.4.2. Hệ số xác định R^2 và hệ số tương quan r
Thước đo độ phù hợp của mô hình đối với dữ liệu là R^2



Hệ số xác định

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

$$R^2 = \frac{\mathrm{ESS}}{TSS} = \frac{8553,0327}{8890} = 0,9621$$

Trong mô hình 2 biến, người ta chứng minh được rằng

$$R^2 = \frac{\hat{\beta}_2^2 \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} y_i^2}$$

=> Có thể nói R^2 phản ánh tỷ lệ mô hình lý thuyết phản ánh thực tế.

*** Tính chất của R^2**

- $0 \leq R^2 \leq 1$. Với $R^2=0$ thể hiện X và Y độc lập thống kê. $R^2 = 1$ thể hiện X và Y phụ thuộc tuyến tính hoàn hảo.

- R^2 không xét đến quan hệ nhân quả

Hệ số tương quan r: Hệ số tương quan r đo lường mức độ phụ thuộc tuyến tính giữa 2 đại lượng X và Y.

$$r = \frac{\sum_{i=1}^{n} y_i x_i}{\sqrt{\sum_{i=1}^{n} \text{?}_i^2 \sum_{i=1}^{n} \text{?}_i^2}}$$

Tính chất của r:

- r > 0: giữa X và Y có quan hệ đồng biến

  r→ ± 1: X và Y có quan hệ tuyến tính chặt chẽ

  r → 0: X và Y có quan hệ tuyến tính không chặt chẽ

  r < 0: X và Y có quan hệ nghịch biến

- Hệ số tương quan có tính chất đối xứng: $r_{XY} = r_{YX}$
- r độc lập với gốc toạ độ và các tỷ lệ. Nghĩa là: với a, c > 0, b, d là hằng số, và:

$$X_i^* = aX_i + b$$

$$Y_i^* = cY_i + d$$

Thì : $r_{XY} = r_{X^*Y^*}$

- Nếu X, Y độc lập theo quan điểm thống kê thì hệ số tương quan giữa chúng bằng 0.
- r chỉ là đại lượng đo sự kết hợp tuyến tính hay phụ thuộc tuyến tính. r không có ý nghĩa để mô tả quan hệ phi tuyến.

$$R^2 = \frac{\left(\sum_{i=1}^{n} x_i y_i\right)^2}{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2} = r_{X,Y}^2$$

$r_{XY} = \pm R$; r_{XY} và $\hat{\beta}_2$ cùng dấu.

$$r_{XY} = \sqrt{0,9621} = 0,9808$$

2.5. Phân bố xác suất của các ước lượng

Giả thiết 6: u_i có phân phối N (0, σ^2),

Với các giả thiết nêu trên, các ước lượng $\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2$ có các tính chất sau:

- Chúng là các ước lượng không chệch
- Có phương sai cực tiểu
- Khi số quan sát đủ lớn thì các ước lượng này xấp xỉ với giá trị thực của phân phối

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1}) \qquad \Rightarrow Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$$

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2}) \qquad \Rightarrow Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \sim N(0,1)$$

2.6. Khoảng tin cậy của các tham số

Ước lượng khoảng cho hệ số hồi quy với mức ý nghĩa α (độ tin cậy 1- α) như sau

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$$

$$, \varepsilon_i = t_{(n-2, \alpha/2)} SE(\hat{\beta}_i)$$

Tìm khoảng tin cậy 95% của β_1 , β_2

$$se(\hat{\beta}_1) = \sqrt{var(\hat{\beta}_1)} = \sqrt{41,0999} = 6,4109$$

$$se(\hat{\beta}_2) = \sqrt{var(\hat{\beta}_2)} = \sqrt{0,001276} = 0,03572$$

- Với độ tin cậy 95% thì $t(n-2, \alpha/2) = t(8; 0,025) = 2,306$
- Vậy khoảng tin cậy của β_1 là

$$24,453 \quad 2,306 \quad 6,4109$$

hay $9,6695 < \beta_1 < 39,2365$

- Vậy khoảng tin cậy của β_2 là

$$0,5091 \quad 2,306 \quad 0,03572$$

hay $0,4267 < \beta_2 < 0,5915$

Ý nghĩa: Với các điều kiện các yếu tố khác không thay đổi, khi thu nhập tăng 1$/tuần thì chi tiêu tiêu dùng trung bình của một gia đình tăng trong khoảng từ 0,4267 đến 0,5914 $/tuần

2.7. Kiểm định giả thiết về hệ số hồi quy

$$H_0: \beta_2 = \beta_2^*$$

$$H_1: \beta_2 \neq \beta_2^*$$

Có 3 cách để kiểm định giả thiết:

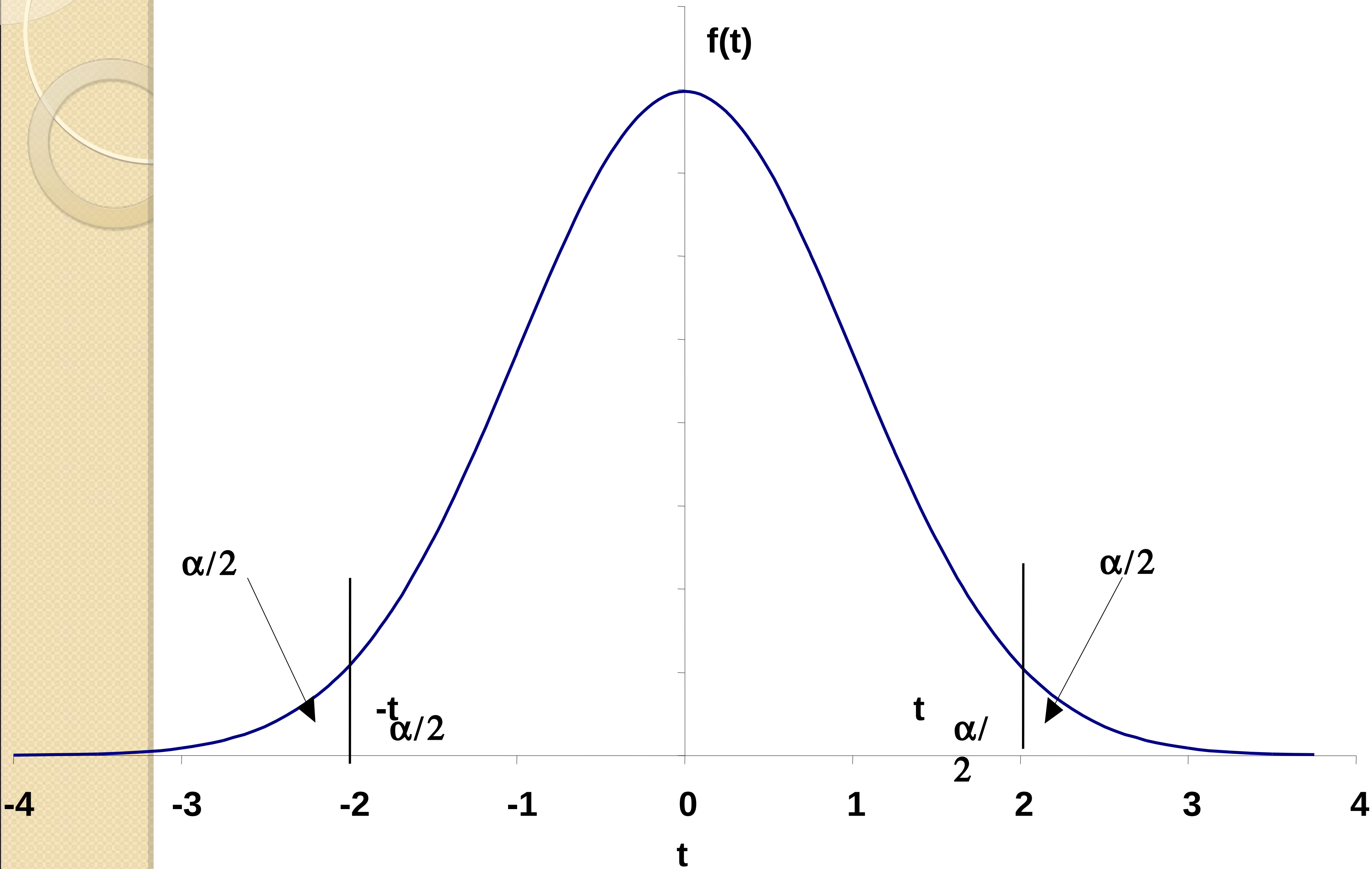
Cách 1: Kiểm định t

$$t = \frac{\hat{\beta}_2 - \beta_2^*}{SE(\hat{\beta}_2)}$$

Quy tắc quyết định

Nếu $|t| > t_{(n-2,\alpha/2)}$ thì bác bỏ H_0.

Nếu $|t| \le t_{(n-2,\alpha/2)}$ thì ta không thể bác bỏ H_0.


f(t)
α/2
α/2
-t
α/2
t
α/
2
-4
-3
-2
-1
0
1
2
3
4
t

Cách 2: Phương pháp khoảng tin cậy

Giả sử ta tìm được khoảng tin cậy của βi là:

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i) \quad \varepsilon_i = t_{(n-2,\alpha/2)} SE(\hat{\beta}_i)$$

với mức ý nghĩa α trùng với mức ý nghĩa của gt H_0

Quy tắc quyết định

- Nếu $\beta_i^* \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$ chấp nhận H_0
- Nếu $\beta_i^* \notin (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$ bác bỏ H_0

<u>Cách 3</u>: Phương pháp P-value

$$t_i = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

Tính

$$P(T > |t_i|) = p$$

Quy tắc quyết định

- Nếu $p \le \alpha$: Bác bỏ H_0
- Nếu $p > \alpha$: Chấp nhận H_0

(Phương pháp này thường dùng khi tiến hành trên máy vi tính)

Kiểm định giả thiết $\beta_2 = 0$ với giả thiết đối $\beta_2 \neq 0$ với mức ý nghĩa $\alpha = 5\%$.

$$\hat{\beta}_2 = 0,5091; \ se(\hat{\beta}_2) = 0,03572$$

$$t = \frac{0,5091 - 0}{0,03572} = 14,2525$$

Với mức ý nghĩa $\alpha = 5\%$ và bậc tự do là $n - 2 = 8$ thì $t_{(n-2;\alpha/2)} = 2,306$

$t > t_{(n-2,\alpha/2)}$ Bác bỏ giả thiết H_0

Ý nghĩa: biến thu nhập thực sự có ảnh hưởng đến chi tiêu.

2.8. Kiểm định sự phù hợp của mô hình – Dự báo

2.8.1. Kiểm định sự phù hợp của mô hình

Kiểm định giả thiết H_0: $R^2 = 0$ với mức ý nghĩa α hay độ tin cậy 1 - α

Xét thống kê

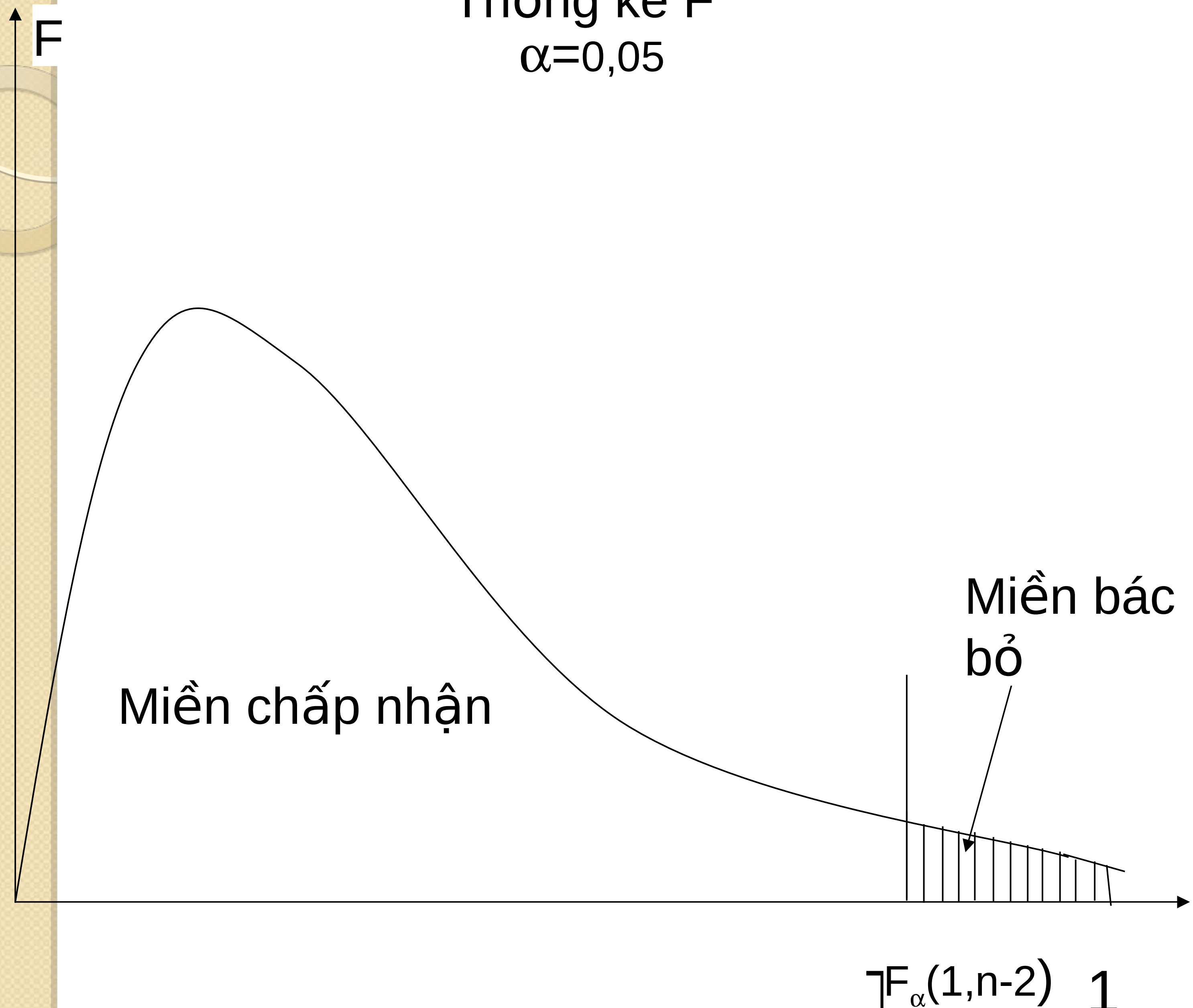
$$F = \frac{R^2(n-2)}{1-R^2}$$

Quy tắc quyết định

- Nếu $F > F_\alpha(1,n-2)$: Bác bỏ H_0
- Nếu $F \le F_\alpha(1,n-2)$: Chấp nhận H_0

Thống kê F

α=0,05



Kiểm định sự phù hợp của hàm hồi quy với mức ý nghĩa α = 5%.

$$F = \frac{R^2\left(n-2\right)}{1-R^2} = \frac{0,9621\left(10-2\right)}{1-0,9621} = 203,08$$

$$F_\alpha(1;n-2) = F_{0,05}(1;8) = 5,32$$

F > F_α(1,n-2): Bác bỏ H_0

Vậy thu nhập thực sự tác động đến tiêu dùng.

2.8.2. Dự báo

Cho trước giá trị X = X_0, hãy dự báo giá trị trung bình và giá trị cá biệt của Y với mức ý nghĩa α hay độ tin cậy 1 - α.

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

*** Dự báo điểm**

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

* Dự báo giá trị trung bình của Y

$$E(Y / X_0) \in (\hat{Y}_0 - \varepsilon_0 ; \hat{Y}_0 + \varepsilon_0)$$

Với:

$$\varepsilon_0 = SE(\hat{Y}_0) t_{(n-2, \alpha/2)}$$

$$SE(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)}$$

$$Var(\hat{Y}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(\overline{X} - X_0)^2}{\sum x_{\grave{i}}^2} \right)$$

* Dự báo giá trị cá biệt của Y

$$Y_0 \in (\hat{Y}_0 - \varepsilon_0 ; \hat{Y}_0 + \varepsilon_0)$$

Với:

$$\varepsilon_0 = SE(Y_0 - \hat{Y}_0) t_{(n-2,\alpha/2)}$$

$$SE(Y_0 - \hat{Y}_0) = \sqrt{Var(Y_0 - \hat{Y}_0)}$$

$$Var(Y_0 - \hat{Y}_0) = \hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(\bar{X} - X_0)^2}{\sum x_i^2}\right)$$

Hãy dự báo giá trị trung bình của chi tiêu cho tiêu dùng khi thu nhập ở mức 100$/tuần với hệ số tin cậy 95%.

$$Y_0 \in (\hat{Y}_0 - \varepsilon_0 ; \hat{Y}_0 + \varepsilon_0)$$

$$\varepsilon_0 = SE(Y_0 - \hat{Y}_0)\, t_{(n-2,\alpha/2)}$$

$$SE(Y_0 - \hat{Y}_0) = \sqrt{Var(Y_0 - \hat{Y}_0)}$$

$$Var(Y_0 - \hat{Y}_0) = \hat{\sigma}^2\left(1 + \frac{1}{n} + \frac{(\bar{X} - X_0)^2}{x_i^2}\right)$$

Hãy dự báo giá trị trung bình của chi tiêu cho tiêu dùng khi thu nhập ở mức 100$/tuần với hệ số tin cậy 95%.

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 = 24,453 + 0,5091 \quad 100 = 75,363$$

$$Var(\hat{Y}_0) = 42,1210 \left[\frac{1}{10} + \frac{(100-170)^2}{33000} \right] = 10,4664$$

$$se(\hat{Y}_0) = \sqrt{Var(\hat{Y}_0)} = 3,2352; t_{(n-2;\alpha/2)} = 2,306$$

$$75,363 \quad 2,306 \quad 3,2352$$

$$67,903 \quad \hat{Y} \quad 82,82$$

KINH TẾ LƯỢNG

Chương 3:
MỞ RỘNG MÔ HÌNH HỒI QUY 2 BIẾN

3.1. Mô hình hồi quy qua gốc tọa độ

Mô hình hồi quy tổng thể:

$$E(Y/X) = \beta_2 X_i$$

$$Y_i = \beta_2 X_i + u_i$$

Mô hình hồi quy mẫu ngẫu nhiên: $Y_i = \hat{\beta}_2 X_i + e_i$

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$$

$$Var(\hat{\beta}_2) = \frac{\hat{\sigma}^2}{\sum X_i^2}, \hat{\sigma}^2 = \frac{\sum e_i^2}{n-1}$$

3.2. Mô hình tuyến tính logarit (log-log)

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i}$$

MHHQTTNN: $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_1 + u_i$

$$\frac{d \ln Y}{dX} = \frac{\beta_2}{X} \Leftrightarrow \frac{dY/Y}{dX} = \frac{\beta_2}{X}$$

$$\beta_2 = \frac{dY/Y}{dX/X} = E_{Y/X} = \frac{dY}{dX}\frac{X}{Y}$$

Ví dụ: $\ln Y_i = 2 - 0{,}75 \ln X_i + u_i$

Khi giá tăng 1% thì lượng cầu của loại hàng hoá này sẽ giảm 0,75%.

3.3. Mô hình bán logarit

3.3.1. Mô hình log-lin

Mô hình bán logarit có dạng:

$$\ln Y_i = \beta_1 + \beta_2.X_i + u_i$$

$$\beta_2 = \frac{d(\ln Y)}{dX} = \frac{(1/Y)dY}{dX} = \frac{dY/Y}{dX}$$

$$\beta_2 = \frac{\text{Thay đổi tương đối của biến phụ thuộc (Y)}}{\text{Thay đổi tuyệt đối của biến độc lập (X)}}$$

Nếu nhân thay đổi tương đối của Y lên 100 thì β_2 (β_2>0) sẽ là tốc độ tăng trưởng (%) của Y đối với thay đổi tuyệt đối của t. Nếu $\beta_2 < 0$ thì β_2 là tốc độ giảm sút.

Ví dụ 3.1: Tổng SP nội địa tính theo giá năm 1987 của Mỹ trong khoảng thời gian 1972-1991

Năm	RGDP	Năm	RGDP	Năm	RGDP
1972	3107.1	1979	3796.8	1986	4404.5
1973	3268.6	1980	3776.3	1987	4539.9
1974	3248.1	1981	3843.1	1988	4718.6
1975	3221.7	1982	3760.3	1989	4838
1976	3380.8	1983	3906.6	1990	4877.5
1977	3533.3	1984	4148.5	1991	4821
1978	3703.5	1985	4279.8		

Với Y = ln(RGDP), và kết quả hồi quy như sau:

$$\hat{Y}_i = 8,0139 + 0,0247t$$

GDP thực tăng với tốc độ 2,47%/năm từ 1972-91.

*** Mô hình xu hướng tuyến tính:**

Mô hình:

$$Y_t = \beta_1 + \beta_2.t + u_t$$

Tức hồi quy Y theo thời gian, và phương trình trên được gọi là mô hình xu hướng tuyến tính và t được gọi là biến xu hướng.

Với số liệu ở VD 3.1,đặt Y=RGDP, ta có kết quả:

$$\hat{Y}_i = 2933{,}054 + 97{,}6806t$$

Mô hình này được giải thích như sau: trong giai đoạn 1972-1991, trung bình GDP thực của Mỹ tăng với tốc độ tuyệt đối 97,68tỷ USD/năm.

3.3.2. Mô hình lin-log

Mô hình lin-log cho biết sự thay đổi tuyệt đối của Y khi X thay đổi 1%.

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \qquad \beta_2 = \frac{dY}{dX/X}$$

Như vậy nếu X thay đổi 0,01 (hay 1%) thay đổi tuyệt đối của Y sẽ là 0,01β_2.

Ví dụ 3.3. lấy bài tập 3.2, ta có

$$\hat{Y}_i = 265678.7 + 24994.11 \ln X_i + u_i$$

β_2=24994.11 có nghĩa là trong khoảng thời gian 1970-84, lượng cung tiền tăng lên 1%, sẽ kéo theo sự gia tăng bình quân của GDP là 249,94 triệu USD.

3.4. Mô hình nghịch đảo

Các mô hình có dạng sau được gọi là mô hình nghịch đảo: $$Y_i = \beta_1 + \beta_2 \frac{1}{X} + u_i$$

Mô hình này phù hợp cho nghiên cứu đường chi phí đơn vị, đường tiêu dùng theo thu nhập Engel hoặc đường cong Philip.

KINH TẾ LƯỢNG

Chương 4: MÔ HÌNH HỒI QUY BỘI

4.1. Mô hình hồi quy tuyến tính 3 biến

Mô hình hồi quy tổng thể

$$E(Y / X_2, X_3) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Mô hình hồi quy tổng thể ngẫu nhiên:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

u_i: sai số ngẫu nhiên của tổng thể

4.1.1. Ước lượng các tham số của mô hình (OLS)

Cho n quan sát của 3 đại lượng Y, X_2, X_3, ký hiệu quan sát thứ i là Y_i, X_{2i}, và X_{3i}.

$e_i = Y_i - \hat{Y}_i$ sai số của mẫu ứng với quan sát thứ i

$$Q = \sum e_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \to \min$$

$$\frac{dQ}{d\hat{\beta}_1} = -2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}) = 0$$

$$\frac{dQ}{d\hat{\beta}_2} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{dQ}{d\hat{\beta}_3} = 2\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}_{2i} - \hat{\beta}_3 \overline{X}_{3i}$$

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{\sum y_i x_{3i} \sum x_{2i}^2 - \sum y_i x_{2i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

$$x_i = X_i - \overline{X} \qquad\qquad y_i = Y_i - \overline{Y}$$

TRANG 1

4.1.2. Phương sai của các ước lượng

$$Var(\hat{\beta}_1) = \left(\frac{1}{n} + \frac{\overline{X}_2^2 \sum x_{3i}^2 + \overline{X}_3^2 \sum x_{2i}^2 - 2\overline{X}_2 \overline{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\right)\sigma^2$$

$$Var(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\sigma^2$$

$$Var(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i}x_{3i})^2}\sigma^2$$

Do σ^2 là phương sai của u_i chưa biết nên trong thực tế người ta dùng ước lượng không chệch của nó:

$$\hat{\sigma}^2 = \frac{\sum e_i^2}{n-3} = \frac{(1-R^2)\sum y_i^2}{n-3}$$

4.1.3. Hệ số xác định và hệ số xác định hiệu chỉnh

Hệ số xác định R^2

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

MH hồi quy 3 biến

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2}$$

Hệ số xác định hiệu chỉnh

Với *k* là tham số của mô hình,
kể cả hệ số tự do

$$\bar{R}^2 = 1 - \frac{\sum e_i^2 / (n-k)}{\sum y_i^2 / (n-1)}$$

Mối quan hệ giữa R^2 và $\overline{R}^2$

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Người ta dùng $\overline{R}^2$ để xem xét việc đưa thêm 1 biến vào mô hình. Biến mới đưa vào mô hình phải thỏa 2 điều kiện:

- Làm $\overline{R}^2$ tăng
- Khi kiểm định giả thiết hệ số của biến này trong mô hình với giả thiết H_0 thì phải bác bỏ H_0.

4.1.4. Khoảng tin cậy của các tham số

Khoảng tin cậy của tham số β_i với mức ý nghĩa α hay độ tin cậy 1- α

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i)$$

$$\varepsilon_i = SE(\hat{\beta}_i) t_{(n-3, \alpha/2)}$$

4.1.5. Kiểm định giả thiết

* Kiểm định giả thiết H_0: $\beta_i = \beta_i^*$

$$t_i = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

Nguyên tắc quyết định:

Nếu $t_i > t_{(n-3,\alpha/2)}$ hoặc $t_i < -t_{(n-3,\alpha/2)}$: bác bỏ H_0

Nếu $-t_{(n-3,\alpha/2)} \leq t_i \leq t_{(n-3,\alpha/2)}$: chấp nhận H_0

* Kiểm định giả thiết đồng thời bằng không:

H_0: $\beta_2 = \beta_3 = 0$; (H_1: ít nhất 1 trong 2 tham số khác 0)

$$F = \frac{R^2(n-k)}{(1-R^2)(k-1)}$$

Nguyên tắc quyết định:

- $F > F_\alpha(2, n-3)$: Bác bỏ H_0: Mô hình phù hợp
- $F \leq F_\alpha(2, n-3)$: Chấp nhận H_0: Mô hình không phù hợp

4.2. Mô hình hồi quy *k* biến

Mô hình hồi quy tổng thể

$$E(Y/X_2,...X_k)=\beta_1+\beta_2X_{2i}+...+\beta_kX_{ki}$$

Mô hình hồi quy mẫu ngẫu nhiên:

$$\hat{Y}_i=\hat{\beta}_1+\hat{\beta}_2X_{2i}+...+\hat{\beta}_kX_{ki}+e_i$$

=> $$e_i=Y_i-\hat{Y}_i=Y_i-\hat{\beta}_1-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_kX_{ki}$$

4.2.1. Ước lượng các tham số của mô hình (OLS)

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right)^2 \to \min$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right) = 0$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_2} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{k,i}\right) X_{2i} = 0$$

...

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_k} = -2\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_k X_{ki}\right) X_{ki} = 0$$

4.2.2. Khoảng tin cậy của các tham số, kiểm định các giả thiết hồi quy

*** Khoảng tin cậy các tham số**

$$\beta_i \in (\hat{\beta}_i - \varepsilon_i ; \hat{\beta}_i + \varepsilon_i) \qquad \varepsilon_i = SE(\hat{\beta}_i) t_{(n-k,\alpha/2)}$$

*** Kiểm định giả thiết**

Kiểm định giả thiết $H_0: \beta_i = \beta_i^*$

$$t_i = \frac{\hat{\beta}_i - \beta_i^*}{SE(\hat{\beta}_i)}$$

Nguyên tắc quyết định:

Nếu $t_i > t_{(n-k,\alpha/2)}$ hoặc $t_i < -t_{(n-k,\alpha/2)}$: bác bỏ H_0

Nếu $-t_{(n-k,\alpha/2)} \le t_i \le t_{(n-k,\alpha/2)}$: chấp nhận H_0

4.2.3. Hệ số xác định và kiểm định sự phù hợp của mô hình

$$R^2 = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \ldots + \hat{\beta}_k \sum y_i x_{ki}}{\sum y_i^2}$$

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$$

Kiểm định sự phù hợp của mô hình tức là kiểm định giả thiết đồng thời bằng không:

H_0: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$; ($H_1$: ít nhất 1 trong k tham số khác 0)

$$F = \frac{R^2(n-k)}{(1-R^2)(k-1)}$$

Nguyên tắc quyết định:

Nếu $F > F_\alpha(k-1, n-k)$: Bác bỏ H_0: Mô hình phù hợp

Nếu $F \leq F_\alpha(k-1, n-k)$: Chấp nhận H_0: Mô hình không phù hợp

KINH TẾ LƯỢNG

CHƯƠNG V HỒI QUY VỚI BIẾN GIẢ

5.1. Sử dụng biến giả trong mô hình hồi quy

Ví dụ 5.1: Xét sự phụ thuộc của thu nhập (Y) (triệu đồng/tháng) vào thời gian công tác (X) (năm) và nơi làm việc của người lao động (DNNN và DNTN).

Z = 1: làm trong DNNN và Z = 0: làm trong DNTN

Trong đó Y và X là biến số lượng, còn Z là chỉ tiêu chất lượng cho biết có hay không một thuộc tính nào đó. Z được gọi là biến giả trong mô hình

$E(Y/X,Z) = \beta_1 + \beta_2 X_i + \beta_3 Z_i$ (5.1)

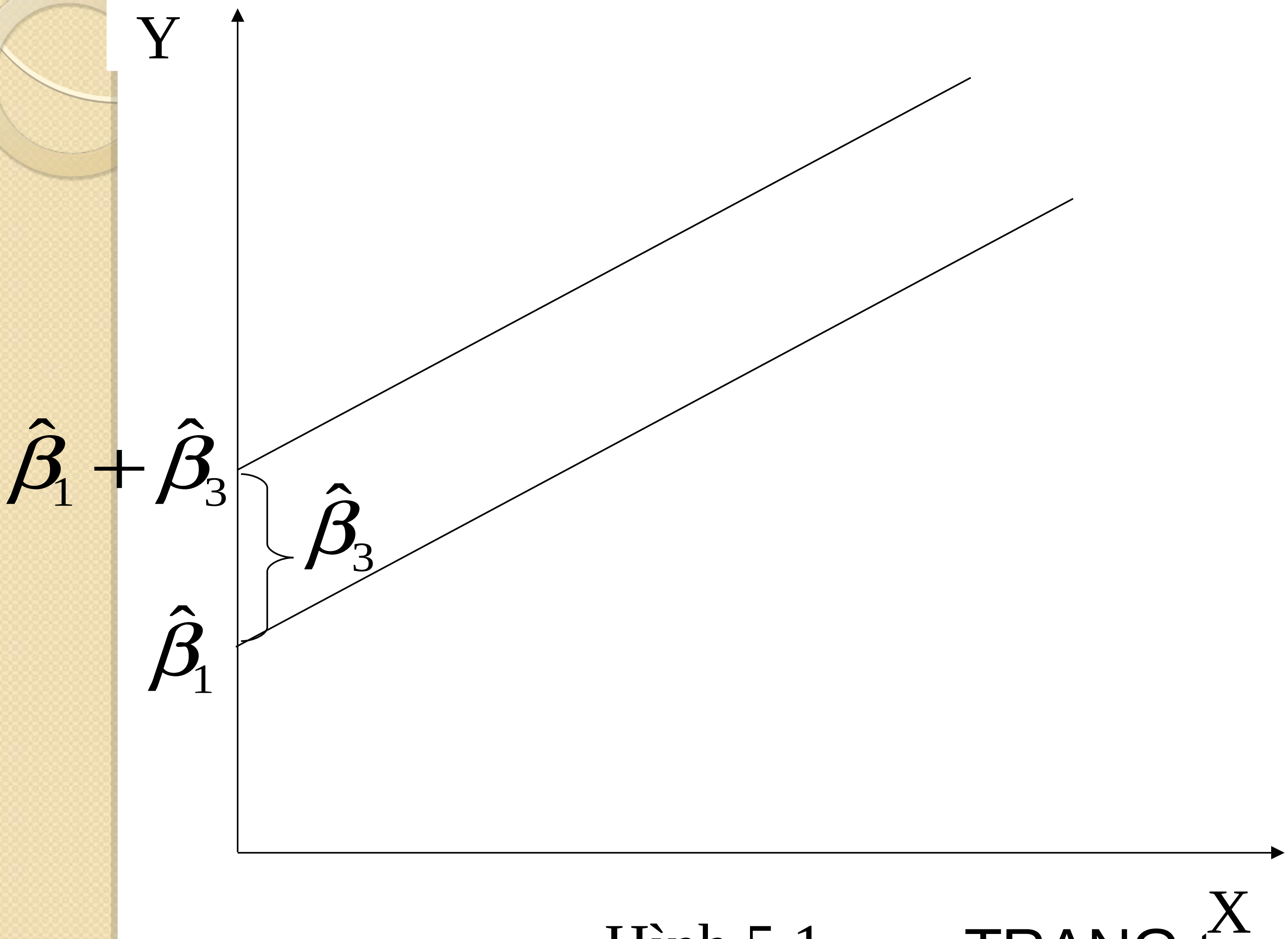
$E(Y/X,Z=0) = \beta_1 + \beta_2 X_i$ (5.2)

$E(Y/X,Z=1) = \beta_1 + \beta_2 X_i + \beta_3$ (5.3)

(5.2): mức thu nhập bình quân tháng của người lao động tại DNTN khi có thời gian công tác là X năm.

$\hat{\beta}_3 = 0,4$: 2 người có cùng thời gian công tác thì trung bình mức thu nhập của người làm tại DNNN cao hơn người làm tại DNTN 0,4 triệu đồng/tháng.

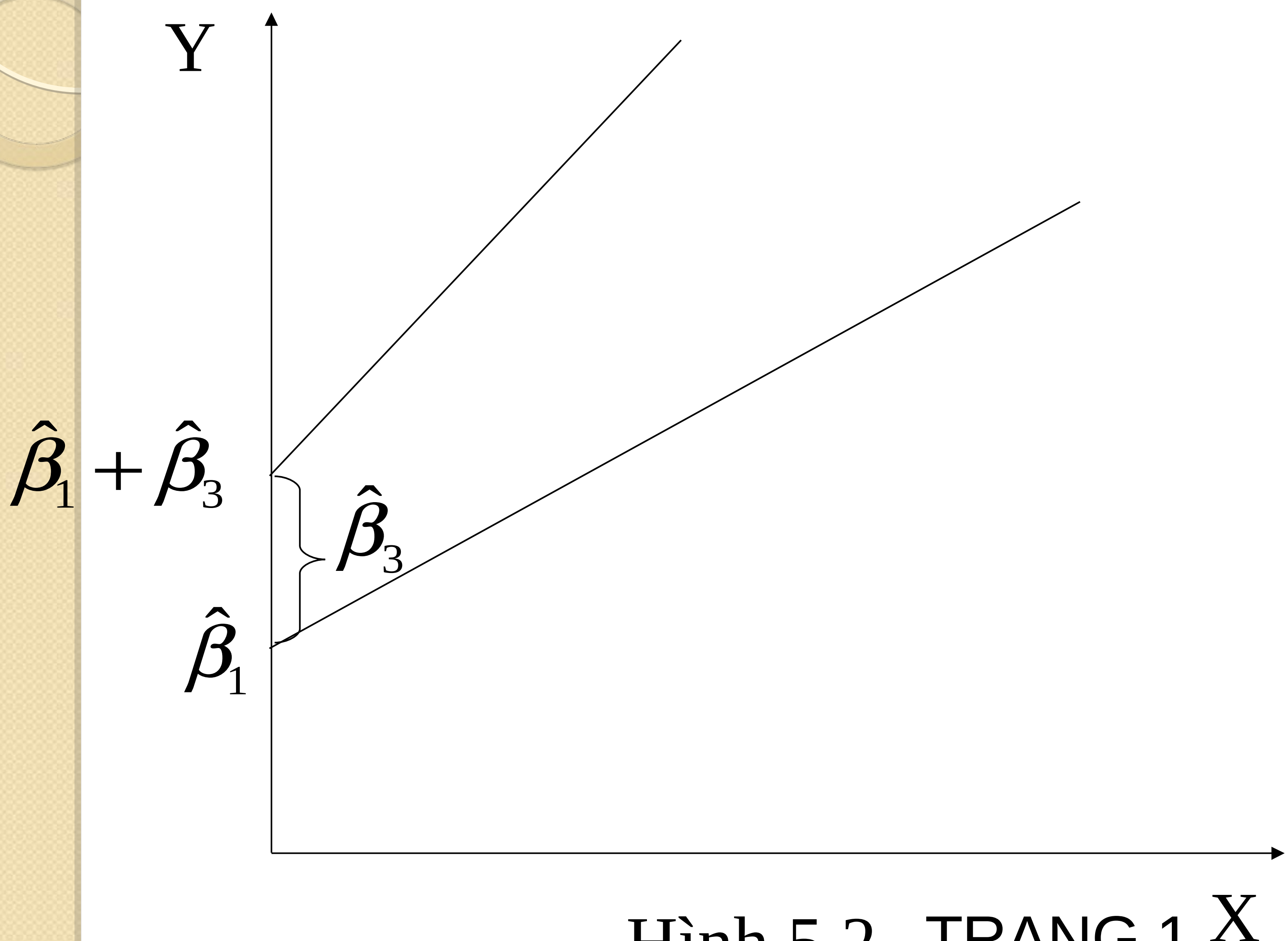
$$E(Y/X,Z) = \beta_1 + \beta_2 X_i + \beta_3 Z_i$$



Hình 5.1 TRANG 1

99

$$E(Y/X,Z) = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \beta_4 X_i Z_i$$



Hình 5.2 TRANG 1

Ví dụ 5.2: Xét sự phụ thuộc của thu nhập (Y) (triệu đồng/tháng) vào thời gian công tác (X) (năm) và nơi làm việc của người lao động (DNNN, DNTN và DNLD)

$$\left.\begin{aligned} Z_{1i} &= 0 \\ Z_{2i} &= 0 \end{aligned}\right\} \quad \text{phạm trù cơ sở}$$

Để lượng hoá chỉ tiêu chất lượng trên, ta phải dùng 2 biến giả Z_1 và Z_2.

$$Z_{1i} = \begin{cases} 1 \in DNNN \\ 0 \notin DNNN \end{cases} \qquad Z_{2i} = \begin{cases} 1 \in DNTN \\ 0 \notin DNTN \end{cases}$$

$E(Y/X,Z1,Z2) = \beta_1 + \beta_2 X_i + \beta_3 Z_{1i} + \beta_4 Z_{2i}$

$E(Y/X,Z1=0,Z2=0) = \beta_1 + \beta_2 X_i$

$E(Y/X,Z1=1,Z2=0) = \beta_1 + \beta_2 X_i + \beta_3$

$E(Y/X,Z1=0,Z2=1) = \beta_1 + \beta_2 X_i + \beta_4$

$\hat{\beta}_3 = 0,4$: 2 người có cùng thời gian công tác thì trung bình mức thu nhập của người làm tại DNNN cao hơn người làm tại DNLD 0,4 triệu đồng/tháng.

$\hat{\beta}_4 = -0,2$: 2 người có cùng thời gian công tác thì trung bình mức thu nhập của người làm tại DNTN thấp hơn người làm tại DNLD 0,2 triệu đồng/tháng.

Lưu ý: *Một chỉ tiêu chất lượng có* ***m*** *phạm trù khác nhau thì ta phải dùng* ***m-1*** *biến giả để lượng hoá cho chỉ tiêu chất lượng đó.*

Ví dụ 5.3. tiếp ví dụ 5.2, thu nhập còn phụ thuộc vào trình độ người lao động (từ đại học trở lên, cao đẳng và khác)

$$D_{1i} = \begin{cases} 1: \text{nếu trình độ từ đại học trở lên} \\ 0: \text{nếu không} \end{cases}$$

$$D_{2i} = \begin{cases} 1: \text{nếu trình độ cao đẳng} \\ 0: \text{nếu không có trình độ cao đẳng} \end{cases}$$

Tổng quát: số biến giả đưa vào mô hình phụ thuộc vào số biến định tính và số phạm trù có ở mỗi biến định tính. Số biến giả đưa vào mô hình có thể được xác định theo công thức sau:

$$n = \sum_{i=1}^{k} (n_i - 1)$$

Trong đó: n – số biến giả đưa vào mô hình; k – số biến định tính; n_i – số phạm trù của biến định tính thứ i.

5.2. Sử dụng biến giả trong phân tích mùa

Z = 1, nếu quan sát trong mùa, và Z=0 nếu quan sát không nằm trong mùa.

Từ tháng 1-6: trong mùa, Tháng 7-12: ngoài mùa.

Y: chi tiêu cho quần áo, X: thu nhập khả dụng

- Nếu yếu tố mùa chỉ ảnh hưởng đến hệ số chặn

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i$$

- Nếu yếu tố mùa có ảnh hưởng đến hệ số góc thì

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i + \hat{\beta}_4 X_i Z_i$$

Mô hình sau có tính tổng quát hơn. Thông qua việc kiểm định giả thiết chúng ta sẽ biết được hệ số góc nào có ý nghĩa.

5.3. Kiểm định sự ổn định cấu trúc của các mô hình hồi quy bằng biến giả

Ví dụ 5.4. Cho số liệu tiết kiệm và thu nhập cá nhân ở nước Anh từ 1946-63 (triệu pounds)

TK I	Tiết kiệm	Thu nhập	TK II	Tiết kiệm	Thu nhập
1946	0.36	8.8	1955	0.59	15.5
1947	0.21	9.4	1956	0.9	16.7
1948	0.08	10	1957	0.95	17.7
1949	0.2	10.6	1958	0.82	18.6
1950	0.1	11	1959	1.04	19.7
1951	0.12	11.9	1960	1.53	21.1
1952	0.41	12.7	1961	1.94	22.8
1953	0.5	13.5	1962	1.75	23.9
1954	0.43	14.3	1963	1.99	25.2

Hàm tiết kiệm

Thời kỳ tái thiết: 1946-54

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$

Thời kỳ hậu tái thiết

$$Y_i = \lambda_1 + \lambda_2 X_i + \varepsilon_i$$

có các trường hợp sau xảy ra:

$$\begin{cases} \alpha_1 = \lambda_1 \\ \alpha_2 = \lambda_2 \end{cases} \quad \begin{cases} \alpha_1 = \lambda_1 \\ \alpha_2 \neq \lambda_2 \end{cases} \quad \begin{cases} \alpha_1 \neq \lambda_1 \\ \alpha_2 = \lambda_2 \end{cases} \quad \begin{cases} \alpha_1 \neq \lambda_1 \\ \alpha_2 \neq \lambda_2 \end{cases}$$

Chúng ta kiểm tra xem hàm tiết kiệm có bị thay đổi cấu trúc giữa 2 thời kỳ hay không. Chúng ta xét hàm tiết kiệm tổng quát của cả 2 thời kỳ:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{\beta}_3 Z_i + \hat{\beta}_4 X_i Z_i + e_i$$

Với $n = n_1 + n_2$

Trong đó Z = 1: quan sát thuộc thời kỳ tái thiết

Z = 0 : quan sát thuộc thời kỳ hậu tái thiết

* Kiểm định giả thiết H_0: $\beta_3=0$

Nếu chấp nhận H_0: loại bỏ Z ra khỏi mô hình

* Kiểm định giả thiết H_0: $\beta_4=0$

Nếu chấp nhận H_0: loại bỏ Z_iX_i ra khỏi mô hình

Từ số liệu ở bảng ta có kết quả hồi quy theo mô hình như sau:

$$Y_i = -1{,}75 + 0{,}15045 X_i + 1{,}4839 Z_i - 0{,}1034 X_i Z_i + e_i$$

t =	(-5,27)	(9,238)	(3,155)	(-3,109)
p_t =	(0,000)	(0,000)	(0,007)	(0,008)

Kết quả trên cho thấy cả tung độ gốc và hệ số góc chênh lệch đều có ý nghĩa thống kê. Điều đó chứng tỏ rằng các hồi quy trong hai thời kỳ là khác nhau.

Từ kết quả trên, chúng ta có thể tính hồi quy cho 2 thời kỳ như sau:

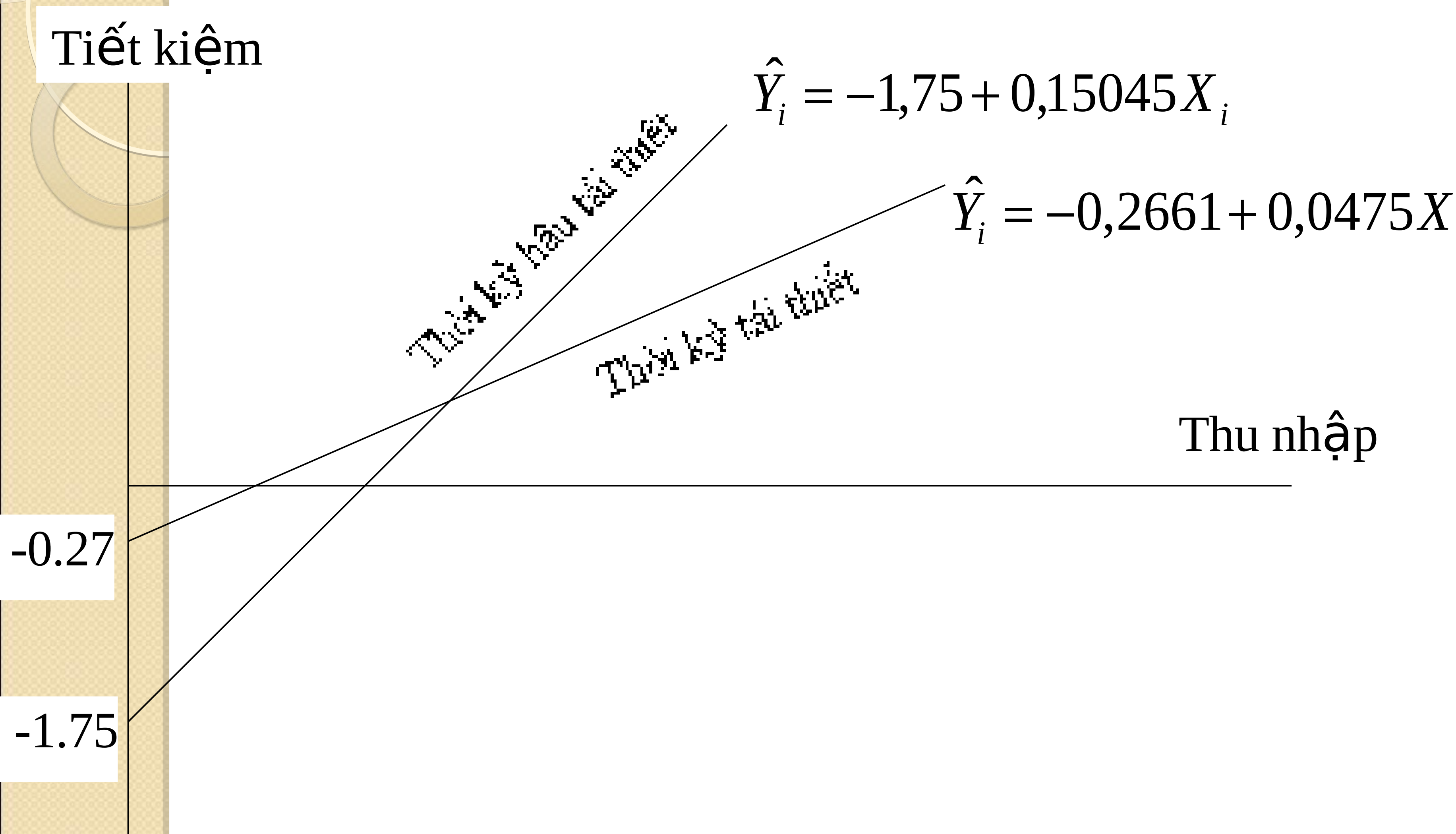
Thời kỳ tái thiết: Z = 1

$$Y_i = -1{,}75 + 0{,}15045 X_i + 1{,}4839 - 0{,}1034 X_i + e_i$$

$$Y_i = -0{,}2661 + 0{,}0475 X_i + ei$$

Thời kỳ hậu tái thiết: Z = 0

$$Y_i = -1{,}75 + 0{,}15045 X_i + e_i$$



Tiết kiệm
$\hat{Y}_i = -1,75 + 0,15045 X_i$
$\hat{Y}_i = -0,2661 + 0,0475 X_i$
Thời kỳ hậu tái thiết
Thời kỳ tái thiết
Thu nhập
-0.27
-1.75

5.4. Hàm tuyến tính từng khúcYX

Ví dụ 5.5: Sản lượng dưới X*, thì chi phí hoa hồng sẽ khác với khi sản lượng trên X*.

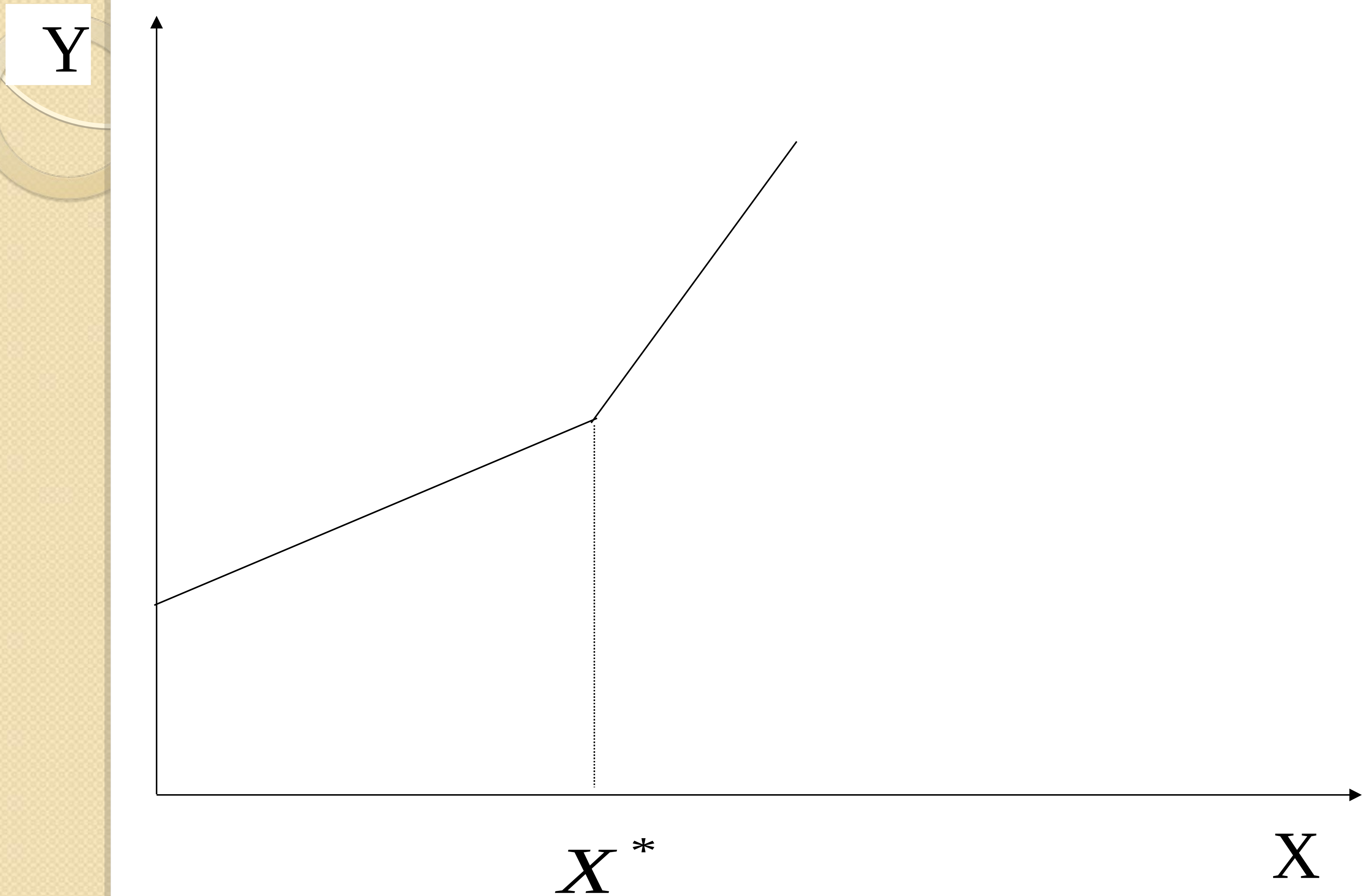
Hàm hồi quy sẽ có dạng:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i - X^*) Z_i + u_i$$

Y: Chi phí; X: sản lượng;

X*: giá trị ngưỡng sản lượng

$$Z_{1i} = \begin{cases} 1: X_i > X^* \\ 0: X_i \le X^* \end{cases}$$


Y
X
X *

Trong đó tổng SL làm thay đổi độ dốc (X*) là 5500 tấn

CP	256	414	634	778	1003
SL	1000	2000	3000	4000	5000
CP	1839	2081	2423	2734	2914
SL	6000	7000	8000	9000	10000

Ta có kết quả hồi quy như sau:

$$Y_i = -145,717 + 0,279X_i + 0,095(X_i - X^*)Z_i + e_i$$

t = (-0,824) (6,607) (1,145)

R^2 = 0,9737 X* = 5500

114

Lưu ý: Nếu biến phụ thuộc là biến giả:

Nếu ta có một biến phụ thuộc là biến giả tức là biến chỉ nhận hai giá trị 0 và 1. Chúng ta không thể sử dụng phương pháp bình phương bé nhất (OLS) để ước lượng hàm hồi quy mà phải dùng các phương pháp khác để ước lượng như:

-Mô hình xác suất tuyến tính (LPM)

-Mô hình Logit (Logit model)

-Mô hình Probit (Probit model)

-Mô hình Tobit (Tobit model)

CHƯƠNG VI
ĐA CỘNG TUYẾN

6.1. Bản chất của đa cộng tuyến

Khi lập mô hình hồi quy bội

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$$

Có sự phụ thuộc tuyến tính cao giữa các biến giải thích gọi là đa cộng tuyến.

a. Đa cộng tuyến hoàn hảo

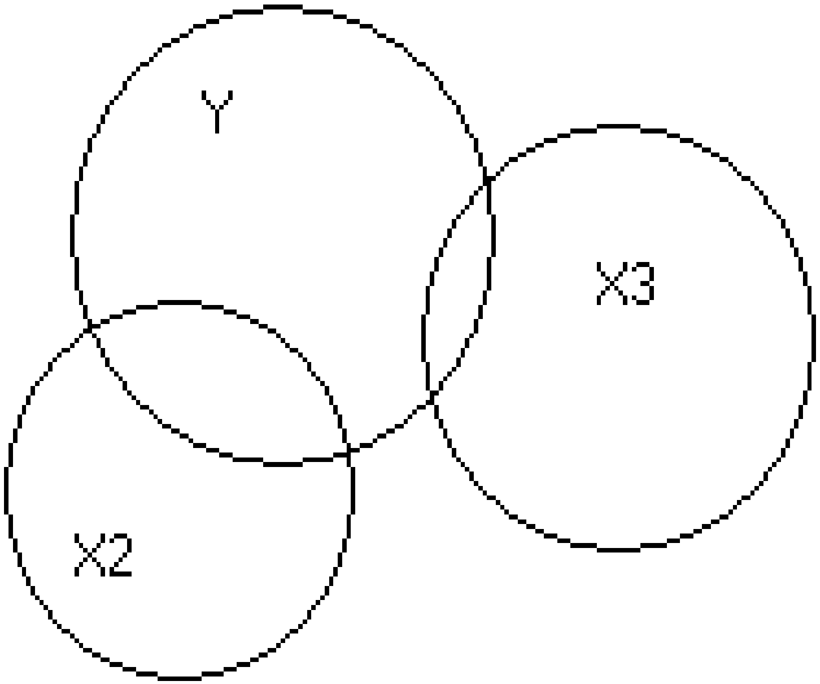
Tồn tại λ_2, λ_3,… λ_k không đồng thời bằng 0 sao cho

$$\lambda_2 X_2 + \lambda_3 X_3 + \ldots + \lambda_k X_k = 0$$

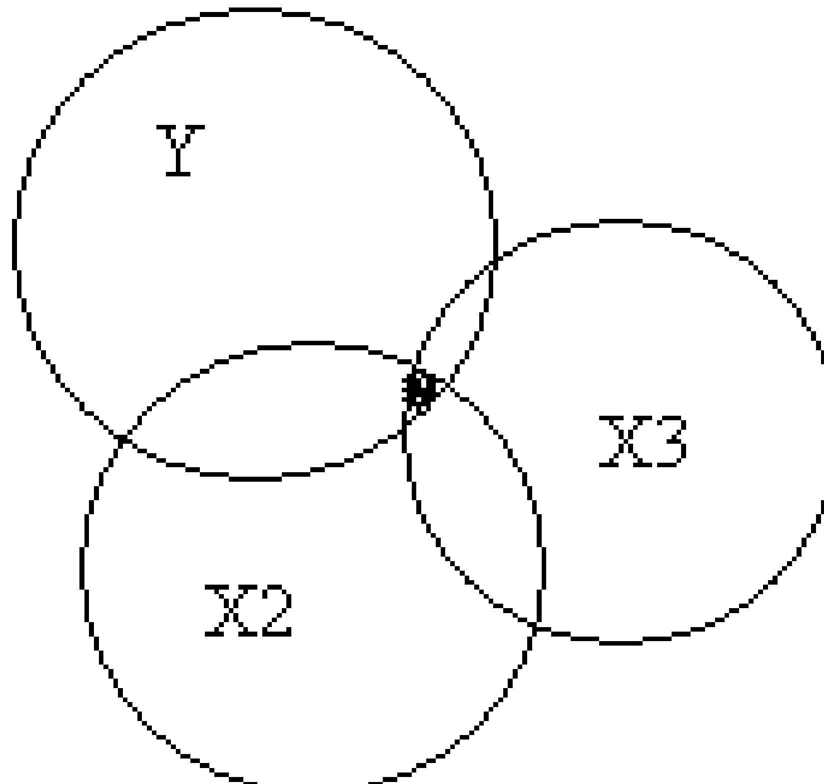
b. Đa cộng tuyến không hoàn hảo

$$\lambda_2 X_2 + \lambda_3 X_3 + \ldots + \lambda_k X_k + v_i = 0$$

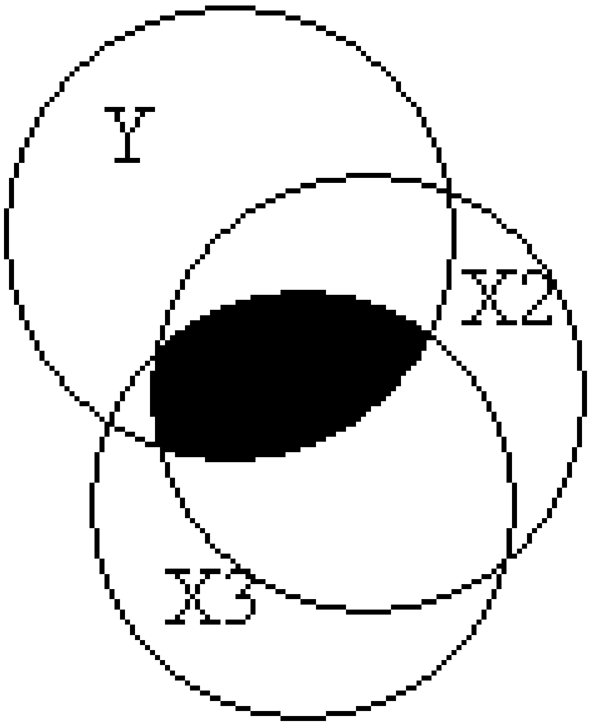
không có đa cộng tuyến



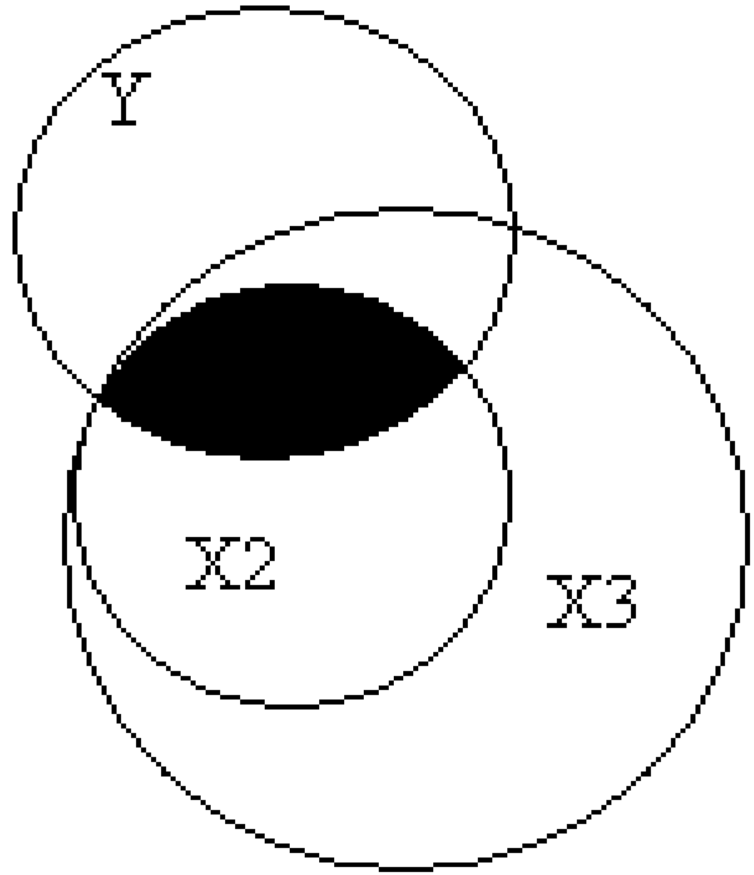
đa cộng tuyến thấp



đa cộng tuyến rất cao



đa cộng tuyến hoàn hảo



6.2. Ước lượng các tham số khi có đa cộng tuyến

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i} \sum x_{3i}^2 - \sum y_i x_{3i} \sum x_{2i} x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - (\sum x_{2i} x_{3i})^2}$$

Nếu $X_{2i} = \lambda X_{3i}$ => $x_{2i} = \lambda x_{3i}$

=> $$\hat{\beta}_2 = \frac{\lambda \sum y_i x_{3i} \sum x_{3i}^2 - \lambda \sum y_i x_{3i} \sum x_{3i} x_{3i}}{\lambda^2 \sum x_{3i}^2 \sum x_{3i}^2 - \lambda^2 \sum x_{3i}^2 \sum x_{3i}^2} = \frac{0}{0}$$

=> không xác định được $\hat{\beta}_2, \hat{\beta}_3$

Một số nguyên nhân gây ra hiện tượng đa cộng tuyến

- Khi chọn các biến độc lập mối quan có quan hệ nhân quả hay có tương quan cao vì đồng phụ thuộc vào một điều kiện khác.
- Khi số quan sát nhỏ hơn số biến độc lập.
- Cách thu thập mẫu.
- Chọn biến X_i có độ biến thiên nhỏ.

6.3. Hậu quả của đa cộng tuyến

- Ước lượng các hệ số không hiệu quả do phương sai của ước lượng lớn.
- Khoảng tin cậy của các ước lượng rộng
- Tỷ số t_i không có ý nghĩa
- R^2 lớn nhưng t nhỏ
- Các ước lượng OLS và sai số chuẩn của chúng trở nên rất nhạy với những thay đổi nhỏ của dữ liệu
- Dấu các ước lượng của các hệ số hồi quy có thể sai
- Thêm vào hay bớt đi các biến cộng tuyến với các biến khác, mô hình sẽ thay đổi về dấu hoặc thay đổi về độ lớn của các ước lượng

6.4. Cách phát hiện đa cộng tuyến

6.4.1. R^2 lớn nhưng tỷ số t nhỏ

6.4.2. Tương quan cặp giữa các biến giải thích cao

$$r_{XZ} = \frac{\sum (X_i - \bar{X})(Z_i - \bar{Z})}{\sqrt{\sum (X_i - \bar{X})^2 (Z_i - \bar{Z})^2}}$$

Trong đó X, Z là 2 biến giải thích trong mô hình

6.4.3. Sử dụng mô hình hồi quy phụ

$$\hat{X}_{2i} = \hat{\beta}_1 + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{mi}$$

H_0: $R^2 = 0$

$$F = \frac{R^2(n - m)}{(1 - R^2)(m - 1)}$$

Nếu $F > F_\alpha(m-1,n-m)$: bác bỏ H_0 => có đa cộng tuyến

Nếu $F < F_\alpha(m-1,n-m)$: chấp nhận H_0 => không có đa cộng tuyến

6.4.4. Sử dụng nhân tử phóng đại phương sai (VIF)

Đối với hàm hồi quy 2 biến giải thích, VIF được định nghĩa như sau:

$$VIF = \frac{1}{(1 - r_{23}^2)}$$

Đối với trường hợp tổng quát, có (k-1) biến giải thích thì:

$$VIF = \frac{1}{(1 - R_j^2)}$$

R^2_j: là giá trị R^2 trong hàm hồi quy của X_j theo (k-1) biến giải thích còn lại.

Thông thường khi VIF > 10, thì biến này được coi là có cộng tuyến cao.

6.5. Biện pháp khắc phục

6.5.1. Dùng thông tin tiên nghiệm

Ví dụ khi hồi quy mô hình sản xuất Cobb-Douglas

$$Y_i = A L_i^{\beta_3} K_i^{\beta_2} e^{u_i}$$

$$Ln(Y_i) = \beta_1 + \beta_2 ln(K_i) + \beta_3 ln(L_i) + u_i$$

Có thể gặp hiện tượng đa cộng tuyến do K và L cùng tăng theo quy mô sản xuất. Nếu ta biết là hiệu suất không đổi theo quy mô tức là $\beta_2 + \beta_3 = 1$.

$$Ln(Y_i) = \beta_1 + \beta_2 ln(K_i) + (1-\beta_2) ln(L_i) + u_i$$

$$Ln(Y_i) - Ln(L_i) = \beta_1 + \beta_2 [ln(K_i) - ln(L_i)] + u_i$$

=> mất đa cộng tuyến (vì đây là mô hình hồi quy đơn).

6.5.2. Loại trừ một biến giải thích ra khỏi mô hình

B1: Xem cặp biến giải thích nào có quan hệ chặt chẽ

B2: Tính R^2 đối với các hàm hồi quy: có mặt cả 2 biến; không có mặt một trong 2 biến

B3: Loại biến mà giá trị R^2 tính được khi không có mặt biến đó là lớn hơn.

6.5.3. Bổ sung thêm dữ liệu hoặc chọn mẫu mới

6.5.4. Dùng sai phân cấp 1

(Phương pháp này chỉ áp dụng cho chuỗi thời gian)

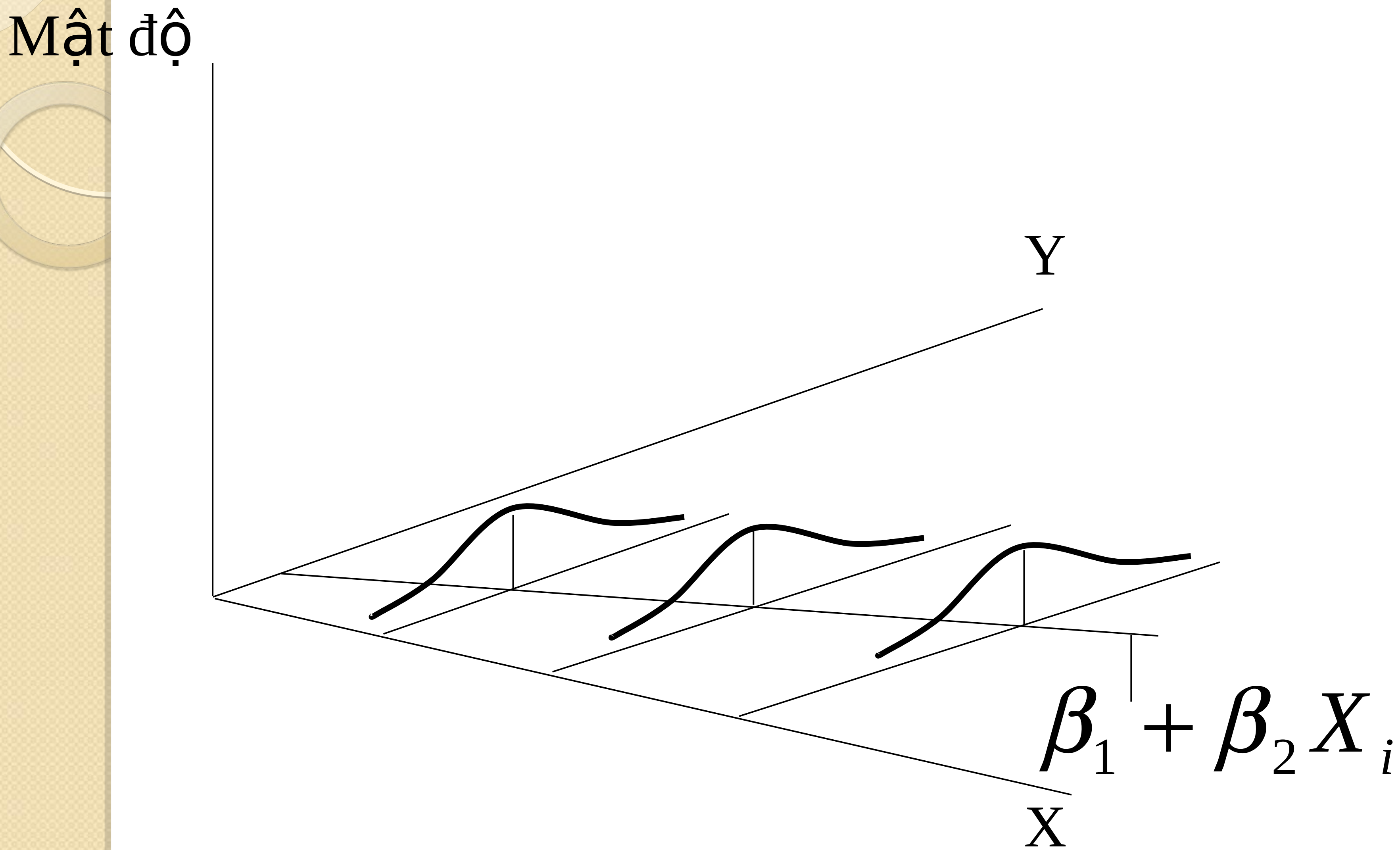
Ví dụ 6.1. xem xét đa cộng tuyến trong mô hình từ số liệu ở file "vi du 6.1 - da cong tuyen"

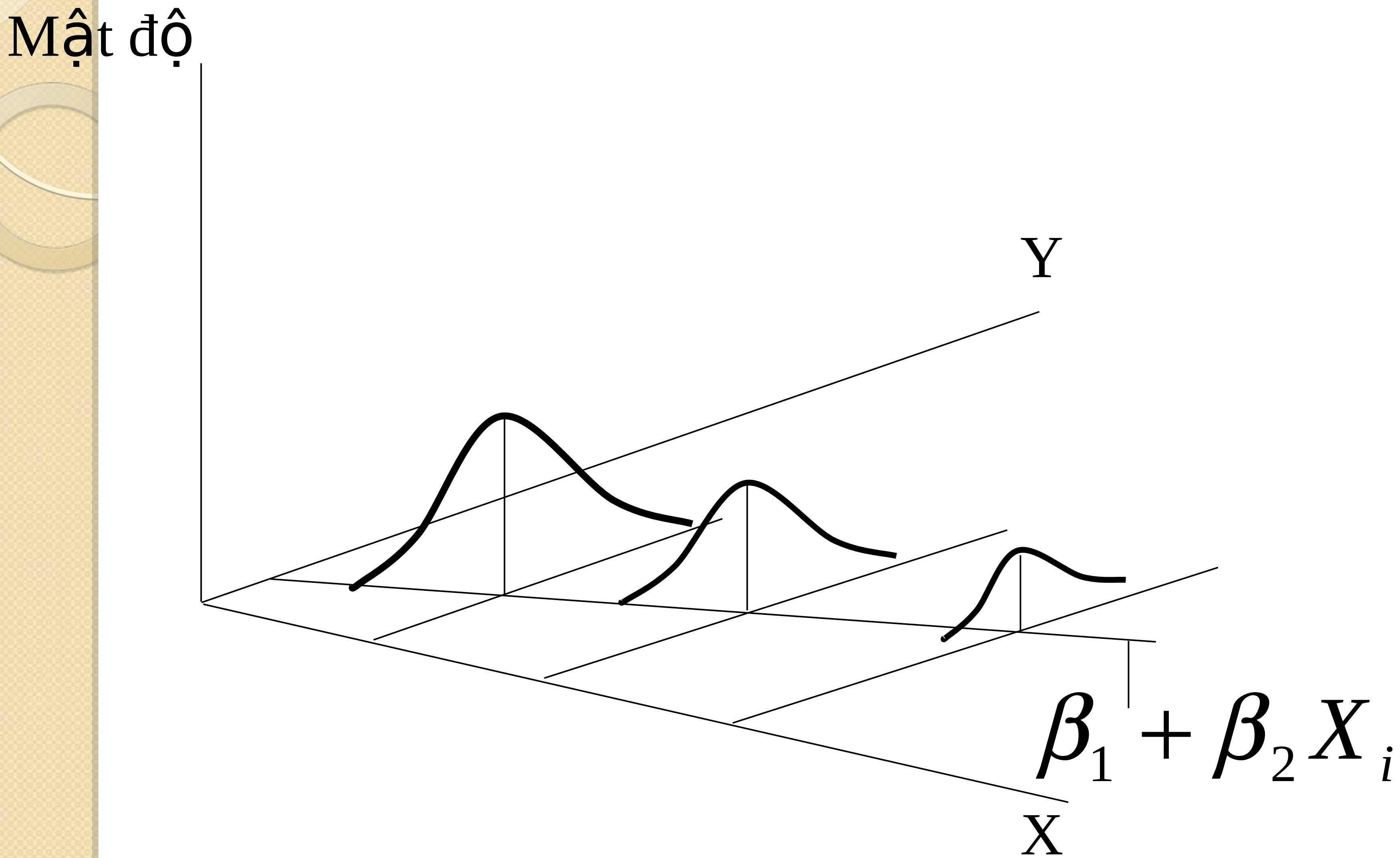
KINH TẾ LƯỢNG

CHƯƠNG VII PHƯƠNG SAI THAY ĐỔI

7.1. Bản chất của phương sai thay đổi

Giả định của mô hình hồi quy tuyến tính cổ điển là phương sai của sai số hồi quy không đổi qua các quan sát. Trong thực tế sai số hồi quy có thể tăng lên hoặc giảm đi khi giá trị biến độc lập X tăng lên => Phương sai thay đổi.


Mật độ
Y
$\beta_1 + \beta_2 X_i$
X



131

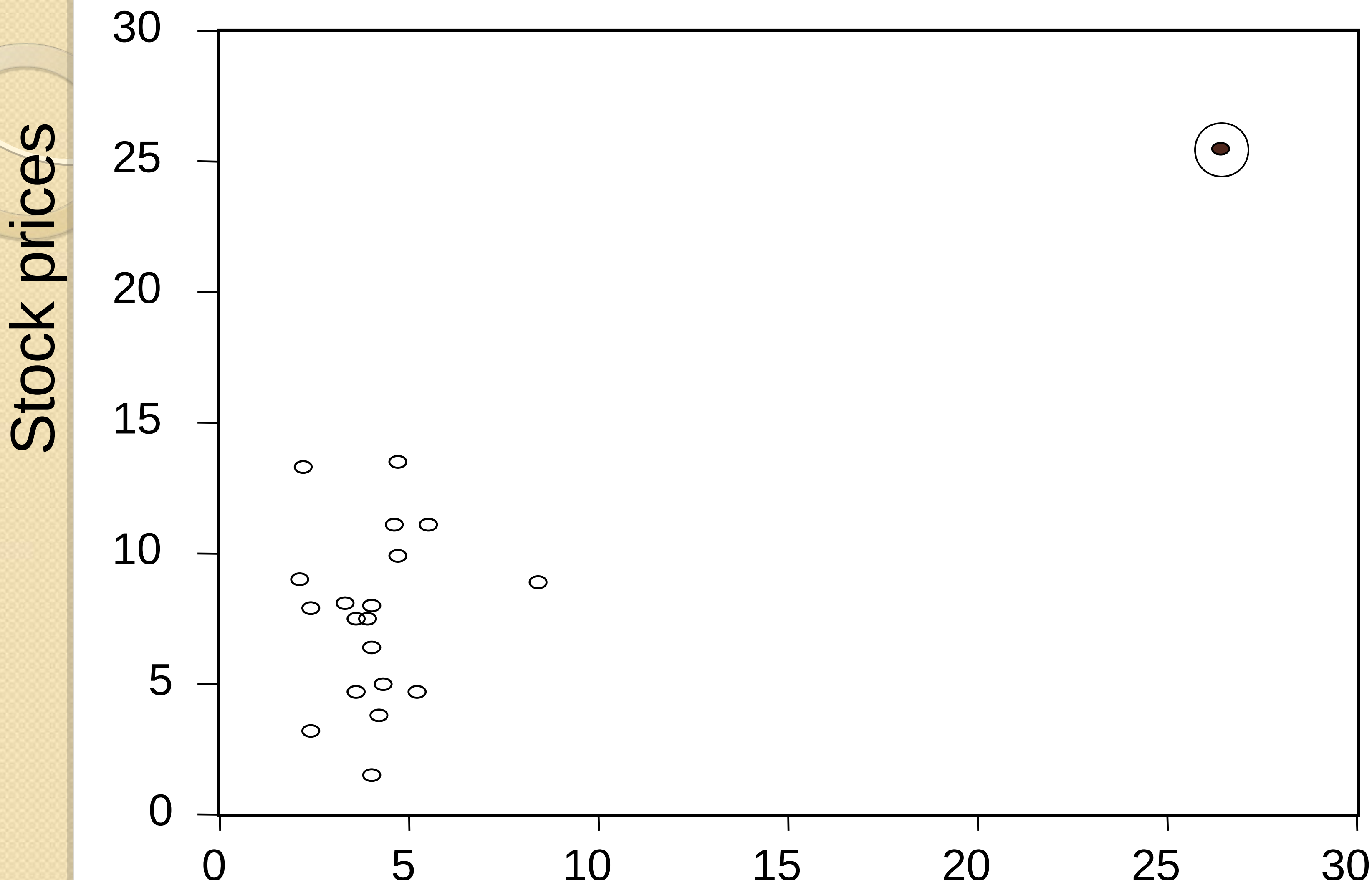
Nguyên nhân phương sai không đồng nhất:

-Gọi Y là số phế phẩm trong 100 sản phẩm của một thợ học việc, X là số giờ thực hành. Khi số giờ thực hành càng lớn thì số phế phẩm càng nhỏ và càng ít biến động. Chúng ta có trường hợp phương sai giảm dần khi X tăng dần.

- Khi thu nhập (X) tăng thì chi tiêu cho các mặt hàng xa xỉ tăng và mức biến động càng lớn. Chúng ta có trường hợp phương sai tăng dần khi X tăng dần.

- Khi cải thiện phương pháp thu thập số liệu thì phương sai giảm.

- Phương sai của sai số tăng do sự xuất hiện của điểm nằm ngoài, đó là các trường hợp bất thường với dữ liệu rất khác biệt (rất lớn hoặc rất nhỏ so với các quan sát khác).
- Phương sai thay đổi khi không xác đúng dạng mô hình, nếu một biến quan trọng bị bỏ sót thì phương sai của sai số lớn và thay đổi. Tình trạng này giảm hẳn khi đưa biến bị bỏ sót vào mô hình.



Source: Gujarati, 1995, p.397

TRANG 1

7.2. Hệ quả của phương sai thay đổi khi sử dụng ước lượng OLS

- Các ước lượng bình phương bé nhất vẫn là ước lượng không chệch nhưng không phải là ước lượng hiệu quả (ước lượng có phương sai nhỏ nhất).
- Ước lượng của các phương sai sẽ bị chệch, do đó các kiểm định mức ý nghĩa và khoảng tin cậy dựa theo phân phối t và F không còn đáng tin cậy nữa.

7.3. Ước lượng bình phương tối thiểu có trọng số (WLS) (SGK)

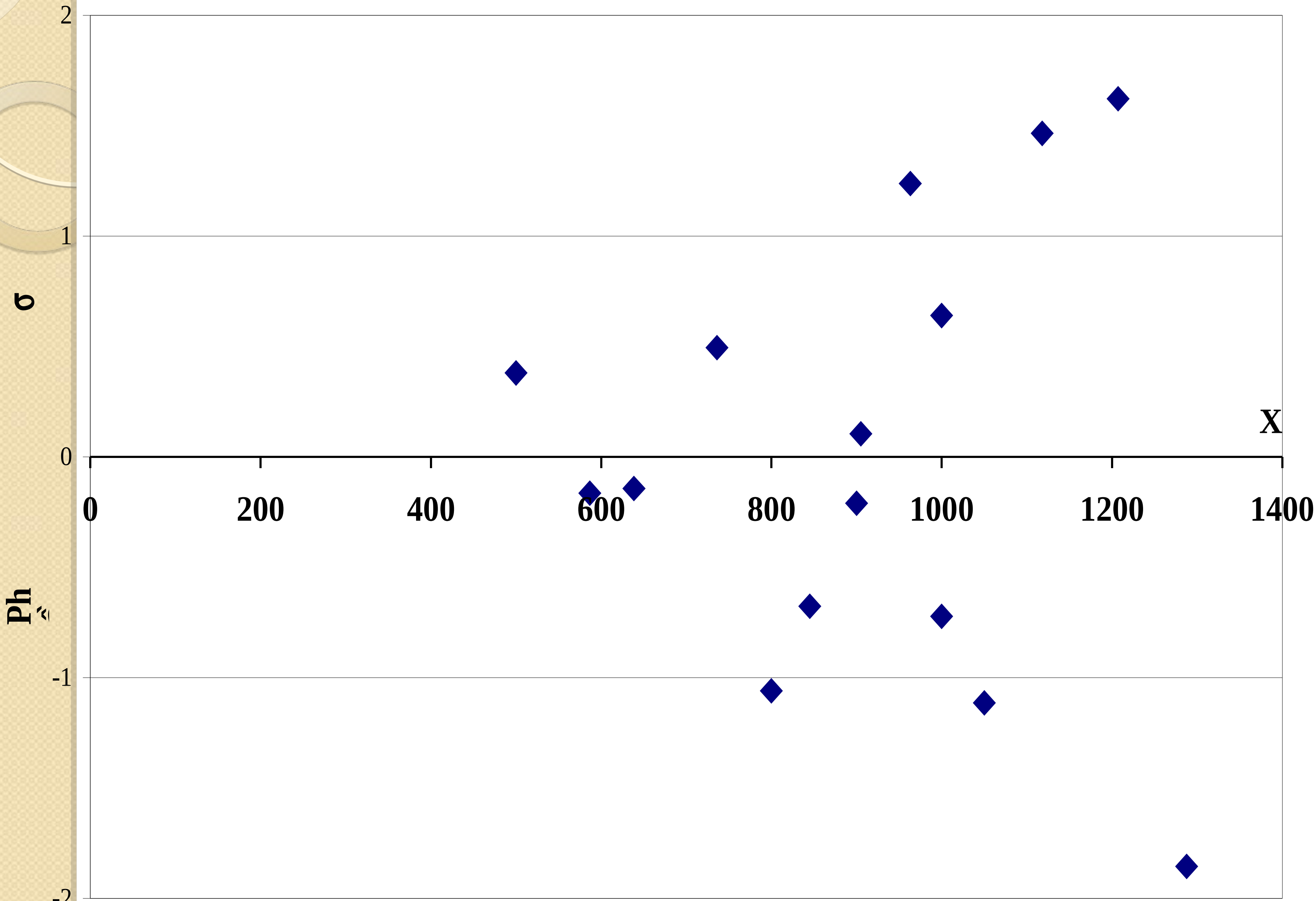
7.4. Cách phát hiện

7.4.1. Bản chất của vấn đề nghiên cứu

Nghiên cứu dữ liệu chéo về chi phí và sản lượng của các doanh nghiệp có quy mô khác nhau.

7.4.2. Phương pháp đồ thị

Xét đồ thị của phần dư theo giá trị Y hoặc X.


2
1
0
-1
-2
σ
Ph
0
200
400
600
800
1000
1200
1400
X

7.4.3. Kiểm định Park

B1: Ước lượng hồi quy gốc dù có tồn tại phương sai thay đổi.

B2: Tính $\mathbf{Lne^2_i}$ từ $\mathbf{e_i}$ của mô hình hồi quy gốc

B3: Ước lượng mô hình: $\mathbf{Lne^2_i = \beta_1 + \beta_2 LnX_i + v_i}$

X_i là biến giải thích của mô hình hồi quy gốc. Trong mô hình đa biến sẽ tiến hành hồi quy $\mathbf{Lne^2_i}$ theo từng biến X_i, hoặc có thể sử dụng Yi-hat làm biến giải thích.

B4: Kiểm định giả thiết H_0: $\beta_2=0$: Không có hiện tượng phương sai thay đổi.

VD: Dữ liệu Hete-Park_Glejser test, Có sự liên hệ

7.4.4. Kiểm định Glejser

B1: Ước lượng hồi quy gốc dù có tồn tại phương sai thay đổi.

B2: Ước lượng các mô hình:

$$|e_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|e_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|e_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

X_i là biến giải thích của mô hình hồi quy gốc. Trong mô hình đa biến sẽ tiến hành hồi quy |ei| theo từng biến Xi.

B3: Kiểm định giả thiết H_0: $\beta_2=0$: Không có hiện tượng phương sai thay đổi.

VD: Dữ liệu Hete-Park_Glejser test, có hiện tượng phương sai thay đổi do chúng ta bác bỏ H_0 trong 2 trường hợp sau:

$$|e_i| = -0.17 + 0.046 X_i + v_i$$

$$|e_i| = -1.07 + 0.423 \sqrt{X_i} + v_i$$

7.4.5. Kiểm định White

Xét mô hình hồi quy 3 biến:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i$$

Bước 1: Ước lượng phương trình trên, thu được e_i

Bước 2: Ước lượng mô hình sau:

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$$

Phương trình trên có thể có số mũ cao hơn và nhất thiết phải có hệ số chặn bất kể mô hình hồi quy gốc có hệ số chặn hay không. R^2 là hệ số xác định thu được từ phương trình trên.

Bước 3: Kiểm định giả thiết H_0: Phương sai của sai số không đổi.

- Nếu $n.R^2 < \chi^2$ với bậc tự do p-1 (hệ số của mô hình trên) => chấp nhận H_0.

-Nếu $n.R^2 \geq \chi^2$: Bác bỏ H_0, tức phương sai của sai số thay đổi.

Ví dụ 7.1. Sử dụng file vi du 7.1–phuong sai thay doi

Từ số liệu trên, Eviews cho ta kết quả

$Y = -1.5999 + 0.409704*X_2 + 1.460808*X_3 + e_i$

Từ phương trình trên ta thu được e_i

Tiến hành hồi quy

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$$

Ta thu được kết quả:

=> $n.R^2 = 50x0.294004 = 14.7002$

Mà $\chi^2_{0.05}(5) = 11.1$ => Bác bỏ H_0, tức phương sai của sai số thay đổi.

Dependent Variable: E^2
Method: Least Squares
Date: 08/22/06 Time: 21:16
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.880149	90.51368	0.042868	0.9660
X2	-6.164398	7.181977	-0.858315	0.3954
X3	47.81948	38.04095	1.257052	0.2154
(X2)^2	0.234514	0.135770	1.727294	0.0911
(X3)^2	-7.498671	4.835527	-1.550745	0.1281
X2*X3	0.175746	0.722439	0.243267	0.8089

R-squared	0.294004	Mean dependent var	79.51277
Adjusted R-squared	0.213777	S.D. dependent var	178.5604
S.E. of regression	158.3281	Akaike info criterion	13.07938
Sum squared resid	1102982.	Schwarz criterion	13.30882
Log likelihood	-320.9846	F-statistic	3.664660
Durbin-Watson stat	2.062843	Prob(F-statistic)	0.007381

7.4.6. Kiểm định Goldfeld-Quandt

Bước 1: Sắp xếp các quan sát theo thứ tự tăng dần về giá trị của biến X.

Bước 2: Bỏ c quan sát ở giữa: c = 4 nếu n ≈ 30, c = 10 nếu n ≈ 60.

Và chia số quan sát còn lại thành 2 nhóm, mỗi nhóm có (n-c)/2 quan sát.

Bước 3: Ước lượng tham số của các hồi quy đối với (n-c)/2 quan sát đầu và quan sát cuối, thu được RSS1 và RSS2, với bậc tự do là (n-c)/2-k.

7.4.6. Kiểm định Goldfeld-Quandt (tt)

Bước 4: Tính:

$$F = \frac{RSS_2 / df}{RSS_1 / df}$$

Bước 5: Quy tắc quyết định

H_0: Phương sai của sai số không đổi.

- $F \geq F(df,df)$: Bác bỏ H_0
- $F < F(df,df)$: Chấp chấp H_0

Các kiểm định khác:

- Kiểm định tương quan hạng của Spearman
- Kiểm định Goldfeld-Quandt
- Kiểm định Breusch-Pagan-Godfrey

7.5. Biện pháp khắc phục

7.5.1. Nếu đã biết δ^2_i

Sử dụng phương pháp bình phương nhỏ nhất có trọng số

7.5.2. Nếu chưa biết δ^2_i

Xét phương trình: $Y_i = \beta_1 + \beta_2 X_i + u_i$

<u>Giả thiết 1</u>: Phương sai của sai số tỷ lệ với bình phương biến giải thích

$$E(u_i^2) = \delta^2 X_i^2$$

Chia cả hai vế của mô hình gốc cho X_i

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{Xi} = \frac{\beta_1}{X_i} + \beta_2 + v_i$$

Ta chứng minh được:

$$E(v_i^2) = E(\frac{u_i}{X_i})^2 = \frac{1}{X_i^2}E(u_i^2) = \delta^2$$

Như vậy phương trình không còn hiện tượng phương sai thay đổi là:

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + v_i$$

Lưu ý: trong phương trình trên, hệ số chặn chính là hệ số góc của mô hình hồi quy gốc, và ngược lại. Để trở lại mô hình hồi quy gốc ta phải nhân 2 vế của phương trình trên với X_i.

Giả thiết 2: Phương sai của sai số tỷ lệ với biến giải thích $E(u_i^2) = \delta^2 X_i$

Chia cả hai vế của mô hình gốc cho $\sqrt{X_i}$

$$\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + v_i$$

Và ta có:

$$E(v_i^2) = E\left(\frac{u_i}{\sqrt{X_i}}\right)^2 = \frac{1}{X_i}E(u_i^2) = \delta^2$$

Như vậy phương trình trên không còn hiện tượng phương sai thay đổi, có thể áp dụng OLS để tìm các tham số hồi quy.

Lưu ý: Phương trình trên không có hệ số tự do nên ta phải sử dụng mô hình hồi quy đi qua gốc tọa độ để ước lượng các tham số, sau đó nhân cả 2 vế với $\sqrt{X_i}$ để trở lại mô hình ban đầu.

Giả thiết 3: Phương sai của sai số tỷ lệ với bình phương giá trị trung bình của Y

$$E(u_i^2) = \delta^2 [E(Y_i)]^2$$

Ta biến đổi như sau

$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \frac{\beta_2 X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \frac{\beta_2 X_i}{E(Y_i)} + v_i$$

Và

$$E(v_i^2) = E\left(\frac{u_i}{E(Y_i)}\right)^2 = \frac{1}{[E(Y_i)]^2} E(u_i^2) = \delta^2$$

Như vậy phương trình trên không còn hiện tượng phương sai thay đổi, thỏa mãn các giả thiết của mô hình hồi quy tuyến tính cổ điển và ta có thể áp dụng OLS để tìm các tham số hồi quy.

Tuy nhiên, do $E(Y_i)$ chưa biết (vì β_1 và β_2 chưa có), chúng ta sẽ dùng ước lượng điểm của chúng là: $\hat{Y}_i$

và phương trình sẽ được viết lại là:

$$\frac{Y_i}{\hat{Y}_i} = \frac{\beta_1}{\hat{Y}_i} + \frac{\beta_2 X_i}{\hat{Y}_i} + v_i$$

Giả thiết 4: Phép biến đổi logarit

$LnY_i = \beta_1 + \beta_2 LnX_i + u_i$

Lưu ý: Phép biến đổi Logarit không dùng được nếu có 1 số giá trị của X (hoặc Y) là âm.

KINH TẾ LƯỢNG

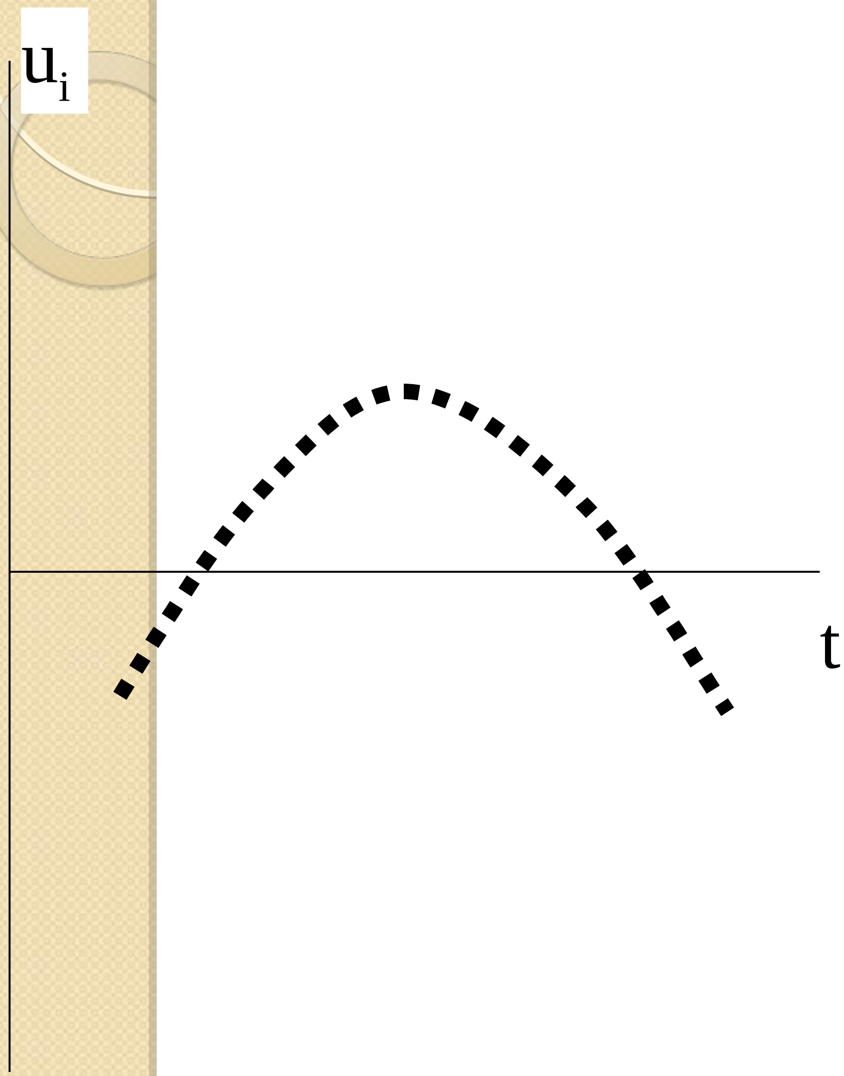
CHƯƠNG VIII
TỰ TƯƠNG QUAN – CHỌN MÔ HÌNH – THẨM ĐỊNH VIỆC CHỌN MÔ HÌNH

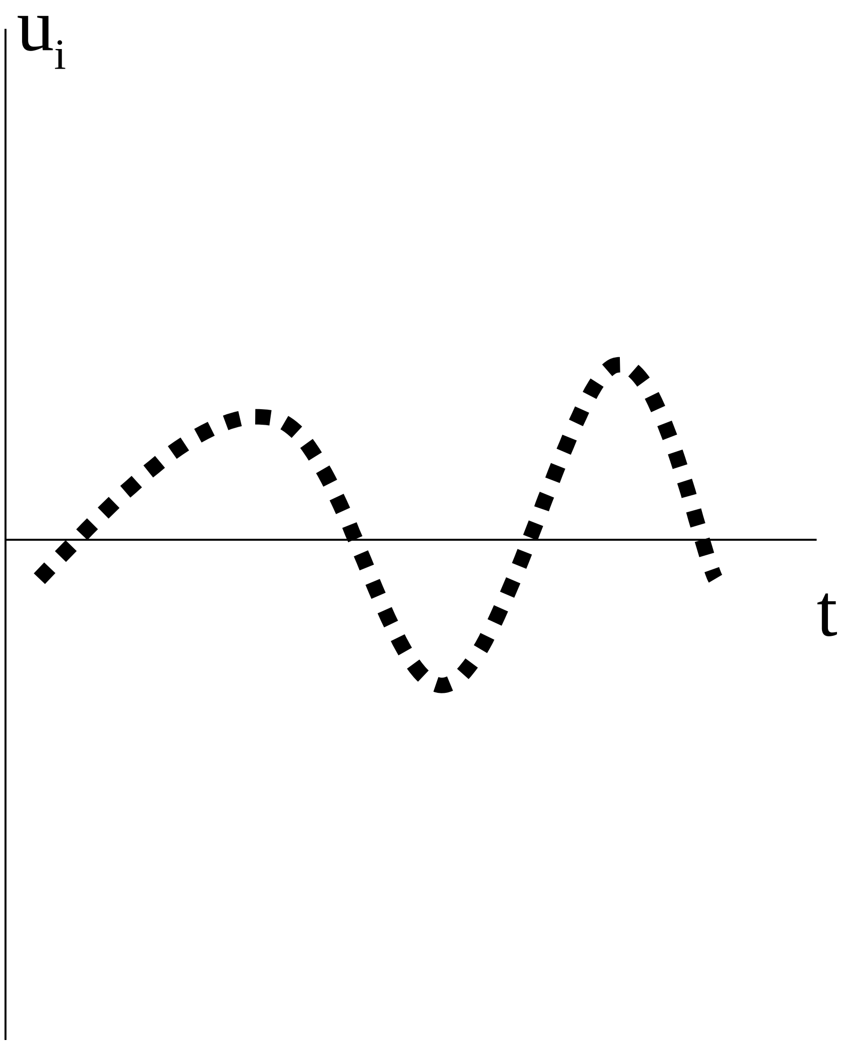
8.1. Tự tương quan (tương quan chuỗi)

8.1.1. Bản chất và nguyên nhân của tự tương quan

Trong mô hình hồi quy tuyến tính cổ điển chúng ta giả định không có tương quan giữa các phần dư hay $Cov(u_i u_j) = 0$ với mọi i, j.

$\Rightarrow Cov(u_i,u_j) \neq 0$: tự tương quan





TRANG 1

156

*** Nguyên nhân khách quan:**

- Chuỗi có tính chất quán tính theo chu kỳ
- Hiện tượng mạng nhện: dãy số cung về café năm nay phụ thuộc vào giá năm trước => u_i không còn ngẫu nhiên nữa.
- Dãy số có tính chất trễ: tiêu dùng ở thời kỳ này chẳng những phụ thuộc vào thu nhập kỳ này mà còn phụ thuộc vào tiêu dùng của kỳ trước nữa.

*** Nguyên nhân chủ quan**

- Chọn dạng mô hình sai (thường xảy ra ở mô hình với chi phí biên)
- Đưa thiếu biến giải thích vào mô hình
- Việc xử lý số liệu.(số liệu tháng = số liệu quý/3)

8.1.2. Hậu quả của tự tương quan

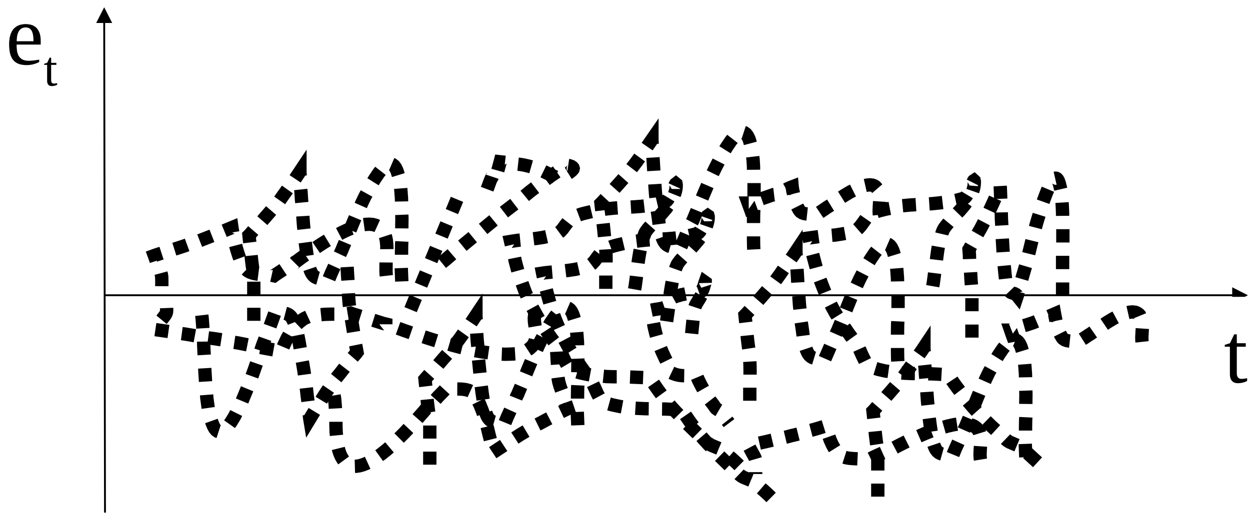
Nếu vẫn áp dụng OLS khi mô hình có hiện tượng tự tương quan thì sẽ có các hậu quả sau:

- Các ước lượng không chệch nhưng đó là không phải là các hiệu quả vì đó không phải là các ước lượng có phương sai nhỏ nhất.
- Phương sai của các ước lượng là các ước lượng chệch vì vậy các kiểm định t và F không còn hiệu quả.
- $\hat{\delta}^2$ là ước lượng chệch của δ^2
- R^2 của mẫu là ước lượng chệch (dưới) của R^2 tổng thể
- Các dư báo về Y không chính xác

8.1.3. Cách phát hiện tự tương quan

a. Đồ thị

Chúng ta có thể phát hiện hiện tượng tự tương quan bằng cách quan sát đồ thị phần dư của mô hình trên dữ liệu chuỗi thời gian.



phần dư phân bố một cách ngẫu nhiên xung quanh giá trị trung bình của nó.

b. Dùng kiểm định d của Durbin – Watson

Thống kê d của Durbin – Watson được định nghĩa như sau:

$$d = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}$$

Khi n đủ lớn thì d ≈ 2(1-ρ)

trong đó: $$\rho = \frac{\sum e_i e_{i-1}}{\sum e_i^2}$$

do -1 ≤ ρ ≤ 1, nên khi:

ρ = -1 => d = 4: tự tương quan hoàn hảo âm

ρ = 0 => d = 2: không có tự tương quan

ρ = 1 => d = 0: tự tương quan hoàn hảo dương

Giả thiết H_0	Quyết định	Nếu
Không có tự tương quan dương	Bác bỏ	$0 < d < d_L$
Không có tự tương quan dương	Không quyết định	$d_L \leq d \leq d_U$
Không có tự tương quan âm	Bác bỏ	$4\text{-}d_L < d < 4$
Không có tự tương quan âm	Không quyết định	$4\text{-}d_U \leq d \leq 4\text{-}d_L$
Không có tự tương quan âm hoặc dương	Không bác bỏ	$d_U < d < 4\text{-}d_L$

Trong đó d_U và d_L là các giá trị tra bảng giá trị d.

*** Chú ý:** trong thực tế khi tiến hành kiểm định Durbin – Watson, người ta thường áp dụng quy tắc kiểm định đơn giản sau:

Nếu $1 < d < 3$ thì kết luận mô hình không có tự tương quan.

Nếu $0 < d < 1$ thì kết luận mô hình có tự tương quan dương.

Nếu $3 < d < 4$ thì kết luận mô hình có tự tương quan âm.

Nếu d thuộc vùng chưa quyết định, chúng ta sẽ sử dụng quy tắc kiểm định cải biên như sau:

1. H_0: ρ = 0; H_1: ρ > 0. Nếu $d < d_U$ thì bác bỏ H_0 và chấp nhận H_1 (với mức ý nghĩa α), nghĩa là có tự tương quan dương.

2. H_0: ρ = 0; H_1: ρ < 0. Nếu $d > 4 - d_U$ thì bác bỏ H_0 và chấp nhận H_1 (với mức ý nghĩa α), nghĩa là có tự tương quan âm.

3. H_0: ρ = 0; H_1: ρ ≠ 0. Nếu $d < d_U$ hoặc $d > 4 - d_U$ thì bác bỏ H_0 và chấp nhận H_1 (với mức ý nghĩa 2α), nghĩa là có tự tương quan (âm hoặc dương).

c. Dùng kiểm định Breusch – Godfrey (BG)

Xét mô hình:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (8.1)$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + v_t$$

ta cần kiểm định giả thiết H_0: $\rho_1 = \rho_2 = \dots = \rho_\rho = 0$, có nghĩa là không tồn tại tự tương quan ở bất kỳ bậc nào trong số từ bậc 1 đến bậc p.

Bước 1: Ước lượng (8.1) bằng OLS, tìm phần dư e_t

Bước 2: Dùng OLS để ước lượng mô hình

$$e_t = \beta_1 + \beta_2 X_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + \varepsilon_t$$

từ đây ta thu được R^2.

Bước 3: với n đủ lớn, $(n-p)R^2$ có phân phối xấp xỉ $\chi^2(p)$.

- Nếu $(n-p)R^2 > \chi^2_\alpha(p)$: Bác bỏ H_0, nghĩa là có tự tương quan ít nhất ở một bậc nào đó.
- Nếu $(n-p)R^2 \leq \chi^2_\alpha(p)$: Chấp nhận H_0, nghĩa là không có tự tương quan.

8.1.4. Cách khắc phục

Phương pháp Durbin – Watson 2 bước để ước lượng ρ

Ước lượng mô hình $Y_t = \beta_1 + \beta_2 X_t + u_t$

Phương trình sai phân dạng tổng quát

$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \rho Y_{t-1} + u_t - \rho u_{t-1}$

Bước 1: Coi đây là phương trình hồi quy bội, hồi quy Y_t theo X_t, X_{t-1} và Y_{t-1}, và coi giá trị ước lượng được đối với hệ số hồi quy của Y_{t-1} $(=\hat{\rho})$ là ước lượng của ρ. Mặc dù là ước lượng chệch nhưng ta có ước lượng vững của ρ.

Bước 2: Sau khi có $\hat{\rho}$, hãy biến đổi $X_t^* = X_t - \rho X_{t-1}$ và $Y_t^* = Y_t - \rho Y_{t-1}$ và ước lượng phương trình ban đầu theo các biến đã được biến đổi ở trên.

8.2. Chọn mô hình và kiểm định việc chọn mô hình

8.2.1. Chọn mô hình

- Tiết kiệm
- Tính đồng nhất
- Tính thích hợp (R^2)
- Tính bền vững về mặt lý thuyết
- Khả năng dự báo cao

8.2.2. Các sai lầm khi chọn mô hình

- Bỏ sót biến thích hợp
- Đưa vào mô hình những biến không phù hợp
- Lựa chọn mô hình không chính xác

8.2.3. Kiểm định việc chọn mô hình

a. Kiểm định sai lầm khi đưa các biến không cần thiết vào mô hình (kiểm định Wald)

Xét mô hình:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

Tiến hành kiểm định giả thiết H_0: $\beta_4 = 0$. Khi đó ta dùng kiểm định Wald.

Kiểm định Wald. Xét các mô hình:

(U) $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_m X_{mi} + \beta_{m+1} X_{(m+1)i} + \ldots + \beta_k X_{ki} + u_i$

(R) $Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_m X_{mi} + v_i$

(U) là MH không giới hạn và (R) là mô hình giới hạn.

Kiểm định giả thiết H_0: $\beta_{m+1} = \ldots = \beta_k = 0$

Bước 1: Ước lượng (U) và (R), từ đó tính được RSS_U và RSS_R thay vào công thức:

$$F_C = \frac{(RSS_R - RSS_U)/(k-m)}{RSS_U/(n-k)}$$

Bước 2: Với mức ý nghĩa α, tìm $F_\alpha(k\text{-}m,n\text{-}k)$

Bước 3: Nếu $F_C > F_\alpha(k\text{-}m,n\text{-}k)$: Bác bỏ H_0, tức là (U)

b. Kiểm định việc bỏ sót biến giải thích trong mô hình

Để kiểm định các biến bỏ sót, ta dùng kiểm định Reset của Ramsey, gồm các bước:

Bước 1: Dùng OLS để ước lượng mô hình

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

Từ đó ta tính $\hat{Y}_i$ và R^2_{old}

Bước 2: dùng OLS để ước lượng mô hình

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 \hat{Y}^2 + \beta_4 \hat{Y}^3 + \ldots + v_i$$

Tính R^2_{new}

Kiểm định giả thiết H_0: $\beta_3 = \beta_4 = \ldots = \beta_k = 0$

Bước 3: Tính $F = \dfrac{(R^2_{new} - R^2_{old})/m}{(1 - R^2_{new})/(n-k)}$

n: số quan sát, k: số tham số trong mô hình mới; m: số biến đưa thêm vào.

Bước 4: Nếu F > F_α(m,n-k): Bác bỏ H_0, tức các hệ số $\beta_3, \beta_4, \ldots \beta_k$ không đồng thời bằng 0, mô hình cũ đã bỏ sót biến.

Ví dụ 8.2. Sử dụng số liệu 8.1 để tiến hành việc kiểm định

8.3. Kiểm định giả thiết phân phối chuẩn của u_i

Để kiểm định phân phối chuẩn của Ui, ta dùng kiểm định χ^2, hay kiểm định Jarque-Bera:

Kiểm định giả thiết H_0: u_i có phân phối chuẩn

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

$$S = \frac{\sum (u_i - \bar{u})^3}{n.SE_u^3} \qquad K = \frac{\sum (u_i - \bar{u})^4}{n.SE_u^4}$$

Nếu JB > $\chi^2_{(2)}$, Bác bỏ H_0, ngược lại, chấp nhận H_0