



TẬP BÀI GIẢNG MÔN HỌC KINH TẾ LƯỢNG



Chương I

Chương I	2
NHẬP MÔN KINH TẾ LƯỢNG	14
1.1 KINH TẾ LƯỢNG LÀ GÌ?	14
1.1.1 Ước lượng các mối quan hệ kinh tế	15
Kinh tế học thực nghiệm cung cấp rất nhiều ví dụ nhằm ước lượng các mối quan hệ.....	15
1.1.2 Kiểm định giả thuyết	15
1.1.3 Dự báo	16
1.2 PHƯƠNG PHÁP THỰC HIỆN MỘT NGHIÊN CỨU KINH TẾ LƯỢNG	17
Để thực hiện một nghiên cứu thực nghiệm, một nhà nghiên cứu phải có những câu trả 17	
LYÛTHUYẾT KINH TẾ KINH NGHIỆM, NGHIÊN CỨU KHÁC	17
XÁC ĐỊNH VẤN ĐỀ	17
THIẾT LẬP MÔ HÌNH	17
ÖÖC LÖÖNG MÔ HÌNH	17
KIỂM NHÌN GIẢ THUYẾT	17
THIẾT LẬP LẠI MÔ HÌNH	17
DIỄN DỊCH KẾT QUẢ	17
DÖI BÄÖD	17
Hình 1.1 : Các bước thực hiện một nghiên cứu kinh tế lượng	17
1.2.1 Xác định vấn đề nghiên cứu:	18
1.2.2 Thiết lập mô hình	18
1.2.3 Thu thập dữ liệu	18
Để ước lượng mô hình kinh tế lượng mà một nhà nghiên cứu đưa ra, cần có mẫu dữ 18	
1.2.4 Ước lượng mô hình kinh tế lượng.	18

Sắp xếp lại các môn học Kinh tế lượng và dữ liệu phù hợp đã được thu thập, nhiệm vụ chủ	18
1.2.5 Kiểm định giả thuyết	19
1.2.6 Diễn dịch kết quả	19
1.3 DỮ LIỆU TRONG CÁC MÔ HÌNH KINH TẾ LƯỢNG	19
Ví dụ: Bộ dữ liệu điều tra mức sống dân cư năm 2002 VLSS-2002	19
Ví dụ: Ta có thể có các quan sát chuỗi thời gian hàng năm cho chỉ tiêu GDP	19
Dữ liệu có thể được thu thập trên các biến "rời rạc" hay "liên tục"	19
1.4 CÁC MỐI QUAN HỆ TRONG NGHIÊN CỨU KINH TẾ LƯỢNG:	20
1.4.1 Phân tích hồi quy và quan hệ hàm số:	20
(1) Phân tích hồi quy là phân tích sự phụ thuộc của biến phụ thuộc vào một hay nhiều biến độc lập	20
☞ Biến độc lập (hay còn gọi là biến giải thích): là giá trị được xác định trước	20
<i>Ví dụ:</i>	20
(2) Quan hệ hàm số	20
Biến phụ thuộc không phải là đại lượng ngẫu nhiên, ứng với một giá trị của biến	20
<i>Ví dụ:</i>	20
Cách tính luợng cơ bản: Luợng cơ bản = Hệ số * Đơn giá tiền luợng	20
1.4.2 Hàm hồi quy và quan hệ nhân quả:	20
<i>Ví dụ:</i>	20
Nhu cầu tiêu dùng (Sản lượng) = F(giá cả, thu nhập, ...) □ lý thuyết kinh tế □	20
1.4.3 Phân tích hồi quy và phân tích tương quan:	21
(2) Phân tích hồi quy:	21
Kỹ thuật: không có tính đối xứng	21
Chương II	22
MÔ HÌNH HỒI QUY ĐƠN BIẾN	22
2.1. MÔ HÌNH CƠ BẢN:	22
<i>Lưu ý: Hàm hồi quy tuyến tính được hiểu là tuyến tính theo tham số.</i>	22

Y	Tập bài giảng môn học: Kinh tế lượng	22
2.1.1.	Các quan sát :	24
	Tập hợp tất cả các quan sát có thể có mà ta quan tâm nghiên cứu trong một vấn đề nào	24
	Mẫu là tập hợp con của tổng thể. Số phần tử của mẫu đã ký hiệu là n (cỡ mẫu).	24
2.1.2.	Các tham số thống kê :	24
	Ta có : PRF : $Y_i = \beta_1 + \beta_2 X_i + u_i$	24
	$E(Y_i / X_i) = \beta_1 + \beta_2 X_i$ và $\beta_2 =$	24
	Ý nghĩa các hệ số hồi quy :	24
	Tư đơ ng tự cho cách giải thích	24
	Ví dụ :	25
	PRF : $PRICE_i = \beta_1 + \beta_2 SQFT_i + u_i$	25
	$PRICE_i = 52,35 + 0,139 * SQFT_i + \hat{u}_i$	25
	$\hat{\beta}_2$	25
	$\hat{\beta}_1$	25
	$\hat{\beta}_1$	25
2.1.3.	Số hạng sai số :	25
	Nguyên nhân gây ra sai số :	26
2.2.	ƯỚC LƯỢNG MÔ HÌNH BẰNG PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (Ordinary Least Of Squares)	26
	Nguyên tắc :	26
	Tiêu chuẩn tối ư u đư ợc sử dụng bởi phư ơ ng pháp bình phư ơ ng tối thiểu thông thư ờng	26
	Vậy : $ESS = \sum \hat{u}^2 =$	26
	$\sum (Y - \hat{Y})^2 \rightarrow \text{Min}$	26
	$\sum (Y - \hat{Y}) = 0$	26
	$Y = \hat{Y} = \beta_1 + \beta_2 X$	28
	Bảng 2.1 : Thực hiện hồi quy đơ n biến	28
	Ta có $\hat{\beta}_1$	28

Việc bài giảng môn học Kinh tế lượng Việc: $PRICE_i = 52,35 + 0,19 \cdot SQFT_i + \hat{u}_i$	28
Lưu ý:	29
Hình 2.3 Thực hiện hồi quy đơn biến trên EXCEL	29
Hình 2.4 Thực hiện hồi quy đơn biến trên EVIEW	29
2.3. CÁC GIẢ THIẾT CƠ BẢN :	30
2.4. TÍNH CHẤT :	30
$S_{XX} = \sum (X_i$	30
Ước lượng OLS là BLUE (Best Linear Unbias Estimation) nếu thỏa mãn giả thiết	30
Nghĩa là trong tất cả các tổ hợp tuyến tính không thiên lệch của Y, ước lượng OLS	30
2.5. ĐỘ CHÍNH XÁC CỦA ƯỚC LƯỢNG	31
$Var(\hat{u}) = \hat{\sigma}^2 =$	31
2	31
S_{XX}	31
XX	31
1	31
1	31
XX	31
$X \hat{\sigma}^2$	31
S_{XX}	31
2.6. ĐỘ THÍCH HỢP CỦA MÔ HÌNH:	31
$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \hat{R}^2$	32
S_{XX}	32
Vậy : $-1 \leq R^2 \leq 1$	32
TSS	32
TSS	32
S_{YY}	32
2.7. KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ:	32
Kiểm định giả thuyết thống kê là một trong những nhiệm vụ chính của nhà kinh tế	32

2.7.1. Đối với tham số độ dốc	32
Kỳ vọng: Bác bỏ H_0	32
Xác Định Trị Thống Kê Kiểm Định	32
S^2	32
Kết luận: Nếu $ t_c > t_{\alpha/2, n-2}$ Bác bỏ H_0	33
Khoảng tin cậy của β_2	33
Nếu a, b cùng dấu Bác bỏ H_0	33
Ví dụ: Quay lại với ví dụ	33
$PRICE_i = 52,35 + 0,139 * SQFT_i + \hat{u}_i$	33
2.7.2. Đối với tham số tung độ góc	34
Kiểm định độ thích hợp của mô hình sẽ được trình bày trong chương Hồi quy đa biến	34
2.8. TRÌNH BÀY KẾT QUẢ HỒI QUY	34
Một cách khác là điền các sai số chuẩn dưới các hệ số hồi quy	34
2.9. THANG ĐO:	35
(2) PRICE *	35
1000	35
Chương III	36
MÔ HÌNH HỒI QUY BỘI	36
3.1. MÔ HÌNH CƠ BẢN:	36
$Q^D = f(P, I, P_s, P_c, \text{Market size}, P_f \text{ (giá kỳ vọng)}, T \text{ (thị hiếu)})$	36
$Q^S = f(K, L, \text{TECH})$	36
PRF : $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \dots + \beta_k X_{ki} + u_i$	36
Trong đó : Y_i là quan sát thứ i của biến Y	36
3.1.1. Ý nghĩa các hệ số:	37
Ta có : $E(Y_i / X_{si}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \dots + \beta_k X_{ki}$	37
Tương tự cho cách giải thích thông số ước lượng	37
3.2. ƯỚC LƯỢNG MÔ HÌNH CƠ BẢN BẰNG OLS:	38
Cũng như mô hình hồi quy tuyến tính đơn, các tham số của mô hình hồi quy tuyến tính	38

Tập bài giảng môn học: Kinh tế lượng	
ESS = $\sum u_i^2 =$	38
$2 \sum X_{2i}$	38
$3 \sum X_{3i}$	38
Hệ phương trình chuẩn	38
$\sum u_i$	38
Với	38
Ví dụ : Bổ sung dữ liệu về giá và diện tích, số phòng tắm, số phòng ngủ của 14 căn	39
SUMMARY OUTPUT	39
ANOVA	39
Hình 3.1: Hồi quy đa biến trên EXCEL	39
$PRICE_i = 129.0616 + 0.1548 * SQFT_i - 12.19276 * BATHS_i - 21.58752 * BEDRMS_i + \hat{u}_i$	39
3.3. CÁC GIẢ THIẾT CƠ BẢN :	40
Tức là, không tồn tại tập hợp các hệ số thỏa mãn biểu thức sau với mọi i:	40
$1 + \sum_2 X_{2i} + \sum_3 X_{3i}$	40
3.4. TÍNH CHẤT :	40
Ước lượng OLS cho mô hình hồi quy tuyến tính đa biến là BLUE (Best Linear Unbias	40
3.5. ĐỘ CHÍNH XÁC CỦA ƯỚC LƯỢNG	40
$Var(\hat{u}) = \hat{\sigma}^2 =$	41
$VAR(\hat{\beta})$	41
3.6. ĐỘ THÍCH HỢP CỦA MÔ HÌNH:	41
Trong mô hình hồi quy tuyến tính bội để loại bỏ ảnh hưởng việc thêm biến làm bậc tự	41
2	41
3.7. KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ:	41
3.7.1. Kiểm định các tham số riêng lẻ:	41
Kỳ vọng: Bác bỏ H_0	43
<i>Xác Định Trị Thống Kế Kiểm Định</i>	43
S^2	43
Kết luận: Nếu $ t_c > t_{\alpha/2, n-k} \square$ Bác bỏ H_0	43

3.7.2. Kiểm định Wald (Kiểm định tổ hợp các tham số):	43
<i>Xác Định Trị Thống Kê Kiểm Định</i>	43
$F^* = F_{\alpha, k-m, n-k}$	43
Ví dụ:	44
<i>Xác Định Trị Thống Kê Kiểm Định</i>	44
$F_C = 0.471106 < F^* = F_{5\%, 2, 10} = 4,103$ □ Không bác bỏ H_0 □ Chọn mô hình (R).....	44
Hình 3.3: Thực hiện kiểm định Wald trên EVIEW	44
Ta thấy: $p\text{-value} = 0.6375 > 5\%$ □ Không bác bỏ H_0 với mức ý nghĩa $\alpha = 5\%$	44
3.7.3. Kiểm định độ thích hợp của mô hình:	44
$H_0: \beta_2 = \beta_3 = \beta_4 = 0$	45
H_1 : Có ít nhất 1 số β_j khác 0 (với $j = 2;4$).....	45
Kết quả : $F_C = 16.98894$ hay $p\text{-value} = 0.000299 < 5\%$ □ Bác bỏ H_0 với $\alpha = 5\%$	45
3.8. CHIẾN LƯỢC XÂY DỰNG MÔ HÌNH :	45
3.8.1. Hậu quả của việc thiếu biến quan trọng hoặc thừa biến không quan trọng:	45
3.8.2. Chiến lược xây dựng mô hình:	46
Từ mô hình nhiều biến sau đó loại dần các biến không quan trọng để được mô hình	46
Hình 3.4: Mô hình tổng quát (không giới hạn)	47
$H_0: \beta_2 = \beta_3 = \beta_7 = 0$	47
$H_1: \beta_2 \neq 0$ hoặc $\beta_3 \neq 0$ hoặc $\beta_7 \neq 0$	47
Chạy EVIEW ta có kết quả:.....	47
Hình 3.5: Kiểm định Wald để chọn mô hình	47
Hình 3.6: Mô hình đơn giản (giới hạn).....	48
$BUSTRAVL = 2815.7 - 0.2013INCOME + 1.5766POP + 0.1534DENSITY + u^i$	48
3.9. HỒI QUY VỚI BIẾN ĐỊNH TÍNH:	48
$E(WAGE_i/D_i=1) = \beta_1 + \beta_2$: Lương trung bình của nam	49
3.9.1. Các dạng mô hình:	49
Nam: $WAGE_i = (\beta_1 + \beta_1^{TM}) + (\beta_2 + \beta_2^{TM})EXPER_i + u_i$	49
3.9.2. Biến giả nhiều thuộc tính:	51
Với một đặc trưng có m thuộc tính ta sẽ đặt $m-1$ biến giả.....	51

Vấn đề giảng môn học: Kinh tế lượng, hạ, thu, đông □ 3 biến giả	51
Chương IV	52
4.1. GIỚI THIỆU:	52
<i>Ví dụ</i> : Ước lượng hàm tiêu dùng. Y : Tiêu dùng, X ₂ : Thu nhập và X ₃ : của cải.....	52
Dạng hàm: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$	52
Theo dữ liệu ta thấy : X ₂ và X ₃ có mối quan hệ tuyến tính chính xác.....	52
4.2. NGUỒN GỐC CỦA ĐA CỘNG TUYẾN (Multicollinearity)	53
4.2.1. Do phương pháp thu thập dữ liệu	53
4.2.2. Dạng hàm mô hình:	53
4.3. HẬU QUẢ :	53
4.3.1. Đa cộng tuyến hoàn hảo	53
4.3.2. Hệ quả khi có đa cộng tuyến không hoàn hảo	53
(1) Ước lượng OLS vẫn BLUE.....	53
(2) Sai số chuẩn của các hệ số sẽ lớn.	53
(3) R ² rất cao cho dù thống kê t ít ý nghĩa.....	54
(4) Các ước lượng và sai số chuẩn của ước lượng rất nhạy cảm với sự thay đổi của dữ liệu.....	54
R ² =0.96, F = 92.40.....	54
Y : Tiêu dùng, X ₂ : Thu nhập và X ₃ : của cải.....	54
Hầu như chúng ta có đa cộng tuyến hoàn hảo giữa X ₂ và X ₃	54
Biến thu nhập trở nên có ý nghĩa thống kê, nhưng trước lúc đó trong mô hình.....	54
4.4. NHẬN DẠNG :	55
(1) R ² cao và thống kê t thấp.....	55
(2) Tương quan tuyến tính mạnh giữa các biến độc lập.....	55
(3) Thực hiện hồi qui phụ.....	55
F _c =.....	55
(4) Thừa số tăng phương sai (Variance inflation factor-VIF).....	55
VIF =.....	55
Khi r _{ij} tăng làm VIF tăng và làm tăng mức độ đa cộng tuyến.....	55

4.3. Tập bài giảng mô hình Kinh tế lượng CÁC GIẢI PHÁP KHÁC PHỤC ĐA CỘNG TUYẾN	55
Ví dụ: bỏ biến của cái ra khỏi mô hình hàm tiêu dùng.	56
Sử dụng kết quả của các mô hình kinh tế lượng trước ít có đa cộng tuyến	56
Ước lượng hàm hồi qui theo thời gian	57
$Y_t = \beta_1 + \beta_2 \ln P_t + u_t$	57
Chương V	58
DẠNG HÀM	58
5.1. HÀM TUYẾN TÍNH:	58
Đây là dạng hàm đơn giản nhất, tuy nhiên, do tính đơn giản này nên khả năng mô tả.....	58
5.2. HÀM ĐA THỨC.	58
<i>Tác dụng:</i>	60
Điểm cực trị (cực đại hoặc cực tiểu): $MC = 0 \Rightarrow Q = \sqrt[3]{1}$	60
<i>Ví dụ :</i>	60
SRF : $LRAC = 2,38 - 0,615Q + 0,054 Q^2 + \hat{u}$	60
Trong đó: Y = sản lượng	60
<i>Dạng tổng quát:</i>	61
<i>Tác dụng:</i>	61
X_k	61
Y	61
Trong hàm Cobb-Douglas: $\beta_2 =$	61
Y	61
X_k	61
<i>Kết quả:</i>	62
<i>Hàm Cobb-Douglas Tổng quát</i>	62
Trong đó: Y = sản lượng	62
β_t	62
β_t	62

5. Tập bài giảng Kinh tế lượng	62
5.4. HÀM BÀN LOG	
Sau một năm, số tiền này sẽ tăng lên đến: $Y_1 = 100.000.000(1 + 0.08) = 108.000.000$	62
Sau hai năm, số tiền này sẽ tăng lên đến: $Y_2 = 108.000.000(1 + 0.08) =$	63
Y_0 là giá trị ban đầu của Y	63
Dạng hàm tổng quát:	63
$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$	63
X	63
X	63
Ví dụ:	63
$\ln(\text{GDP thực}) = 6,9636 + 0,0269t$	63
$R^2 = 0,95$	63
5.5. CÁC DẠNG HÀM KHÁC:	63
Chương VI	64
HIỆN TƯỢNG PHỤ Ơ NG SAI CỦA SAI SỐ THAY ĐỔI	64
6.1. GIỚI THIỆU:	64
6.2. HẬU QUẢ:	64
6.2.1. Tác động lên tính chất của các ước lượng OLS:	64
6.2.2. Tác Động Lên Các Kiểm Định Giả Thuyết	64
6.2.3. Tác Động Lên Việc Dự Báo	65
6.3. NGUYÊN NHÂN XẢY RA HIỆN TƯỢNG HET	65
6.4.1. Bảng trực giác và kinh nghiệm:	65
Hình 6.2: Đồ thị ($X_i u_i$) nhận biết hiện tượng HET	66
$\text{EXPTRAV}_i = \beta_1 + \beta_2 \text{INCOME}_i + u_i$	66
Hình 6.4: Kết quả mô hình hồi quy cơ bản của ví dụ	66
Vậy: $\text{EXPTRAV}_i = 0,49812 + 0,055573 \text{INCOME}_i + \hat{u}_i$	67
Hình 6.5: Đồ thị nhận biết hiện tượng HET	67
Dựa vào đồ thị nói trên, ta có thể nói có dấu hiệu có hiện tượng HET trong mô hình.	67

6.4.3. Kiểm định nhân tử Lagrange (Lagrange Multiplier Test – LM Test):	67
<i>Ví dụ:</i> Kiểm tra HET của ví dụ trên bằng kiểm định WHITE với chọn $\alpha = 10\%$	68
Phụ σ ng trình hồi quy tổng thể: $EXPTRAV_i = \beta_1 + \beta_2 INCOME_i + u_i$	68
Hình 6.6: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.4.....	68
Giả thiết kiểm định: $H_0 : \beta_2 = \beta_3 = 0$	68
$R_{hqp} = 4.876820 >$	68
6.5.1. Bình phương tối thiểu tổng quát (hoặc trọng số):	69
$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_K X_{Ki} + u_i$	69
X_{2i}	69
X_{3i}	69
X_{ki}	69
$Var(u_i)$	69
Mô hình (*) không có số hạng sai số không đổi nên sẽ có tính BLUE.....	69
Thủ tục GLS được áp dụng cho trường hợp phương sai thay đổi thì cũng giống như thủ.....	69
f_i	69
Phương pháp này gọi là Bình Phương Tối Thiểu Tổng Quát Khả Thi (FGLS) hay Bình.....	69
Thực hiện hồi quy: $\sum u_i^2$	69
Hình 6.7: Kiểm định White để nhận dạng hiện tượng HET (cách trực tiếp).....	70
Tạo biến: u_i^2 từ forecast.....	70
Hình 6.8: Thực hiện hồi quy theo WLS theo trọng số từ hồi quy phụ White.....	70
Hình 6.9: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.8.....	71
Giả sử tính phương sai của sai số thay đổi được tính với thông qua một biến Z_i	71
$Var(u_i) = \sigma_i^2 = \sigma^2 Z_i$. Vậy: $f_i = \sigma^2 Z_i$	71
Ta có: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_K X_{Ki} + u_i$	71
X_{2i}	71
Z_i	71
X_{3i}	71

Tập bài giảng môn học: Kinh tế lượng	71
X_{ki}	71
Z_i	71
$\text{Var}(u)$	71
$\leq Z_i f$	71
Hình 6.10: Thực hiện hồi quy theo WLS theo trọng số 1/INCOME	72
Hình 6.11: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.10	72
Trong đó: VA = sản lượng	73
$VA = \beta_0 + \beta_1 K + \beta_2 L + u_i$	74
$L^3 e^{u_i}$	74
Chương VII	75
TƯƠNG QUAN CHUỖI	75
7.1. GIỚI THIỆU:	75
AR(p): tương quan chuỗi bậc p	75
7.2. HẬU QUẢ:	75
7.2.1. Tác động lên tính chất của các ước lượng:	75
7.2.2. Tác Động Lên Các Kiểm Định Giả Thuyết	75
7.2.3. Tác Động Lên Việc Dự Báo	76
7.3. NGUYÊN NHÂN XẢY RA HIỆN TƯỢNG AR	76
7.4. NHẬN DẠNG HIỆN TƯỢNG HET	76
7.4.1. Bằng trực giác và kinh nghiệm:	76
Để phát hiện AR người ta thường dùng các đồ thị phân tán giữa: (X_t, Y_t) (X_t, u_t) hoặc	76
<i>Ví dụ:</i> DATA6-6 có dữ liệu hàng năm về dân số nông trại theo phần trăm tổng dân số	76
Mô hình hồi quy: $FARMPOP = \beta_1 + \beta_2 TIME + u$, trong đó TIME là biến xu thế.	76
Hình 7.1: Đồ thị nhận biết hiện tượng AR	76
7.4.3. Kiểm định tương quan chuỗi bậc nhất (Durbin - Watson):	77
Kiểm định Durbin – Watson chỉ nhận dạng được hiện tượng tương quan chuỗi bậc 1	77
AR(1): $u_t = \rho u_{t-1} + \epsilon_t$	77

Danh sách bài giảng môn học: Kinh tế lượng	77
Bước 4: Tra bảng tìm giá trị : d_U và d_L . Xác định xem DW nằm trong khoảng nào để	77
Tự tương quan dương	77
$H_1: \rho > 0$	77
$H_0: \rho = 0$	77
$H_1: \rho < 0$	77
Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ	78
Hình 7.2: Mô hình hồi quy cơ bản.	78
Mô hình ước lượng: $FARMPOP_t = 13.77727 - 0.324848TIME_t + \hat{u}_t$	78
Nếu chọn mô hình (U) : nghĩa là mô hình có AR(p).	78
Bước 1: Chạy mô hình hồi quy (R).	78
Bước 3: Tính $t_{tt} = (n - p)R^2_{hqp}$	79
Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ	79
Sau khi chạy mô hình (R) : $FARMPOP_t = 13.77727 - 0.324848TIME_t + \hat{u}_t$	79
Hình 7.3: Kiểm định LM để nhận dạng AR(1).	79
$AC_k = \rho(u_t, u_{t-k})$: Hệ số tương quan giữa u_t và u_{t-k}	80
Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ	80
7.5. BIỆN PHÁP KHẮC PHỤC	81
7.5.1. Thay đổi dạng hàm:	81
7.5.2. Thủ tục ước lượng:	81
Hồi quy Y^*	81
Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ	82
Hình 7.5: Thực hiện thủ tục ước lượng để khắc phục AR(1).	82
$FARMPOP_t = -5.110588 + 0.087428TIME_t - 0.956023 \hat{u}_{t-1} + \hat{u}_t$	82

NHẬP MÔN KINH TẾ LƯỢNG

1.1 KINH TẾ LƯỢNG LÀ GÌ?

Diễn giải theo nghĩa đơn giản, kinh tế lượng (ECONOMETRICS) liên quan đến việc

~~Tập bài giảng môn học Kinh tế lượng~~ kê trong kinh tế học. Tuy nhiên, trong thống kê kinh tế, các dữ liệu thống kê là chính yếu còn **kinh tế lượng được là sự hợp nhất của lý thuyết kinh tế, công cụ toán học và các phương pháp luận thống kê.**

Mở rộng hơn, kinh tế lượng quan tâm đến :

- (1) Ước lượng các mối quan hệ kinh tế,
- (2) Đối chiếu lý thuyết kinh tế với thực tế và kiểm định các giả thuyết liên quan đến hành vi kinh tế,
- (3) Dự báo các hành vi của các biến số kinh tế.

Sau đây là những ví dụ thực tế minh họa mỗi hoạt động này của kinh tế lượng :

1.1.1 Ước lượng các mối quan hệ kinh tế

Kinh tế học thực nghiệm cung cấp rất nhiều ví dụ nhằm ước lượng các mối quan hệ kinh tế từ dữ liệu. Sau đây là một số các ví dụ :

- Các nhà phân tích và các công ty thường quan tâm ước lượng cung/cầu của các sản phẩm, dịch vụ.
- Một công ty thường quan tâm đến việc ước lượng ảnh hưởng của các mức độ quảng cáo khác nhau đến doanh thu và lợi nhuận.
- Các nhà phân tích thị trường chứng khoán tìm cách liên hệ giá của cổ phiếu với các đặc trưng của công ty phát hành cổ phiếu đó, cũng như với tình hình chung của nền kinh tế.
- Nhà nước muốn đánh giá tác động của các chính sách tiền tệ tài chính đến các biến quan trọng như thất nghiệp, thu nhập, xuất nhập khẩu, lãi suất, tỷ lệ lạm phát, và thâm hụt ngân sách ...

1.1.2 Kiểm định giả thuyết

Một điểm tốt của kinh tế lượng là quan tâm đến việc kiểm định giả thuyết về các hành vi kinh tế. Ví dụ minh họa :

- Các nhà phân tích thường quan tâm xem nhu cầu có co giãn theo giá và thu nhập hay không ?

- Các công ty cũng muốn xác định xem chiến dịch quảng cáo của mình có thực sự tác động làm tăng doanh thu hay không ?
- Công ty muốn biết lợi nhuận tăng hay giảm theo qui mô hoạt động ?
- Các công ty kinh doanh thuốc lá và các nhà nghiên cứu y khoa đều quan tâm đến các báo cáo phẫu thuật tổng quát về hút thuốc và ung thư phổi (và các bệnh về hô hấp khác) có dẫn đến việc giảm tiêu thụ thuốc lá đáng kể hay không ?
- Các nhà kinh tế vĩ mô muốn đánh giá hiệu quả của các chính sách nhà nước.

1.1.3 Dự báo

Khi các biến số được xác định và chúng ta đánh giá được tác động cụ thể của chúng đến chủ thể nghiên cứu, chúng ta có thể muốn sử dụng các mối quan hệ ước lượng để dự đoán các giá trị trong tương lai. Ví dụ minh họa :

- Các công ty dự báo doanh thu, lợi nhuận, chi phí sản xuất ... cần thiết.
- Chính phủ dự đoán nhu cầu về năng lượng để có chiến lược đầu tư xây dựng hoặc các thỏa thuận mua năng lượng từ bên ngoài cần được ký kết.
- Các công ty dự báo các chỉ số thị trường chứng khoán và giá cổ phiếu.
- Chính phủ dự đoán những con số như thu nhập, chi tiêu, lạm phát, thất nghiệp, và thâm hụt ngân sách và thương mại.
- Các địa phương dự báo định kỳ mức tăng trưởng của địa phương qua các mặt: dân số; việc làm; số nhà ở, tòa nhà thương mại và các xưởng công nghiệp; nhu cầu về trường học, đường xá, trạm cảnh sát, trạm cứu hỏa, và dịch vụ công cộng; ...v.v

Do ba bước tổng quát được xác định trong phần mở đầu của chương này thường căn cứ vào dữ liệu mẫu hơn là dựa vào dữ liệu điều tra của tổng thể, vì vậy trong những cuộc điều tra chuẩn này sẽ có yếu tố bất định:

- Các mối quan hệ ước lượng không chính xác.
- Các kết luận từ kiểm định giả thuyết hoặc là phạm vào sai lầm do chấp nhận một giả thuyết sai hoặc sai lầm do bác bỏ một giả thuyết đúng.
- Các dự báo dựa vào các mối liên hệ ước lượng thường không chính xác.

Để giảm mức độ bất định, một nhà kinh tế lượng sẽ luôn luôn ước lượng nhiều mối quan hệ khác nhau giữa các biến nghiên cứu. Sau đó, họ sẽ thực hiện một loạt các kiểm tra để xác định mối quan hệ nào mô tả hoặc dự đoán gần đúng nhất hành vi của biến số quan tâm.

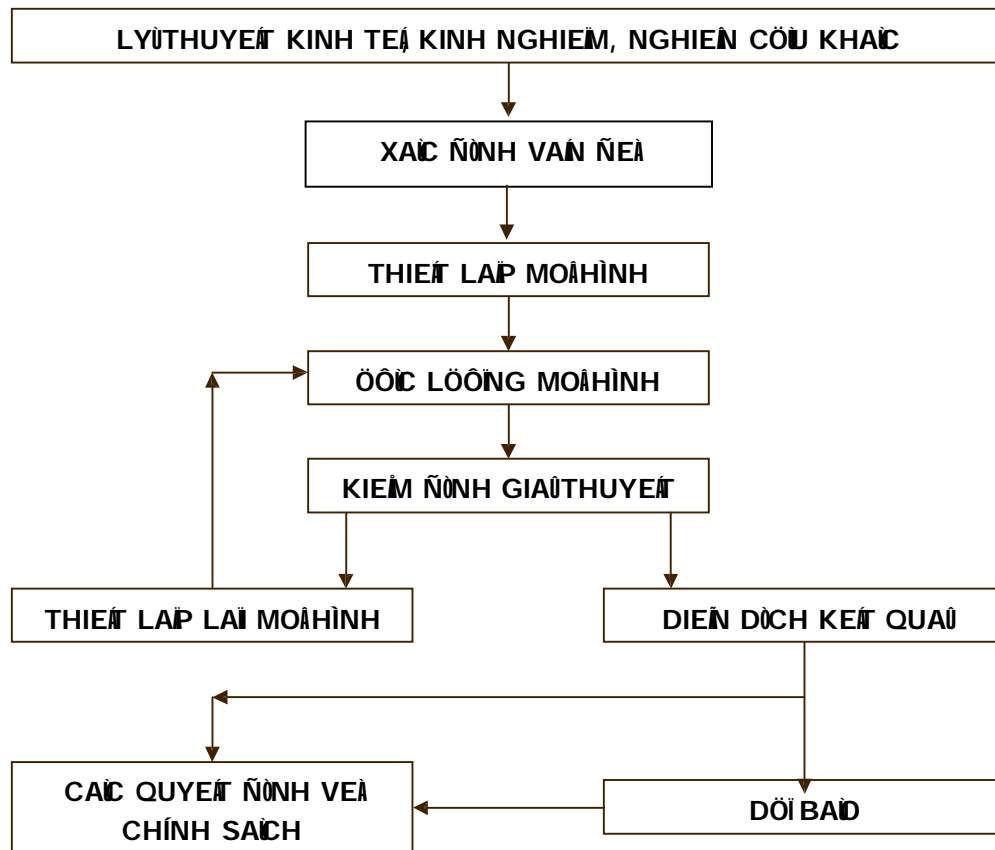
Tính bất định này khiến cho phương pháp thống kê trở nên rất quan trọng trong môn kinh tế lượng.

1.2 PHƯƠNG PHÁP THỰC HIỆN MỘT NGHIÊN CỨU KINH TẾ LƯỢNG

Để thực hiện một nghiên cứu thực nghiệm, một nhà nghiên cứu phải có những câu trả lời thỏa đáng cho các câu hỏi sau:

- (1) Mô hình có ý nghĩa kinh tế không? Cụ thể, mô hình có thể hiện mọi quan hệ tương thích ẩn trong quá trình phát dữ liệu hay không?
- (2) Dữ liệu có tin cậy không?
- (3) Phương pháp ước lượng sử dụng có phù hợp không? Có sai lệch trong các ước lượng tìm được không?
- (4) Các kết quả của mô hình so với các kết quả từ những mô hình khác như thế nào?
- (5) Kết quả thể hiện điều gì? Kết quả có như mong đợi dựa trên lý thuyết kinh tế hoặc cảm nhận trực giác không?

Do đó, mặc dù có nhiều quan điểm khác nhau, nhưng nói chung đều chia một nghiên cứu kinh tế lượng thành các bước sau:



Hình 1.1 : Các bước thực hiện một nghiên cứu kinh tế lượng

1.2.1 Xác định vấn đề nghiên cứu:

Vấn đề nghiên cứu thường được xác định bởi yêu cầu của công việc và/hoặc do cấp trên của nhà nghiên cứu chỉ định.

Ví dụ, một trong những nhiệm vụ chính của nhà phân tích trong bộ phận dự báo của ngành điện lực là ước lượng liên hệ giữa nhu cầu về điện và các yếu tố ảnh hưởng như thời tiết và tiêu thụ theo mùa, giá điện, thu nhập, loại máy móc gia dụng, đặc điểm địa lý, công nghiệp của nơi phục vụ ...v.v. Mối liên hệ ước lượng sau đó sẽ được dùng để tính các giá trị dự báo lượng điện. Các giá trị dự báo này được ngành điện lực khu vực xem xét để quyết định cấu trúc giá mới như thế nào và có cần phải xây dựng thêm nhà máy năng lượng mới để đáp ứng nhu cầu người dân trong khu vực hay không. Trong ví dụ này, dễ dàng nêu ra vấn đề nghiên cứu là liên hệ giữa nhu cầu điện với các yếu tố ảnh hưởng đến nhu cầu này, và phát ra các dự báo.

1.2.2 Thiết lập mô hình

Mọi phân tích hệ thống kinh tế, xã hội, chính trị hoặc vật lý dựa trên một cấu trúc logic (gọi là mô hình), cấu trúc này mô tả hành vi của các phần tử trong hệ thống và là khung phân tích chính. Trong kinh tế học, cũng như trong các ngành khoa học khác, mô hình này được thiết lập dưới dạng phương trình, trong trường hợp này, các phương trình này mô tả hành vi kinh tế và các biến liên quan. Một mô hình được nhà nghiên cứu thiết lập có thể là một phương trình hoặc là hệ gồm nhiều phương trình.

Dựa trên lý thuyết kinh tế, kinh nghiệm, các nghiên cứu khác, nhà nghiên cứu sẽ đưa ra mô hình lý thuyết đề nghị. Chẳng hạn, một nhà kinh tế có thể xác định một hàm tiêu dùng có dạng như sau:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \text{ (quan hệ không xác định, có tính ngẫu nhiên)}$$

Trong đó: Y_t : Tiêu dùng (\$billion)

X_t : GDP (\$billion)

u_t : là sai số, một biến ngẫu nhiên (stochastic)

1.2.3 Thu thập dữ liệu

Để ước lượng mô hình kinh tế lượng mà một nhà nghiên cứu đưa ra, cần có mẫu dữ liệu về các biến phụ thuộc và biến độc lập.

1.2.4 Ước lượng mô hình kinh tế lượng.

Sau khi mô hình đã được thiết lập và dữ liệu phù hợp đã được thu thập, nhiệm vụ chủ yếu của nhà điều tra là ước lượng những thông số chưa biết của mô hình.

Trong ví dụ trên ta sẽ các ước lượng của số hạng tung độ gốc β_1 , số hạng độ dốc β_2 , và các thông số (như trung bình và phương sai) của phân bố xác suất của sai số u_t .

1.2.5 Kiểm định giả thuyết

Sau khi ước lượng mô hình, nhà nghiên cứu cần kiểm định các giả thuyết hoặc dự báo các giá trị của biến phụ thuộc, với những giá trị của các biến độc lập cho trước.

Việc kiểm định chẩn đoán mô hình nhiều lần nhằm chắc chắn là những giả định đặt ra và các phương pháp ước lượng được sử dụng phù hợp với dữ liệu đã thu thập. Mục tiêu của kiểm định là tìm được những kết luận thuyết phục nhất, đó là những kết luận không thay đổi nhiều đối với các đặc trưng của mô hình.

Kiểm định giả thuyết không chỉ được thực hiện nhằm cải tiến các đặc trưng của mô hình mà còn nhằm kiểm định tính đúng đắn của các lý thuyết.

1.2.6 Diễn dịch kết quả

Bước cuối cùng của một nghiên cứu là diễn dịch các kết quả: ra quyết định về chính sách hay dự báo.

1.3 DỮ LIỆU TRONG CÁC MÔ HÌNH KINH TẾ LƯỢNG

Có ba dạng dữ liệu kinh tế cơ bản: dữ liệu chéo (cross-sectional data), dữ liệu chuỗi thời gian (time series data), và dữ liệu dạng bảng (panel data).

- Dữ liệu chéo (cross-sectional data): bao gồm các quan sát cho nhiều đơn vị kinh tế tại một thời điểm cho trước. Các đơn vị kinh tế có thể là các cá nhân, các hộ gia đình, các hãng, các tỉnh thành, các quốc gia v.v...

Ví dụ: Bộ dữ liệu điều tra mức sống dân cư năm 2002 VLSS-2002

- Dữ liệu chuỗi thời gian (time series data): bao gồm các quan sát trên một đơn vị kinh tế cho trước tại nhiều thời điểm.

Ví dụ: Ta có thể có các quan sát chuỗi thời gian hàng năm cho chỉ tiêu GDP của một quốc gia từ năm 1960 đến 2005.

- Dữ liệu dạng bảng (panel data): là sự kết hợp giữa các quan sát của các đơn vị kinh tế về một chỉ tiêu nào đó theo thời gian.

Ví dụ: chúng ta thực hiện điều tra về hộ gia đình cho cùng những hộ gia đình trong vài năm để đánh giá sự thay đổi của những hộ này theo thời gian.

Dữ liệu có thể được thu thập trên các biến "rời rạc" hay "liên tục".

- Biến rời rạc là biến có một tập hợp các kết quả nhất định có thể đếm được. Ví dụ: số thành viên trong một hộ gia đình là một biến rời rạc.
- Biến liên tục là biến có một số vô hạn các kết quả, như là chiều cao của một đứa trẻ.

Nhiều biến kinh tế được đo bằng những đơn vị đủ nhỏ để chúng ta coi chúng như là liên tục, mặc dù thực ra chúng là rời rạc.

1.4 CÁC MÔI QUAN HỆ TRONG NGHIÊN CỨU KINH TẾ LƯỢNG:

1.4.1 Phân tích hồi quy và quan hệ hàm số:

(1) Phân tích hồi quy là phân tích sự phụ thuộc của biến phụ thuộc vào một hay nhiều biến độc lập.

☞ Biến phụ thuộc (hay còn gọi là biến được giải thích): là đại lượng ngẫu nhiên có phân bố xác suất.

☞ Biến độc lập (hay còn gọi là biến giải thích): là giá trị được xác định trước.

Ví dụ:

Nghiên cứu sự phụ thuộc chiều cao của con trai vào chiều cao của người cha (Galton Karl Pearson): Biến độc lập là chiều cao của người cha còn biến phụ thuộc là chiều cao của người con trai, ta không thể dự báo một cách chính xác chiều cao của người con trai thông qua chiều cao của người cha vì sai số và còn nhiều yếu tố không có trong mô hình.

Nói cách khác, từ chiều cao của một người cha X_i nào đó ta sẽ xác định được chiều cao trung bình của người con trai (giữa chiều cao thực sự của người con trai và chiều cao trung bình này sẽ có một khoảng cách gọi là sai số).

(2) Quan hệ hàm số

Biến phụ thuộc không phải là đại lượng ngẫu nhiên, ứng với một giá trị của biến độc lập ta xác định được duy nhất một biến phụ thuộc:

Ví dụ:

Cách tính lương cơ bản: Lương cơ bản = Hệ số * Đơn giá tiền lương

Ứng với mỗi hệ số và đơn giá tiền lương ta chỉ có mức lương cơ bản chính xác duy nhất.

1.4.2 Hàm hồi quy và quan hệ nhân quả:

Phân tích hồi quy nghiên cứu quan hệ một biến phụ thuộc với một hay nhiều biến độc lập, điều này không đòi hỏi giữa biến độc lập và biến phụ thuộc và biến độc lập phải có mối quan hệ nhân quả.

Ví dụ:

Nhu cầu tiêu dùng (Sản lượng) = $F(\text{giá cả, thu nhập, ...})$ □ lý thuyết kinh tế □
quan hệ nhân quả.

Lượng mưa = $F(\text{số con chuồn chuồn, ...})$

1.4.3 Phân tích hồi quy và phân tích tương quan:

Phân tích hồi quy và tương quan khác nhau về mục đích và kỹ thuật.

(1) Phân tích tương quan:

Mục đích: đo lường mức độ kết hợp tuyến tính giữa 02 biến. Ví dụ: mức độ nghiện thuốc lá và ung thư phổi, điểm thi môn toán và thống kê.

Kỹ thuật: có tính đối xứng.

(2) Phân tích hồi quy:

Mục đích: ước lượng hoặc dự báo một (hay nhiều) biến trên cơ sở giá trị đã cho của một (hay nhiều) biến khác.

Kỹ thuật: không có tính đối xứng.

Lưu ý: Để chuẩn bị cho các chương sau, đề nghị sinh viên về ôn tập lại các kiến thức về xác suất và thống kê.

Chương II

MÔ HÌNH HỒI QUY ĐƠN BIẾN (Simple Linear Regression Model)

2.1. MÔ HÌNH CƠ BẢN:

Mô hình hồi quy tổng thể (Population Regression Function – PRF) cho biến giá trị trung bình của biến phụ thuộc thay đổi như thế nào khi các biến độc lập nhận các giá trị khác nhau.

Nếu PRF có một biến độc lập thì gọi là mô hình hồi quy tuyến tính đơn biến (gọi tắt là mô hình hồi quy đơn biến).

Lưu ý: Hàm hồi quy tuyến tính được hiểu là tuyến tính theo tham số.

Mô hình hồi quy tổng thể PRF đơn biến có dạng như sau :

$$\text{PRF : } Y_i = \beta_1 + \beta_2 X_i + u_i$$

Trong đó : X_i, Y_i là quan sát thứ i của các biến X và Y

X là Biến độc lập, Y là biến phụ thuộc.

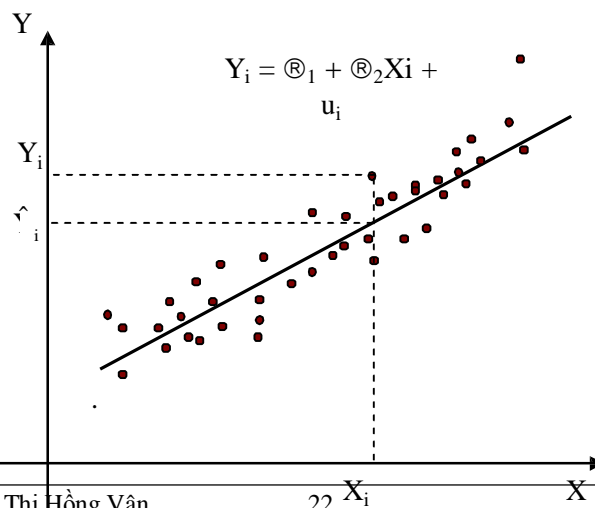
u_i là sai số của mô hình.

β_1, β_2 tham số của mô hình.

Dạng xác định của mô hình

$$E(Y_i/X_i) = \beta_1 + \beta_2 X_i + u_i$$

Đồ thị biểu diễn :



Hình 2.1 : Đường biểu diễn mô hình hồi quy tổng thể

2.1.1. Các quan sát :

Ví dụ : Để tìm mối liên hệ giữa giá bán của một ngôi nhà và diện tích sử dụng của nó ta sẽ đi thu thập dữ liệu này của từng ngôi nhà. Dữ liệu về giá bán và diện tích sử dụng của một căn nhà nào đó ta gọi là một quan sát.

Tập hợp tất cả các quan sát có thể có mà ta quan tâm nghiên cứu trong một vấn đề nào đó gọi là tổng thể. Số phần tử của tập hợp chính được ký hiệu là N.

Mẫu là tập hợp con của tổng thể. Số phần tử của mẫu đã ký hiệu là n (cỡ mẫu).

Để tìm được mô hình PRF ta phải có dữ liệu của tổng thể về các quan sát X_i và Y_i . Nhưng trong thực tế điều này rất khó khả thi vì khả năng và chi phí. Do đó thông thường ta chỉ có dữ liệu về các biến X_i và Y_i của một mẫu lấy ra từ tổng thể nên ta có thể xây dựng được mô hình hồi quy mẫu (Sample Regression Function – SRF)

Mô hình hồi quy mẫu SRF đơn biến có dạng như sau :

$$\text{SRF : } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

Trong đó : \hat{u}_i là phần dư của mô hình.

$\hat{\beta}_1, \hat{\beta}_2$ là tham số ước lượng của mô hình.

2.1.2. Các tham số thống kê :

Thuật ngữ tuyến tính ở đây được hiểu là tuyến tính theo các tham số ước lượng, nó có thể hoặc không tuyến tính với các biến.

Ta có : PRF : $Y_i = \beta_1 + \beta_2 X_i + u_i$

$$\beta_2 E(Y_i/X_i) = \beta_1 + \beta_2 X_i \text{ và } \beta_2 = \frac{\text{cov}(Y, X)}{\text{var}(X)}$$

Ý nghĩa các hệ số hồi quy :

β_2 : Độ dốc (Slope) của đường hồi quy tổng thể, là lượng thay đổi của Y, ở mức trung bình, trên một đơn vị thay đổi của X. Vì vậy β_2 được diễn dịch là ảnh hưởng cận biên của X lên Y.

β_1 : Tung độ gốc (Intercept) của đường hồi quy tổng thể, và là giá trị của trị trung bình Y khi X bằng 0. Tuy nhiên sẽ không có cách giải thích cho β_1 vì nguyên nhân là vì β_1 còn ẩn chứa biến bỏ sót (ngoài mô hình).

Tương tự cho cách giải thích $\hat{\beta}_1, \hat{\beta}_2$ của hàm hồi quy mẫu SRF.

$$\text{SRF : } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

$\hat{\beta}_2$: Độ dốc của đường SRF, là lượng thay đổi của Y, ở mức trung bình theo thông tin của mẫu, trên một đơn vị thay đổi của X.

$\hat{\beta}_1$: Tung độ gốc của đường SRF.

Ví dụ :

Tìm mối liên hệ giữa giá của một ngôi nhà (PRICE – đơn vị tính : ngàn USD) và diện tích sử dụng (SQFT – đơn vị tính : m²), ta thiết lập một mô hình hồi quy đơn giản sau :

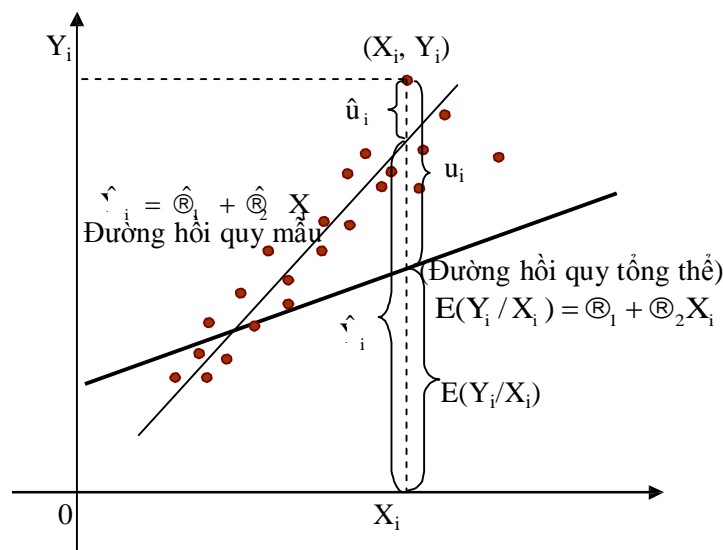
$$\begin{aligned} \text{PRF : } & \text{PRICE}_i = \beta_1 + \beta_2 \text{SQFT}_i + u_i \\ \text{SRF : } & \text{PRICE}_i = \hat{\beta}_1 + \hat{\beta}_2 \text{SQFT}_i + \hat{u}_i \\ & \text{PRICE}_i = 52,35 + 0,139 * \text{SQFT}_i + \hat{u}_i \end{aligned}$$

Cách giải thích các thông số ước lượng của mô hình :

$\hat{\beta}_2 = 0,139$, điều đó có nghĩa là : theo thông tin của mẫu, một mét vuông diện tích tăng thêm sẽ làm tăng giá bán nhà lên, ở mức trung bình, 0.139 ngàn USD hay 139 USD. Một cách thực tế hơn, khi diện tích sử dụng nhà tăng thêm 100 mét vuông thì hy vọng rằng giá bán trung bình của ngôi nhà sẽ tăng thêm khoảng \$14.000 USD.

$\hat{\beta}_1 = 52,35$, không có cách giải thích cho $\hat{\beta}_1$ vì không thể cho rằng khi không ngôi nhà đất thì người mua vẫn phải trả 01 khoản tiền là 52,35 ngàn \$. Nguyên nhân là vì $\hat{\beta}_1$ còn ẩn chứa biến bỏ sót.

2.1.3. Số hạng sai số :



Hình 2.2 : Mô hình hồi quy mẫu và tổng thể

Số hạng sai số u_i (hay còn gọi là số hạng ngẫu nhiên - stochastic disturbance) là thành phần ngẫu nhiên không quan sát được và là sai biệt giữa Y_i và phần xác định $\beta_1 + \beta_2 X_i$

$$u_i = Y_i - \hat{Y}_i \quad (\hat{Y}_i : \text{ giá trị ước lượng (dự báo) của } Y_i)$$

Trong PRF : $u_i = Y_i - E(Y_i/X_i)$ được gọi là sai số (Error), :

còn trong SRF $\hat{u}_i = Y_i - \hat{Y}_i$ được gọi là phần dư (Residual),

Nguyên nhân gây ra sai số :

- **Biến giải thích bị bỏ sót.** Giả sử mô hình thực sự là $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \epsilon_i$, trong đó Z_i là một biến giải thích khác và ϵ_i là số hạng sai số thực sự, nhưng ta lại sử dụng mô hình là $Y_i = \beta_1 + \beta_2 X_i + u_i$. Vì thế, $u_i = \beta_3 Z_i + \epsilon_i$ bao hàm cả ảnh hưởng của biến Z bị bỏ sót.

Trong ví dụ trên, nếu mô hình thực sự bao gồm cả ảnh hưởng của số phòng ngủ và phòng tắm của ngôi nhà lên giá bán và ta đã bỏ qua hai ảnh hưởng này mà chỉ xét đến diện tích sử dụng thì số hạng u sẽ bao hàm cả ảnh hưởng của phòng ngủ và phòng tắm lên giá bán

- **Tính phi tuyến tính :** u_i có thể bao gồm ảnh hưởng phi tuyến tính trong mối quan hệ giữa Y và X . Vì thế, nếu mô hình thực sự là $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \epsilon_i$, nhưng

ta lại sử dụng mô hình là $Y_i = \beta_1 + \beta_2 X_i + u_i$. Vì thế, $u_i = \beta_3 X_i^2 + \epsilon_i$ bao hàm cả

ảnh hưởng của thành phần phi tuyến

- **Sai số đo lường :** Sai số trong việc đo lường X và Y có thể được thể hiện qua u . Sai số này là do : sử dụng các biến thay thế X, Y , cách lấy mẫu, ...
- **Những ảnh hưởng không thể dự báo :** Dù là một mô hình kinh tế lượng tốt cũng có thể chịu những ảnh hưởng ngẫu nhiên không thể dự báo được. Những ảnh hưởng này sẽ luôn được thể hiện qua số hạng sai số u_i .

2.2. ƯỚC LƯỢNG MÔ HÌNH BẰNG PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU THÔNG THƯỜNG (Ordinary Least Of Squares)

Nguyên tắc :

Tiêu chuẩn tối ưu được sử dụng bởi phương pháp bình phương tối thiểu thông thường (OLS) là cực tiểu hóa hàm mục tiêu :

$$ESS = \sum_{i=1}^n \hat{u}_i^2 \rightarrow \text{Min} \quad (\text{viết tắt } \sum \hat{u}_i^2).$$

(ESS = Error Sum of Squares : Tổng bình phương sai số)

Ta có : $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad \text{với } \hat{u}_i = Y_i - \hat{Y}_i$

i

Vậy : $ESS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \rightarrow \text{Min}$

Để tính $\hat{\beta}_1$ và $\hat{\beta}_2$ ta lấy đạo hàm bậc nhất theo $\hat{\beta}_1$ và $\hat{\beta}_2$ và được hệ phương trình chuẩn

$$\begin{aligned} \frac{\partial ESS}{\partial \hat{\beta}_1} &= \sum_{i=1}^n 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-1) = 0 & \text{với } \frac{\partial ESS}{\partial \hat{\beta}_2} &= \sum_{i=1}^n 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-X_i) = 0 \\ \text{hoặc} & \sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0 & \text{hoặc} & \sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) X_i = 0 \end{aligned}$$

= 0

$$\square \hat{u}_i X_i = 0$$

Từ hệ phương trình chuẩn ta suy ra:

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

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

Nếu đặt : $x_i = X_i - \bar{X}$ $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$
 $y_i = Y_i - \bar{Y}$ \textcircled{R}

Ví dụ : Với dữ liệu về giá và diện tích của 14 căn (trong DATA3-1 – Bộ dữ liệu Ramanathan). Tìm $\hat{\beta}_1$, $\hat{\beta}_2$ trong mô hình hồi quy ước lượng sau :

PRF : $PRICE_i = \beta_1 + \beta_2 SQFT_i + u_i$

SRF : $PRICE_i = \hat{\beta}_1 + \hat{\beta}_2 SQFT_i + \hat{u}_i$

Trong đó: $PRICE_i$: Giá mỗi căn nhà (ngàn USD)

$SQFT_i$: diện tích căn nhà (m²)

Từ dữ liệu ta tính được $\bar{Y} = 317,49$ $\bar{X} = 1.910,93$

Bảng 2.1 : Thực hiện hồi quy đơn biến

i	Y _i = PRICE _i	X _i = SQFT _i	Y _i - \bar{Y}	X _i - \bar{X}	(X _i - \bar{X})(Y _i - \bar{Y})	(X _i - \bar{X}) ²
1	199,9	1065	-117,59	-845,93	99.475,16	715.595,15
2	228	1254	-89,49	-656,93	58.790,41	431.555,15
3	235	1300	-82,49	-610,93	50.397,24	373.233,72
4	285	1577	-32,49	-333,93	10.850,29	111.508,29
5	239	1600	-78,49	-310,93	24.405,67	96.676,58
6	293	1750	-24,49	-160,93	3.941,60	25.898,01
7	285	1800	-32,49	-110,93	3.604,39	12.305,15
8	365	1870	47,51	-40,93	-1.944,40	1.675,15
9	295	1935	-22,49	24,07	-541,44	579,43
10	290	1948	-27,49	37,07	-1.019,20	1.374,29
11	385	2254	67,51	343,07	23.159,77	117.698,01
12	505	2600	187,51	689,07	129.205,81	474.819,43
13	425	2800	107,51	889,07	95.581,53	790.448,01
14	415	3000	97,51	1.089,07	106.192,24	1.186.076,58
Tổng cộng					602.099,09	4.339.442,93

Ta có : $\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{602.099,09}{4.339.442,93} = 0,1388$

$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 317,49 - 0,1388 \cdot 1.910,93 = 52,351$

Vậy : $PRICE_i = 52,35 + 0,139 \cdot SQFT_i + \hat{u}_i$

Lưu ý:

✚ *Tính bằng hàm sẵn có của EXCEL :*

$$\hat{\beta}_2 = \text{SLOPE}(\text{known_y's}, \text{known_x's})$$

$$\hat{\beta}_1 = \text{INTERCEPT}(\text{known_y's}, \text{known_x's})$$

✚ *Tính bằng Data Analysis của EXCEL:*

Regression Statistics						
Multiple R	0.905827					
R Square	0.820522					
Adjusted R Square	0.805565					
Standard Error	39.023036					
Observations	14					
ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	83541.44	83541.44	54.8605128	8.19906E-06	
Residual	12	18273.57	1522.797			
Total	13	101815				
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	52.350907	37.2855	1.404056	0.185651	-28.887191	133.5890
SQFT	0.138750	0.018733	7.406788	0.000008	0.097935	0.179566

Hình 2.3 Thực hiện hồi quy đơn biến trên EXCEL

✚ *Tính bằng phần mềm EVIEW :*

Dependent Variable: PRICE				
Method: Least Squares				
Date: 01/04/06 Time: 12:08				
Sample: 1 14				
Included observations: 14				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	52.35091	37.28549	1.404056	0.1857
SQFT	0.138750	0.018733	7.406788	0.0000
R-squared	0.820522	Mean dependent var	317.4929	
Adjusted R-squared	0.805565	S.D. dependent var	88.49816	
S.E. of regression	39.02304	Akaike info criterion	10.29774	
Sum squared resid	18273.57	Schwarz criterion	10.38904	
Log likelihood	-70.08421	F-statistic	54.86051	
Durbin-Watson stat	1.975057	Prob(F-statistic)	0.000008	

Hình 2.4 Thực hiện hồi quy đơn biến trên EVIEW

Lưu ý: Khi có các kết quả tính toán của mô hình từ các phần mềm, sinh viên cần viết ra PRF và SRF.

2.3. CÁC GIẢ THIẾT CƠ BẢN :

- (1) Mô hình hồi quy tuyến tính với các tham số β_1 và β_2 .
- (2) Tất cả các giá trị quan sát X_i không được giống nhau; phải có ít nhất một giá trị khác biệt, nghĩa là $\text{Var}(X_i) > 0$
- (3) Sai số u_i là biến ngẫu nhiên với trung bình bằng không, nghĩa là $E(u_i) = 0$.
- (4) Các giá trị quan sát X_i được cho và không ngẫu nhiên, điều này ngầm định rằng không tương quan với u_i nghĩa là $\text{Cov}(X_i, u_i) = 0$.
- (5) Sai số u_i có phương sai không đổi với mọi i ; nghĩa là $\text{Var}(u_i) = \sigma^2 = \text{const}$.
- (6) Hai sai số u_i và u_s bất kỳ độc lập với nhau đối với $i \neq s$, nghĩa là $\text{Cov}(u_i, u_s) = 0$
- (7) Số quan sát (cỡ mẫu) phải lớn hơn số hệ số hồi quy ước lượng (ở đây $n > 2$).
- (8) Sai số u_i tuân theo phân phối chuẩn $u_i \sim N(0, \sigma^2)$

2.4. TÍNH CHẤT :

Tính chất 1:

$$S_{XX} = \sum (X_i - \bar{X})^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n} = \sum X_i^2 - n\bar{X}^2$$

Tính chất 2:

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} = \sum X_i Y_i - n\bar{X}\bar{Y}$$

Từ tính chất (1) và (2) ta suy ra
$$\hat{\beta}_2 = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

Tính chất 3:

Không thiên lệch : $E(\hat{\beta}_1) = \beta_1$ và $E(\hat{\beta}_2) = \beta_2$ (Do giả thiết 3 và 4)

Tính chất 4:

Tính nhất quán: $\lim_{n \rightarrow \infty} \hat{\beta} = \beta$ (Do giả thiết 2, 3 và 4)

Tính chất 5 (Định lý Gauss–Markov)

Ước lượng OLS là BLUE (Best Linear Unbias Estimation) nếu thỏa mãn giả thiết 2, 3, 4, 5 và 6.

Nghĩa là trong tất cả các tổ hợp tuyến tính không thiên lệch của Y , ước lượng OLS của tham số ước lượng có phương sai bé nhất.

2.5. ĐỘ CHÍNH XÁC CỦA ƯỚC LƯỢNG

Từ lý thuyết xác suất ta biết rằng phương sai của một biến ngẫu nhiên đo lường sự phân tán xung quanh giá trị trung bình. Phương sai càng bé, ở mức trung bình, từng giá trị riêng biệt càng gần với giá trị trung bình. Tương tự, khi đề cập đến khoảng tin cậy, ta biết rằng phương sai của biến ngẫu nhiên càng nhỏ, khoảng tin cậy của các tham số càng bé.

Như vậy, phương sai của một ước lượng là thông số để chỉ độ chính xác của một ước lượng. Do đó việc tính toán phương sai của $\hat{\beta}_1$ và $\hat{\beta}_2$ là luôn cần thiết.

Từ các tính chất và giả thiết ta có các công thức tính toán sau :

✚ Phương sai của sai số :

$$\text{Var}(\hat{u}_i) = \hat{\sigma}^2 = \frac{\sum u_i^2}{n - 2}$$

$$\textcircled{R} \quad \hat{\sigma} = \sqrt{\frac{\sum u_i^2}{n - 2}}$$

✚ Phương sai của hệ số độ dốc :

$$\text{Var}(\hat{\beta}_2) = S_{\hat{\beta}_2}^2 = E(\hat{\beta}_2 - \beta_2) = \frac{\hat{\sigma}^2}{S_{XX}}$$

$$\textcircled{R} \quad \text{SE}(\hat{\beta}_2) = S_{\hat{\beta}_2} = \sqrt{\frac{\hat{\sigma}^2}{S_{XX}}} \quad \text{SE: Standard Error (sai số chuẩn).}$$

✚ Phương sai của hệ số tung độ góc :

$$\text{Var}(\hat{\beta}_1) = S_{\hat{\beta}_1}^2 = E(\hat{\beta}_1 - \beta_1) = \frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{S_{XX}}$$

$$\textcircled{R} \quad \text{SE}(\hat{\beta}_1) = S_{\hat{\beta}_1} = \sqrt{\frac{\sum X_i^2}{n} \cdot \frac{\hat{\sigma}^2}{S_{XX}}}$$

✚ Đồng phương sai của hệ số độ dốc và tung độ góc :

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = S_{\hat{\beta}_1 \hat{\beta}_2} = \frac{\bar{X} \hat{\sigma}^2}{S_{XX}}$$

2.6. ĐỘ THÍCH HỢP CỦA MÔ HÌNH:

Hình 2.2 cho thấy rõ rằng không có đường thẳng nào hoàn toàn “thích hợp” với các dữ liệu bởi vì có nhiều giá trị dự báo bởi đường thẳng cách xa với giá trị thực tế. Để có thể đánh giá một mối quan hệ tuyến tính mô tả những giá trị quan sát có tốt hơn một mối quan hệ tuyến tính khác hay không, cần phải có các chỉ tiêu toán học đo lường độ thích hợp. Hệ số xác định R^2 là một chỉ tiêu được đề nghị.

📌 Tổng bình phương toàn phần (Total Sum of Squares – TSS)

$$TSS = \sum (Y_i - \bar{Y})^2$$

📌 Tổng bình phương sai số (Error Sum of Squares – ESS)

$$ESS = \sum (Y_i - \hat{Y}_i)^2$$

📌 Tổng bình phương hồi quy (Regression Sum of Squares – RSS)

$$RSS = \sum (\hat{Y}_i - \bar{Y})^2$$

Người ta đã chứng minh được : $TSS = ESS + RSS$

📌 Hệ số xác định:

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = \hat{\beta}_2 \frac{S_{XX}}{S_{YY}}$$

Vậy : $-1 \leq R^2 \leq 1$

Ý nghĩa của R^2 : % sự thay đổi của biến phụ thuộc được giải thích bởi mô hình.

Trong mô hình hồi quy đơn biến: $R^2 = r^2$ (r là hệ số tương quan mẫu)

2.7. KHOẢNG TIN CẬY VÀ KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ:

Kiểm định giả thuyết thống kê là một trong những nhiệm vụ chính của nhà kinh tế lượng. Như ta vừa tính toán trong ví dụ trên, thông số ước lượng $\hat{\beta}_1$ và $\hat{\beta}_2$ là một đại lượng khác không nhưng thông số hồi quy của tổng thể β_1 và β_2 liệu có khác 0 hay không, điều này có ý nghĩa gì. Kiểm định các giả thuyết thống kê sẽ giúp ta trả lời được câu hỏi này.

2.7.1. Đối với tham số độ dốc:

Trong nghiên cứu kinh tế lượng, người ta hay dùng kiểm định 02 phía đối với kiểm định các tham số riêng lẻ. Kết quả kiểm định này trên các phần mềm chuyên dụng là kiểm định 02 phía.

Giả thuyết kiểm định:

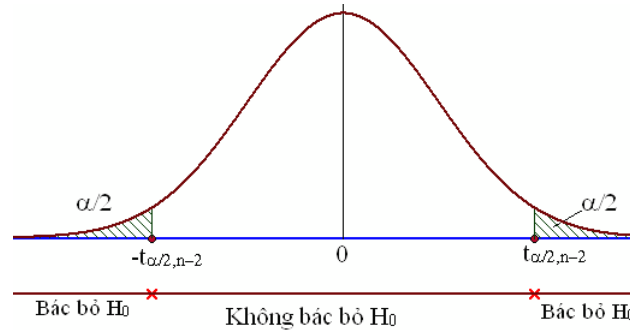
$H_0: \beta_2 = 0$ □ X không ảnh hưởng đối với Y.

$H_1: \beta_2 \neq 0$ □ X có ảnh hưởng đối với Y.

Kỳ vọng: Bác bỏ H_0 .

Xác Định Trị Thống Kế Kiểm Định

$$t_c = \frac{\hat{\beta}_2}{S_{\hat{\beta}_2}}, \quad t^* = t_{(2, n-2)}$$



Kết luận: Nếu $|t_c| > t_{\alpha/2, n-2} \Rightarrow$ Bác bỏ H_0

Lưu ý: Trong các phần mềm ứng dụng giá trị p -value (hay Significant) là giá trị xác suất tương ứng với t_c với bậc tự do $n-2$ thường được tính toán sẵn, ta có thể dùng giá trị này để kết luận nhanh.

Nếu p -value $\delta < \alpha \Rightarrow$ Bác bỏ H_0

Khoảng tin cậy của β_2 :

$$\hat{\beta}_2 \pm t_{\alpha/2, n-2} \cdot S_{\hat{\beta}_2}$$

Đặt $a = \hat{\beta}_2 - t_{\alpha/2, n-2} \cdot S_{\hat{\beta}_2}$; $b = \hat{\beta}_2 + t_{\alpha/2, n-2} \cdot S_{\hat{\beta}_2}$ $\Rightarrow a < \beta_2 < b$

Nếu a, b cùng dấu \Rightarrow Bác bỏ H_0

Ví dụ: Quay lại với ví dụ

$$PRF : PRICE_i = \beta_1 + \beta_2 SQFT_i + u_i$$

$$SRF : PRICE_i = \hat{\beta}_1 + \hat{\beta}_2 SQFT_i + \hat{u}_i$$

$$PRICE_i = 52,35 + 0,139 \cdot SQFT_i + \hat{u}_i$$

Giả thuyết kiểm định với $\alpha = 5\%$:

$H_0: \beta_2 = 0 \Rightarrow$ Diện tích sử dụng không ảnh hưởng đến giá nhà.

$H_1: \beta_2 \neq 0 \Rightarrow$ Diện tích sử dụng ảnh hưởng đến giá nhà.

Theo kết quả hình 2.3 (hoặc 2.4) ta có 3 cách để đưa ra kết luận:

- $\square t_c = 7,406788 < t_{\alpha/2, n-2} = t_{2,5\%, 12} \Rightarrow$ Bác bỏ H_0 .
- \square hay p -value = 0,000 $< \alpha \Rightarrow$ Bác bỏ H_0 .
- $\square 0,979 < \beta_2 < 0,1796 \Rightarrow$ Bác bỏ H_0 .

Vậy: Diện tích sử dụng ảnh hưởng đến giá nhà có ý nghĩa về mặt thống kê.

Lưu ý: Chỉ sử dụng 01 trong 03 cách trong thực hành.

📌 Kiểm định 01 phía:

Nếu biết dấu kỳ vọng của β_2 ta có thể thực hiện kiểm định 01 phía với các giả thiết :

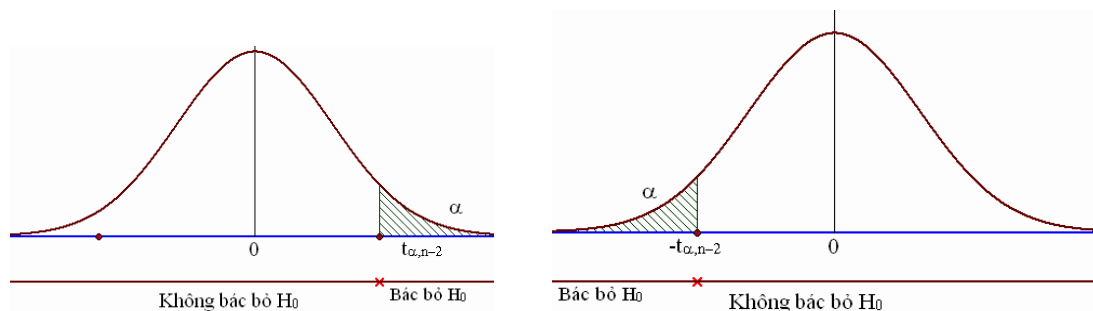
$$H_0: \beta_2 = 0 \quad (\beta_2 \geq 0) \quad \text{hoặc} \quad H_0: \beta_2 = 0 \quad (\beta_2 \leq 0)$$

$$H_1: \beta_2 < 0 \quad \quad \quad \hat{H}_1: \beta_2 > 0$$

Khi đó miền bác bỏ là :

$$t_c < -t_{\alpha, n-2}$$

$$t_c > t_{\alpha, n-2}$$



2.7.2. Đối với tham số tung độ góc.

Trình bày tương tự cho kiểm định các giả thiết thống kê và tính khoảng tin cậy của β_1 . Tuy nhiên thông thường ta không quan tâm đến việc kiểm định các giả thiết thống kê cho tham số tung độ góc β_1 nguyên nhân là do không có cách giải thích phù hợp cho tham số này.

Kiểm định độ thích hợp của mô hình sẽ được trình bày trong chương Hồi quy đa biến.

2.8. TRÌNH BÀY KẾT QUẢ HỒI QUY

Các kết quả của phân tích hồi quy được trình bày theo nhiều cách. Theo cách thông thường, người ta sẽ viết phương trình ước lượng kèm với các trị thống kê t ở dưới mỗi hệ số hồi quy như sau:

$$\begin{aligned} \text{PRICE} = & 52,315 + 0,139\text{SQFT} + \hat{u}_i \\ & (1,404) \quad (7,41) \\ R^2 = & 0,821 \quad \text{df} = 12 \quad \sigma = 39,023 \end{aligned}$$

Bậc tự do (df: Degree Free).

Một cách khác là điền các sai số chuẩn dưới các hệ số hồi quy.

Tuy nhiên, ngày nay với các tiện ích của các phần mềm máy tính người ta có thể trình bày kết quả hồi quy bằng in các bảng tính từ các

2.9. THANG ĐO:

Giả sử chúng ta đã tính PRICE theo đơn vị đồng đôla thay vì theo ngàn đồng đôla. Cột PRICE ở bảng tính sẽ chứa các giá trị như 199.900, 228.000, v.v. Những ước lượng của hệ số hồi quy, các sai số chuẩn của chúng, R^2 , ... sẽ bị ảnh hưởng như thế nào bởi sự thay đổi đơn vị này?

Ta có :

$$(1) \text{ PRICE}_i = \beta_1 + \beta_2 \text{SQFT}_i + u_i \quad (\text{trong đó PRICE}_i \text{ tính bằng ngàn USD})$$

$$(2) \text{ PRICE}_i^* = \gamma_1 + \gamma_2 \text{SQFT}_i + \epsilon_i \quad (\text{trong đó PRICE}_i^* \text{ tính bằng USD})$$

Ⓜ $\text{PRICE}_i^* = 1000 \cdot \text{PRICE}_i$ thay vào (2) ta được :

$$1000 \cdot \text{PRICE}_i = \gamma_1 + \gamma_2 \text{SQFT}_i + \epsilon_i$$

$$\text{Ⓜ } \text{PRICE}_i = \frac{\gamma_1}{1000} + \frac{\gamma_2}{1000} \text{SQFT}_i + \frac{\epsilon_i}{1000}$$

$$\text{Vậy: } \beta_1 = \frac{\gamma_1}{1000}; \quad \beta_2 = \frac{\gamma_2}{1000} \quad u_i = \frac{\epsilon_i}{1000}$$

$$\text{Hay: } \gamma_1 = 1000\beta_1; \quad \gamma_2 = 1000\beta_2; \quad \epsilon_i = 1000u_i$$

Kết luận : Việc thay đổi đơn vị là cho các hệ số ước lượng và các sai số chuẩn của chúng tăng 1000 lần, còn R^2 không thay đổi.

Làm tương tự với sự thay đổi đơn vị của các biến khác.

Chương III

MÔ HÌNH HỒI QUY BỘI (Multiple Linear Regression Model)

3.1. MÔ HÌNH CƠ BẢN:

Trong lý thuyết cũng như trong thực tế, có nhiều trường hợp mà biến kinh tế cho trước không thể giải thích bằng các mô hình hồi quy đơn, Ví dụ :

- (1) Lượng cầu phụ thuộc vào giá, thu nhập, giá các hàng hoá khác v.v... Nhớ lại lý thuyết về hành vi người tiêu dùng.

$$Q^D = f(P, I, P_s, P_c, \text{Market size}, P_f (\text{giá kỳ vọng}), T (\text{thị hiếu}))$$

- (2) Sản lượng phụ thuộc vào giá, các nhập lượng ban đầu, các nhập lượng trung gian, công nghệ v.v... Nhớ lại lý thuyết về hàm sản xuất.

$$Q^S = f(K, L, \text{TECH})$$

- (3) Tốc độ tăng trưởng của nền kinh tế phụ thuộc vào chi tiêu đầu tư, lượng lao động, thay đổi công nghệ. Nhớ lại lý thuyết về hàm tổng sản xuất.
- (4) Lương phụ thuộc vào trình độ giáo dục, kinh nghiệm, giới tính, độ tuổi . . .
- (5) Giá nhà ở phụ thuộc vào diện tích nhà, số phòng ngủ và số phòng tắm . . .
- (6) Chi tiêu của hộ gia đình về thực phẩm phụ thuộc vào qui mô hộ gia đình, thu nhập, vị trí địa lý . . .
- (7) Tỷ lệ tử vong trẻ em của quốc gia phụ thuộc vào thu nhập bình quân đầu người, trình độ giáo dục . . .
- (8) ...

Trong Chương 2 ta giới hạn trong trường hợp đơn giản của mô hình hồi quy đơn biến. Bây giờ, chúng ta sẽ xem xét mô hình hồi quy tuyến tính đa biến (gọi tắt là mô hình hồi quy bội), nghĩa là tìm mối liên hệ giữa biến phụ thuộc Y với nhiều biến độc lập X.

Mô hình hồi quy bội tổng thể PRF và mô hình hồi quy bội mẫu SRF tổng quát có dạng như sau :

$$\text{PRF : } Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \dots + \beta_k X_{ki} + u_i$$

$$\text{SRF : } Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + \dots + \hat{\beta}_k X_{ki} + \hat{u}_i$$

Trong đó : Y_i là quan sát thứ i của biến Y

$X_{2i}, X_{3i}, \dots, X_{ki}$ là quan sát thứ i của các biến X_2, X_3, \dots, X_k

$u_i, (\hat{u}_i)$ là sai số (phần dư) của mô hình.

3.1.1. Ý nghĩa các hệ số:

Ta có : $E(Y_i/X_{si}) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \dots + \beta_k X_{ki}$

$$\beta_k = \frac{\Delta Y}{\Delta X_k}$$

β_k : Độ dốc riêng phần của đường hồi quy tổng thể đối với biến X_k , nghĩa là khi giữ các biến độc lập khác cố định, khi X_k thay đổi một đơn vị thì ở mức trung bình, Y thay đổi β_k đơn vị. Vì vậy β_2 được diễn dịch là ảnh hưởng cận biên riêng phần của X_k lên Y .

Nói cách khác hệ số độ dốc riêng phần ảnh hưởng ròng (net effect) hoặc ảnh hưởng trực tiếp (direct effect) lên biến phụ thuộc khi biến giải thích thay đổi một đơn vị sau khi đã loại trừ ảnh hưởng của các biến hồi qui khác.

β_1 : Tung độ gốc của đường hồi quy tổng thể, và là giá trị của trị trung bình Y khi X bằng 0. Tuy nhiên sẽ không có cách giải thích cho β_1 vì nguyên nhân là vì β_1 còn ẩn chứa biến bỏ sót (ngoài mô hình).

Tương tự cho cách giải thích thông số ước lượng.

Ví dụ : Cũng như trước, giá được tính bằng đơn vị ngàn đô la, nhưng ngoài diện tích sử dụng, giá còn liên hệ với số phòng ngủ cũng như số phòng tắm nên mô hình hồi quy bội được xác định như sau:

$$PRF : PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3 BATHS_i + \beta_4 BEDRMS_i + u_i$$

$$SRF : PRICE_i = \hat{\beta}_1 + \hat{\beta}_2 SQFT_i + \hat{\beta}_3 BATHS_i + \hat{\beta}_4 BEDRMS_i + \hat{u}_i$$

$$SRF : PRICE_i = 129.062 + 0.1548SQFT_i - 12.193BATHS_i - 21.588BEDRMS_i + \hat{u}_i$$

Trong đó :

PRICE : Giá nhà (ngàn USD)

SQFT : Diện tích sử dụng (m²)

BATHS : Số phòng tắm

BEDRMS: Số phòng ngủ.

Cách giải thích các thông số ước lượng của mô hình :

$\hat{\beta}_2 = 0.1548$, điều đó có nghĩa là : theo thông tin của mẫu, khi số phòng tắm và số phòng ngủ được cố định, một mét vuông diện tích tăng thêm sẽ làm tăng giá bán nhà tăng lên, ở mức trung bình, 0.1548 ngàn USD.

Tương tự với cách giải thích cho $\hat{\beta}_3, \hat{\beta}_4$.

3.2. ƯỚC LƯỢNG MÔ HÌNH CƠ BẢN BẰNG OLS:

Cũng như mô hình hồi quy tuyến tính đơn, các tham số của mô hình hồi quy tuyến tính đa biến cũng được ước lượng bằng phương pháp OLS.

$$ESS = \sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2 \rightarrow \text{Min}$$

$$\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki})^2 \rightarrow \text{Min}$$

Hệ phương trình chuẩn

$$\frac{\partial ESS}{\partial \hat{\beta}_1} = 0 \quad \rightarrow \quad 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki}) = 0$$

$$\frac{\partial ESS}{\partial \hat{\beta}_2} = 0 \quad \rightarrow \quad 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki}) X_{2i} = 0$$

$$\dots$$

$$\frac{\partial ESS}{\partial \hat{\beta}_k} = 0 \quad \rightarrow \quad 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_k X_{ki}) X_{ki} = 0$$

Để tìm được các hệ số của SRF ta sẽ giải hệ phương trình chuẩn.

Các tham số ước lượng của mô hình hồi quy 02 biến độc lập:

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

SRF : $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3$$

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

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

Với $\sum X_{ki} = \sum X_k$
 $\sum y_i = \sum Y$

Việc tính toán các tham số ước lượng sẽ càng trở nên khó khăn nếu mô hình hồi qui của chúng ta càng có nhiều biến giải thích. Tuy nhiên với sự trợ giúp của phần mềm chuyên dụng như EXCEL, EVIEWS, SPSS, MINITAB ... chúng ta có thể tìm được giá trị các ước lượng của mô hình hồi qui một cách nhanh chóng.

Ví dụ : Bổ sung dữ liệu về giá và diện tích, số phòng tắm, số phòng ngủ của 14 căn
 Tìm $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4$ trong mô hình hồi quy ước lượng sau :

$$PRF : PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3 BATHS_i + \beta_4 BEDRMS_i + u_i$$

$$SRF : PRICE_i = \hat{\beta}_1 + \hat{\beta}_2 SQFT_i + \hat{\beta}_3 BATHS_i + \hat{\beta}_4 BEDRMS_i + \hat{u}_i$$

📊 Dùng phần mềm EXCEL ta được:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.914317
R Square	0.835976
Adjusted R Square	0.786769
Standard Error	40.865717
Observations	14

ANOVA

	df	SS	MS	F	Significance F
Regression	3	85114.941	28371.647	16.988940	0.000299
Residual	10	16700.069	1670.007		
Total	13	101815.009			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	129.0616	88.3033	1.4616	0.1746	-67.6903	325.8136
SQFT	0.1548	0.0319	4.8465	0.0007	0.0836	0.2260
BATHS	-12.1928	43.2500	-0.2819	0.7838	-108.5598	84.1743
BEDRMS	-21.5875	27.0293	-0.7987	0.4430	-81.8126	38.6376

Hình 3.1: Hồi quy đa biến trên EXCEL

📊 Dùng phần mềm EVIEWS ta được:

Dependent Variable: PRICE				
Method: Least Squares				
Date: 05/29/05 Time: 17:50				
Sample: 1 14				
Included observations: 14				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	129.0616	88.30326	1.461573	0.1746
SQFT	0.154800	0.031940	4.846516	0.0007
BATHS	-12.19276	43.25000	-0.281913	0.7838
BEDRMS	-21.58752	27.02933	-0.798670	0.4430
R-squared	0.835976	Mean dependent var	317.4929	
Adjusted R-squared	0.786769	S.D. dependent var	88.49816	
S.E. of regression	40.86572	Akaike info criterion	10.49342	
Sum squared resid	16700.07	Schwarz criterion	10.67600	
Log likelihood	-69.45391	F-statistic	16.98894	
Durbin-Watson stat	1.970415	Prob(F-statistic)	0.000299	

Hình 3.2: Hồi quy đa biến trên EVIEWS

Mô hình hồi quy mẫu:

$$PRICE_i = 129.0616 + 0.1548 * SQFT_i - 12.19276 * BATHS_i - 21.58752 * BEDRMS_i + \hat{u}_i$$

3.3. CÁC GIẢ THIẾT CƠ BẢN :

Mô hình hồi quy tuyến tính đa biến tuân thủ các giả thiết cơ bản như mô hình hồi quy tuyến tính đơn biến như:

- (1) Mô hình hồi quy tuyến tính với các tham số.
- (2) Tất cả các giá trị quan sát của mỗi biến X_{ki} không được giống nhau; phải có ít nhất một giá trị khác biệt, nghĩa là $\text{Var}(X_{ki}) > 0$
- (3) Sai số u_i là biến ngẫu nhiên với trung bình bằng không, nghĩa là $E(u_i/X_s) = 0$.
- (4) Các giá trị quan sát X_{ki} được cho và không ngẫu nhiên, điều này ngầm định rằng không tương quan với u_i nghĩa là $\text{Cov}(X_{ki}, u_i) = 0$.
- (5) Sai số u_i có phương sai không đổi với mọi i ; nghĩa là $\text{Var}(u_i/X_s) = \sigma^2 = \text{const}$.
- (6) Hai sai số u_i và u_s bất kỳ độc lập với nhau đối với mọi $i \neq s$, nghĩa là $\text{Cov}(u_i, u_s) = 0$
- (7) Số quan sát (cỡ mẫu) phải lớn hơn số hệ số hồi quy ước lượng (ở đây $n > k$).
- (8) Sai số u_i tuân theo phân phối chuẩn $u_i \sim N(0, \sigma^2)$

Ngoài ra, mô hình hồi quy tuyến tính đa biến có các giả thiết bổ sung :

- (9) Không nhận dạng sai mô hình (không sai dạng hàm, không thiếu biến quan trọng và thừa biến không quan trọng).
- (10) Không có hiện tượng đa cộng tuyến hoàn hảo (perfect multicollinearity hoặc exact linear relationship) trong mô hình.

Tức là, không tồn tại tập hợp các hệ số thỏa mãn biểu thức sau với mọi i :

$$1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} = 0$$

3.4. TÍNH CHẤT :

Ước lượng OLS cho mô hình hồi quy tuyến tính đa biến là BLUE (Best Linear Unbias Estimation).

Nghĩa là trong tất cả các tổ hợp tuyến tính không thiên lệch của Y , ước lượng OLS của tham số ước lượng có phương sai bé nhất.

3.5. ĐỘ CHÍNH XÁC CỦA ƯỚC LƯỢNG

Do sự phức tạp của công thức tính phương sai của các tham số ước lượng và sự hỗ trợ tính toán nhanh chóng của các phần mềm, nên trong tập bài giảng này chỉ trình bày công thức tính phương sai và sai số chuẩn của các thông số ước lượng cho mô hình hồi quy tuyến tính 02 biến giải thích.

$$\text{Var}(\hat{u}_i) = \hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n - k}$$

$$\textcircled{R} \quad \hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n - k}}$$

$$\text{VAR}[\hat{\beta}_2] = \frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \hat{\sigma}^2$$

$$\text{SE}_{\hat{\beta}_2} = \sqrt{\frac{\sum x_{3i}^2}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}} \hat{\sigma}$$

3.6. ĐỘ THÍCH HỢP CỦA MÔ HÌNH:

Trong mô hình hồi quy tuyến tính bội để loại bỏ ảnh hưởng việc thêm biến làm bậc tự do của mô hình giảm và làm hệ số xác định tăng người ta hay dùng một đại lượng đo lường độ thích hợp của mô hình là hệ số xác định điều chỉnh \bar{R}^2

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}}$$

$$\bar{R}^2 = 1 - \frac{\text{ESS}/(n - k)}{\text{TSS}/(n - 1)} = 1 - \frac{n - 1}{n - k} (1 - R^2)$$

Lưu ý: \bar{R}^2 có thể xảy ra trường hợp âm.

Ngoài \bar{R}^2 , người ta có đề nghị một số tiêu chuẩn khác để đo lường độ thích hợp của mô hình, sau đây là 02 tiêu chuẩn được tính toán sẵn trong phần mềm EVIEW:

○ Akaike Info Criterion: $\text{AIC} = \frac{\sum \text{ESS}}{n} e^{\frac{2k}{n}}$

○ Schwarz criterion: $\text{SCHWARZ} = \frac{\sum \text{ESS}}{n} e^{\frac{k}{n}}$

Với 2 tiêu chuẩn AIC và SCHWARZ, mô hình có 02 tiêu chuẩn này càng nhỏ thì độ thích hợp của dữ liệu đối với mô hình càng cao.

3.7. KIỂM ĐỊNH CÁC GIẢ THUYẾT THỐNG KÊ:

3.7.1. Kiểm định các tham số riêng lẻ:

Như trong Chương 2, kiểm định giả thuyết về một tham số hồi qui riêng lẻ được tiến hành bằng kiểm định t.

Giả thuyết kiểm định:

$H_0: \beta_k = 0$ \square X_k không có ảnh hưởng đối với Y .

$H_1: \beta_k \neq 0$ \square X_k có ảnh hưởng đối với Y .

Kỳ vọng: Bác bỏ H_0 .

Xác Định Trị Thống Kế Kiểm Định

$$t_c = \frac{\hat{\beta}_k}{S_{\hat{\beta}_k}}, \quad t^* = t_{\alpha/2, n-k}$$

Kết luận: Nếu $|t_c| > t_{\alpha/2, n-k} \square$ Bác bỏ H_0

Hay p-value $< \alpha \square$ Bác bỏ H_0

Nếu biết dấu kỳ vọng của β_k ta có thể thực hiện kiểm định 01 phía với các giả thiết :

$$\begin{array}{ll} H_0: \beta_k = 0 (\beta_k \geq 0) & \text{hoặc} & H_0: \beta_k = 0 (\beta_k \leq 0) \\ H_1: \beta_k < 0 & & H_1: \beta_k > 0 \end{array}$$

Khi đó miền bác bỏ là :

$$t_c < - t_{\alpha, n-k} \qquad t_c > t_{\alpha, n-k}$$

3.7.2. Kiểm định Wald (Kiểm định tổ hợp các tham số):

Kiểm định này giúp xác định nên thêm vào hay bớt đi một nhóm biến trong mô hình, hay nói cách khác giúp chọn một trong hai mô hình sau:

- Mô hình nhiều biến - gọi là mô hình không giới hạn (Unrestricted model – U)
- Mô hình ít biến - gọi là mô hình giới hạn (Restricted model – Ký hiệu là R)

Phương trình hồi quy tổng thể của 02 mô hình này như sau :

$$\begin{array}{l} (U): \quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_m X_{mi} + \beta_{m+1} X_{(m+1)i} + \dots + \beta_k X_{ki} + u_i \\ (R): \quad Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_m X_{mi} + e_i \end{array}$$

Giả thuyết kiểm định:

$$H_0: \beta_{m+1} = \beta_{m+2} = \dots = \beta_k = 0 \qquad \square \text{ Chọn mô hình}$$

$$(R). H_1: \text{Có ít nhất 01 số } \beta_j \text{ khác 0 (với } j = (m+1), k) \square \text{ Chọn mô hình}$$

(U).

Xác Định Trị Thống Kế Kiểm Định

$$F_c = \frac{(ESS_R - ESS_U)/(k - m)}{ESS_U/(n - k)} = \frac{(R_U^2 - R_R^2)/(k - m)}{(1 - R_U^2)/(n - k)}$$

$$F^* = F_{\alpha, k-m, n-k}$$

Kết luận:

Nếu : $F_c > F^*$ (hay p-value $< \alpha$) \square Bác bỏ H_0

Ví dụ:

(U): $PRICE_i = \beta_1 + \beta_2 SQFT_i + \beta_3 BATHS_i + \beta_4 BEDRMS_i + u_i$

(R): $PRICE_i = \gamma_1 + \gamma_2 SQFT_i + e_i$

Giả thuyết kiểm định:

$H_0: \beta_3 = \beta_4 = 0$ Chọn mô hình (R).

$H_1: \beta_3 \neq 0$ hoặc $\beta_4 \neq 0$ Chọn mô hình (U).

Xác Định Trị Thống Kế Kiểm Định

$F_c = 0.471106 < F^* = F_{5\%, 2; 10} = 4,103$ Không bác bỏ H_0 Chọn mô hình (R).

🔧 Thực hành trong EVIEW:

- Thực hiện hồi quy mô hình U (hình 3.2)
- Thực hiện kiểm định Wald: View/Coefficient Tests/ Wald Test
- Kết quả:

Wald Test:			
Equation: EQ01			
Test Statistic	Value	df	Probability
F-statistic	0.471106	(2, 10)	0.6375
Chi-square	0.942211	2	0.6243
Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
C(3)	-12.19276	43.25000	
C(4)	-21.58752	27.02933	
Restrictions are linear in coefficients.			

Hình 3.3: Thực hiện kiểm định Wald trên EVIEW

Ta thấy: $p\text{-value} = 0.6375 > 5\%$ Không bác bỏ H_0 với mức ý nghĩa $\alpha = 5\%$

Chọn mô hình (R).

3.7.3. Kiểm định độ thích hợp của mô hình:

Kiểm định này giúp xác định các biến độc lập trong mô hình có ảnh hưởng lên biến phụ thuộc Y hay không. Đây là một trường hợp đặc biệt của kiểm định Wald.

Ta có:

(PRF): $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$

Giả thuyết kiểm định:

$H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$ Tất cả các biến không tác động đến


Y $H_1: \text{Có ít nhất 1 số } \beta_j \text{ khác } 0 \text{ (với } j = \overline{2; k})$

Đây chính là trường hợp đặc biệt của kiểm định Wald. Trị thống kê kiểm định được tính toán như sau:

$$F_c = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \quad F^* = F_{\langle k-1, n-k \rangle}$$

Kết luận:

Nếu : $F_c > F^*$ (hay p-value $< \langle \rangle$) \square Bác bỏ H_0

 Trên EVIEW, kiểm định này được thực hiện kèm trong bảng kết quả chạy mô hình (hình 3.1 và 3.2):

$$(PRF): \text{PRICE}_i = \beta_1 + \beta_2 \text{SQFT}_i + \beta_3 \text{BATHS}_i + \beta_4 \text{BEDRMS}_i + u_i$$

Giả thuyết kiểm định:

$$H_0: \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1: \text{Có ít nhất 1 số } \beta_j \text{ khác 0 (với } j = \overline{2;4} \text{)}$$


Kết quả : $F_c = 16.98894$ hay p-value = 0.000299 $< 5\%$ \square Bác bỏ H_0 với $\langle = 5\%$

\square Có ít nhất một biến trong mô hình tác động đến giá nhà.


3.8. CHIẾN LƯỢC XÂY DỰNG MÔ HÌNH:

Trong giả thiết bổ sung cho mô hình hồi quy tuyến tính đa biến, ta có một giả thiết là: “Không nhận dạng sai mô hình”. Trong phần này ta sẽ khảo sát một số vấn đề liên quan đến nhận dạng sai mô hình là: mô hình thiếu biến quan trọng hoặc thừa biến không quan trọng.

3.8.1. Hậu quả của việc thiếu biến quan trọng hoặc thừa biến không quan trọng:

 Mô hình thiếu biến quan trọng :

- \square Nếu một biến độc lập quan trọng của mô hình bị bỏ sót (tham số hồi qui khác không có ý nghĩa thống kê) thì các giá trị ước lượng của tất cả các tham số hồi qui còn lại sẽ bị thiên lệch.
- \square Các giá trị dự báo cũng bị thiên lệch.
- \square Ước lượng phương sai của các tham số hồi qui cũng sẽ bị thiên lệch, và vì vậy các kiểm định giả thuyết sẽ không có ý nghĩa.

 Mô hình thừa biến không quan trọng :

- \square Nếu một biến độc lập có tham số hồi qui bằng không (nghĩa là, biến này là thừa) được đưa vào mô hình, các giá trị ước lượng của các tham số hồi qui khác vẫn sẽ không thiên lệch và nhất quán.

- Phương sai của tham số hồi qui sẽ cao hơn các giá trị khi không có biến không liên quan, và vì vậy các hệ số sẽ không hiệu quả.
- Vì các phương sai ước lượng của các hệ số hồi qui là không thiên lệch, các kiểm định giả thuyết vẫn có hiệu lực.

3.8.2. Chiến lược xây dựng mô hình:

Từ hai sai lầm này, để xác định mô hình đúng người ta thường xây dựng mô hình theo một trong hai chiến lược:

✚ Chiến lược từ tổng quát đến đơn giản:

Từ mô hình nhiều biến sau đó loại dần các biến không quan trọng để được mô hình đúng có ít biến hơn.

✚ Chiến lược từ đơn giản đến tổng quát

Từ mô hình ít biến sau đó thêm dần các biến quan trọng để được mô hình đúng có nhiều biến hơn.

Ví dụ: Lập mô hình xác định các nhân tố ảnh hưởng đến việc di chuyển bằng xe buýt của 40 thành phố khắp nước Mỹ. Các biến được đề nghị như sau:

BUSTRAVL : Mức độ giao thông bằng xe buýt ở đô thị tính (ngàn hành khách/ giờ)

FARE : Giá vé xe buýt (USD)

GASPRICE : Giá một ga lông nhiên liệu (USD)

INCOME : Thu nhập bình quân đầu người (USD)

POP : Dân số thành phố (ngàn người)

DENSITY : Mật độ dân số tính (người/dặm vuông)

LANDAREA : Diện tích thành phố (dặm vuông)

✚ Đặc trưng tổng quát của mô hình, được cho dưới đây:

$$BUSTRAV = \beta_1 + \beta_2FARE + \beta_3GASPRICE + \beta_4INCOME + \beta_5POP + \beta_6DENSITY + \beta_7LANDAREA + u$$

Trước khi ước lượng mô hình, ta xác định dấu kỳ vọng của các biến như sau:

Giá vé xe buýt (FARE) tăng sẽ làm nhu cầu đi xe buýt giảm (lý thuyết kinh tế) nghĩa là làm mức độ giao thông bằng xe buýt ở đô thị (BUSTRAV) giảm, do đó ta kỳ vọng β_2 mang dấu âm.

Tương tự với các hệ số khác ta kỳ vọng: β_2 (âm), β_3 (dương), β_4 (có thể âm hoặc dương), β_5 (dương), β_6 (dương) và β_7 (âm).

🚩 Mô hình tổng quát sẽ được chạy như sau:

Dependent Variable: BUSTRAVL				
Method: Least Squares				
Date: 01/04/06 Time: 13:16				
Sample: 1 40				
Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2744.680	2641.672	1.038994	0.3064
FARE	-238.6544	451.7281	-0.528314	0.6008
GASPRICE	522.1132	2658.228	0.196414	0.8455
INCOME	-0.194744	0.064887	-3.001294	0.0051
POP	1.711442	0.231364	7.397176	0.0000
DENSITY	0.116415	0.059570	1.954253	0.0592
LANDAREA	-1.155230	1.802638	-0.640855	0.5260
R-squared	0.921026	Mean dependent var	1933.175	
Adjusted R-squared	0.906667	S.D. dependent var	2431.757	
S.E. of regression	742.9113	Akaike info criterion	16.21666	
Sum squared resid	18213267	Schwarz criterion	16.51221	
Log likelihood	-317.3332	F-statistic	64.14338	
Durbin-Watson stat	2.082671	Prob(F-statistic)	0.000000	

Hình 3.4: Mô hình tổng quát (không giới hạn)

Theo kết quả ta thấy với mức ý nghĩa $\alpha = 10\%$, có 04 biến FARE, GASPRICE, LANDAREA không có ý nghĩa thống kê.

🚩 Thực hiện kiểm định Wald để bỏ 03 biến này ra khỏi mô hình:

$$H_0: \beta_2 = \beta_3 = \beta_7 = 0$$

$$H_1: \beta_2 \neq 0 \text{ hoặc } \beta_3 \neq 0 \text{ hoặc } \beta_7 \neq 0$$

Chạy EVIEW ta có kết quả:

Wald Test:			
Equation: Untitled			
Test Statistic	Value	df	Probability
F-statistic	0.315845	(3, 33)	0.8138
Chi-square	0.947535	3	0.8139
Null Hypothesis Summary:			
Normalized Restriction (= 0)	Value	Std. Err.	
C(2)	-238.6544	451.7281	
C(3)	522.1132	2658.228	
C(7)	-1.155230	1.802638	
Restrictions are linear in coefficients.			

Hình 3.5: Kiểm định Wald để chọn mô hình

Ta thấy $F_c = 0.315845$ và $p\text{-value} = 0.8138 > \alpha$ □ Không bác bỏ H_0 □ Chọn mô hình R, hay nói cách khác nên bỏ 03 biến FARE, GASPRICE, LANDAREA ra khỏi mô hình.

🚩 Mô hình mới (R) là:

Dependent Variable: BUSTRAVL				
Method: Least Squares				
Date: 01/04/06 Time: 13:31				
Sample: 1 40				
Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2815.703	976.3007	2.884053	0.0066
INCOME	-0.201273	0.062101	-3.241076	0.0026
POP	1.576575	0.120612	13.07148	0.0000
DENSITY	0.153421	0.034898	4.396311	0.0001
R-squared	0.918759	Mean dependent var	1933.175	
Adjusted R-squared	0.911989	S.D. dependent var	2431.757	
S.E. of regression	721.4228	Akaike info criterion	16.09497	
Sum squared resid	18736228	Schwarz criterion	16.26386	
Log likelihood	-317.8993	F-statistic	135.7080	
Durbin-Watson stat	1.878671	Prob(F-statistic)	0.000000	

Hình 3.6: Mô hình đơn giản (giới hạn)

Tất cả các biến trong mô hình đều có ý nghĩa thống kê.

Vậy:

$$BUSTRAVL = 2815.7 - 0.2013INCOME + 1.5766POP + 0.1534DENSITY + u^{\wedge}$$

Lưu ý: Trong thực hành trên EVIEWS, khi thực hiện bỏ bớt biến từ mô hình (U) ta nên bỏ từng biến một (Gợi ý: biến không quan trọng có p-value lớn nhất).

3.9. HỘI QUY VỚI BIẾN ĐỊNH TÍNH:

Tất cả các biến chúng ta gặp trước đây đều có bản chất định lượng; nghĩa là các biến này có các đặc tính có thể đo lường bằng số. Tuy nhiên, hành vi của các biến kinh tế cũng có thể phụ thuộc vào các nhân tố định tính như giới tính, trình độ học vấn, mùa, công cộng hay cá nhân v.v...

Chúng ta bắt đầu với việc xem xét trường hợp đơn giản nhất trong đó một biến định tính chỉ có hai lựa chọn.

Ví dụ: Tìm hiểu xem có sự khác biệt về lương giữa nam và nữ không ta xem xét mô hình đơn giản sau:

$$WAGE_i = \beta_1 + \beta_2 D_i + u_i$$

$$WAGE_i = \hat{\beta}_1 + \hat{\beta}_2 D_i + \hat{u}_i \quad (*)$$

Trong đó: $WAGE_i$: Lương của người i.

D_i Giới tính: $D_i = 1$, nếu người i là nam

$D_i = 0$, nếu người i là nữ

Lấy kỳ vọng có điều kiện 2 vế của (*) ta được :

$$E(WAGE_i/D_i=1) = \beta_1 + \beta_2 : \quad \text{Lương trung bình của nam}$$

$$E(WAGE_i/D_i=0) = \beta_1 \quad \text{Lương trung bình của nữ}$$

Vậy sự khác biệt về lương trung bình của nam và nữ là β_2 .

3.9.1. Các dạng mô hình:

Mô hình tổng quát:

$$(PRF) : WAGE_i = \beta_1 + \beta_2 EXPER_i + \gamma_1 D_i + \gamma_2 EXPER_i \cdot D_i + u_i$$

Các phương trình:

$$\text{Nam: } WAGE_i = (\beta_1 + \gamma_1) + (\beta_2 + \gamma_2) EXPER_i + u_i$$

$$\text{Nữ: } WAGE_i = \beta_1 + \beta_2 EXPER_i + u_i$$

$$(SRF) \quad WAGE_i = \hat{\beta}_1 + \hat{\beta}_2 EXPER_i + \hat{\gamma}_1 D_i + \hat{\gamma}_2 EXPER_i \cdot D_i + \hat{u}_i$$

Các kiểm định liên quan đến biến định tính:

🚩 Kiểm định 1:

$H_0: \gamma_1 = 0$ Không có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do các yếu tố ngoài mô hình.

$H_1: \gamma_1 \neq 0$ Có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do các yếu tố ngoài mô hình.

🚩 Kiểm định 2:

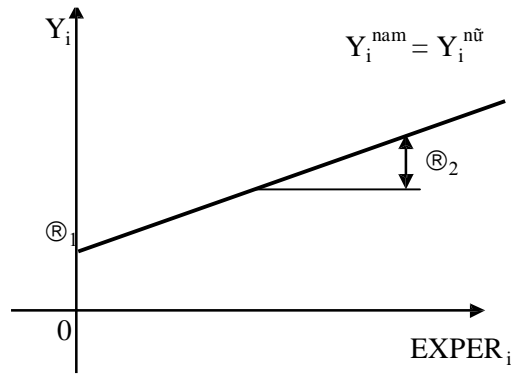
$H_0: \gamma_2 = 0$ Không có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do số năm kinh nghiệm gây nên.

$H_1: \gamma_2 \neq 0$ Có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do số năm kinh nghiệm gây nên

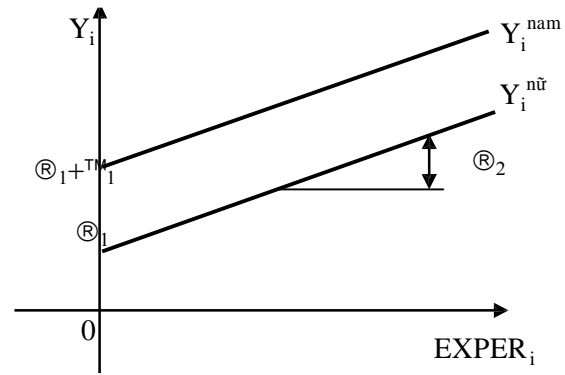
Các dạng mô hình:

- Dạng 1: $\gamma_1 = 0$ và $\gamma_2 = 0$: Không có sự khác biệt về lương giữa nam và nữ.
- Dạng 2: $\gamma_1 \neq 0$ và $\gamma_2 = 0$: Có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do các yếu tố ngoài mô hình
- Dạng 3: $\gamma_1 = 0$ và $\gamma_2 \neq 0$: Có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do số năm kinh nghiệm gây nên
- Dạng 4: $\gamma_1 \neq 0$ và $\gamma_2 \neq 0$: Có sự khác biệt về lương giữa nam và nữ, sự khác biệt này là do số năm kinh nghiệm và các yếu tố ngoài mô hình gây nên.

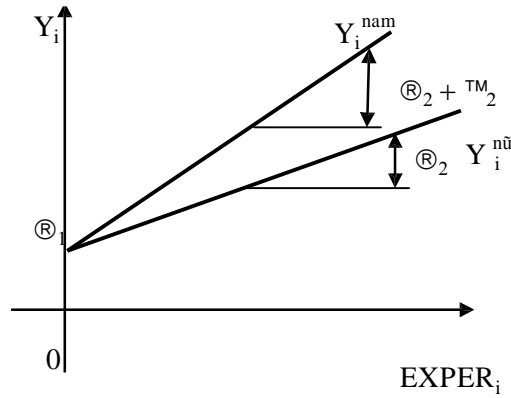
Minh họa:



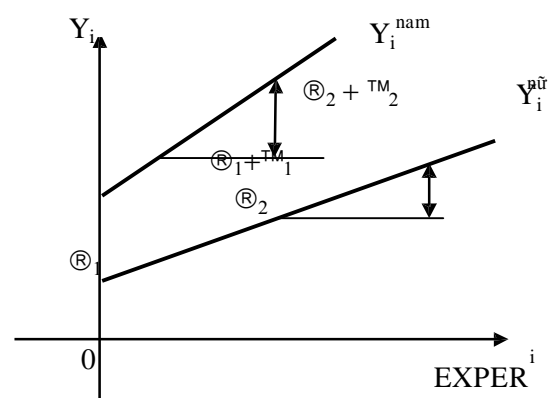
Dạng 1: $TM_1 = 0$ và $TM_2 = 0$



Dạng 2: $TM_1 = \bar{0}$ và $TM_2 = 0$



Dạng 3: $TM_1 = 0$ và $TM_2 = \bar{0}$



Dạng 4: $TM_1 = \bar{0}$ và $TM_2 = \bar{0}$

📌 Ví dụ:

Dependent Variable: WAGE				
Method: Least Squares				
Date: 01/09/06 Time: 17:13				
Sample: 1 49				
Included observations: 49				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1178.287	222.2657	5.301252	0.0000
EXPER	44.23393	24.27003	1.822574	0.0750
GENDER	822.2153	296.6991	2.771209	0.0081
EXPER*GENDER	-35.45679	29.24061	-1.212587	0.2316
R-squared	0.255122	Mean dependent var	1820.204	
Adjusted R-squared	0.205464	S.D. dependent var	648.2687	
S.E. of regression	577.8457	Akaike info criterion	15.63460	
Sum squared resid	15025756	Schwarz criterion	15.78903	
Log likelihood	-379.0477	F-statistic	5.137534	
Durbin-Watson stat	1.727763	Prob(F-statistic)	0.003853	

3.9.2. Biến giả nhiều thuộc tính:

Số các lựa chọn có thể có của một biến định tính có thể nhiều hơn hai. Ví dụ, đặt Y doanh số của một công ty, X là chi phí cho quảng cáo. Chúng ta kỳ vọng quan hệ giữa tiền tiết kiệm và thu nhập sẽ khác nhau theo các mùa khác nhau:

Với một đặc trưng có m thuộc tính ta sẽ đặt $m - 1$ biến giả.

Ví dụ: Biến mùa vụ: 4 mùa: xuân, hạ, thu, đông \square 3 biến giả.

$$D_1 = \begin{cases} 1, & \text{nếu là mùa xuân} \\ 0, & \text{nếu không là mùa xuân} \end{cases}$$

$$D_2 = \begin{cases} 1, & \text{nếu là mùa hạ} \\ 0, & \text{nếu không là mùa hạ} \end{cases}$$

$$D_3 = \begin{cases} 1, & \text{nếu là mùa thu} \\ 0, & \text{nếu không là mùa thu} \end{cases}$$

Y_i : Doanh số,

X_i : Chi phí quảng cáo

Vậy phương trình hồi quy tổng quát:

$$Y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + \beta_4 D_{3i} + \beta_5 X_i + \beta_6 X_i D_{1i} + \beta_7 X_i D_{2i} + \beta_8 X_i D_{3i} + u_i$$

Vậy phương trình hồi quy cho từng mùa:

Mùa xuân: $Y_i = (\beta_1 + \beta_2) + (\beta_5 + \beta_6)X_i + u_i$

Mùa hạ: $Y_i = (\beta_1 + \beta_3) + (\beta_5 + \beta_7)X_i + u_i$

Mùa thu: $Y_i = (\beta_1 + \beta_4) + (\beta_5 + \beta_8)X_i + u_i$

Mùa đông: $Y_i = \beta_1 + \beta_5 X_i + u_i$

Chương IV

ĐA CỘNG TUYẾN (Multicollinearity)

4.1. GIỚI THIỆU:

Một trong những giả thiết của mô hình hồi qui tuyến tính đa biến là không có hiện tượng đa cộng tuyến hoàn hảo. Vậy hiện tượng đa cộng tuyến, đa cộng tuyến hoàn hảo là gì, nó sẽ ảnh hưởng như thế nào đến mô hình và cách khắc phục hiện tượng này sẽ được trình bày ở chương này.

- Hiện tượng đa cộng tuyến là hiện tượng các biến độc lập trong mô hình phụ thuộc lẫn nhau và có quan hệ gần như tuyến tính.
- Hiện tượng đa cộng tuyến hoàn hảo là hiện tượng các biến độc lập có mối quan hệ tuyến tính chính xác (exact linear relationship) được thể hiện được dưới dạng hàm số tuyến tính.

$$\beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} = 0$$

Ví dụ : Ước lượng hàm tiêu dùng. Y : Tiêu dùng, X₂ : Thu nhập và X₃ : của cải

X ₂	X ₃	Y
10	50	52
15	75	75
18	90	97
24	120	129

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

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

Theo dữ liệu ta thấy : X₂ và X₃ có mối quan hệ tuyến tính chính xác

$$X_{3i} - 5X_{2i} = 0$$

$$\beta_3 X_{3i} = 5\beta_2 X_{2i}$$

Vậy $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 5X_{2i} + \hat{u}_i$

$$Y_i = \hat{\beta}_1 + (\hat{\beta}_2 + 5\hat{\beta}_3)X_{2i} + \hat{u}_i$$

Chúng ta chỉ có thể ước lượng $(\hat{\beta}_2 + 5\hat{\beta}_3)$ chứ không ước lượng riêng từng tham số hồi qui $\hat{\beta}_2$ và $\hat{\beta}_3$ được.

Đa cộng tuyến hoàn hảo thường rất ít khi xảy ra trong thực tế. Còn đa cộng tuyến không hoàn hảo thường hay xảy ra trong thực tế (Near collinearity)

4.2. NGUỒN GỐC CỦA ĐA CỘNG TUYẾN (Multicollinearity)

4.2.1. Do phương pháp thu thập dữ liệu

Các giá trị của các biến độc lập phụ thuộc lẫn nhau trong mẫu, nhưng không phụ thuộc lẫn nhau trong tổng thể

Ví dụ: Người có thu nhập cao hơn khuynh hướng sẽ có nhiều của cải hơn. Điều này có thể đúng với mẫu mà không đúng với tổng thể. Trong tổng thể sẽ có các quan sát về các cá nhân có thu nhập cao nhưng không có nhiều của cải và ngược lại.

4.2.2. Dạng hàm mô hình:

Ví dụ các dạng hàm sau dễ xảy ra hiện tượng đa cộng tuyến:

- Hồi qui dạng các biến độc lập được bình phương sẽ xảy ra đa cộng tuyến và đặc biệt khi phạm vi giá trị ban đầu của biến độc lập là nhỏ.
- Các biến độc lập vĩ mô được quan sát theo dữ liệu chuỗi thời gian
- Nhập khẩu quốc gia phụ thuộc vào GDP và CPI (các chỉ số này được thu thập từ dữ liệu thời gian). Giải thích đa cộng tuyến theo ý nghĩa vĩ mô?

4.3. HẬU QUẢ :

4.3.1. Đa cộng tuyến hoàn hảo

Chúng ta không thể ước lượng được mô hình. Các phần mềm máy tính sẽ báo các tín hiệu sau:

- “Matrix singular”: ma trận khác thường mà máy tính không thể thực hiện được khi ước lượng các hệ số hồi qui
- “Exact collinearity encountered”.

4.3.2. Hệ quả khi có đa cộng tuyến không hoàn hảo

(1) Ước lượng OLS vẫn BLUE

- Ước lượng không chệch: trung bình các ước lượng từ mẫu lặp lại sẽ hội tụ đến giá trị ước lượng của tổng thể.
- Phương sai của hệ số ước lượng vẫn đạt minimum nhưng không có nghĩa nhất thiết là nhỏ so với giá trị của ước lượng

(2) Sai số chuẩn của các hệ số sẽ lớn.

- Khoảng tin cậy lớn và thống kê t ít ý nghĩa.
- Các ước lượng không thật chính xác.

- Do đó chúng ta dễ đi đến không có cơ sở bác bỏ giả thiết “không” và điều này có thể không đúng.

(3) R^2 rất cao cho dù thống kê t ít ý nghĩa

- Tại sao hệ số xác định lại cao? Do không có nhiều những biến đổi khác biệt giữa các biến số độc lập vì chúng thực sự có mối quan hệ với nhau
- Dễ dàng bác bỏ giả thuyết “không” của thống kê F và cho rằng mô hình ước lượng có giá trị

(4) Các ước lượng và sai số chuẩn của ước lượng rất nhạy cảm với sự thay đổi của dữ liệu

- Chỉ cần một sự thay đổi nhỏ trong mẫu dữ liệu sẽ kéo theo sự thay đổi lớn các hệ số ước lượng.
- Bởi vì các hệ số ước lượng chứa đựng những mối quan hệ mạnh giữa các biến độc lập

Ví dụ : Xem kết quả ước lượng hàm tiêu dùng:

$$Y = 24.77 + 0.94X_{2i} - 0.04X_{3i} + u$$

$$t \quad (3.67) \quad (1.14) \quad (-0.53)$$

$$R^2=0.96, \quad F = 92.40$$

Y : Tiêu dùng, X_2 : Thu nhập và X_3 : của cải

Ta thấy:

- R^2 rất cao giải thích 96% biến đổi của hàm tiêu dùng,
- Không có biến độc lập nào có ý nghĩa (thống kê t quá thấp).
- Có một biến sai dấu.
- Giá trị thống kê F rất cao dẫn đến bác bỏ giả thuyết “không” và cho rằng mô hình ước lượng có ý nghĩa.
- Biến thu nhập và của cải tương quan rất mạnh với nhau do đó không ước lượng được tác động biên chính xác cho thu nhập hoặc của cải lên tiêu dùng.

Thực hiện các hồi quy sau:

- Thực hiện hồi quy X_3 theo X_2

$$X_3 = 7.54 + 10.19X_2 + u$$

$$(0.26) \quad (62.04) \quad R^2 = .99$$

Hầu như chúng ta có đa cộng tuyến hoàn hảo giữa X_2 và X_3

- Hồi quy tiêu dùng theo thu nhập:

$$Y = 24.45 + 0.51X_2 + u$$

$$(3.81) \quad (14.24) \quad R^2 = 0.96$$

Biến thu nhập trở nên có ý nghĩa thống kê, nhưng trước lúc đó trong mô hình đầu thì không có ý nghĩa.

- Tương tự hồi qui thu nhập Y theo của cải:

$$Y = 24.41 + 0.05X_3 + u$$

$$t \quad (3.55) \quad (13.29) \quad R^2 = 0.96$$

Biến của cải trở nên có ý nghĩa thống kê, nhưng trước lúc đó trong mô hình đầu thì không có ý nghĩa.

4.4. NHÂN DẠNG :

(1) **R² cao và thống kê t thấp.**

(2) **Tương quan tuyến tính mạnh giữa các biến độc lập**

- Xây dựng ma trận hệ số tương quan cặp và quan sát để nhận diện độ mạnh của các tương quan từng cặp biến số độc lập
- Xét về ý nghĩa kinh tế các biến có khả năng tương quan cao

(3) **Thực hiện hồi qui phụ**

- Hồi qui giữa một biến độc lập với tất cả các biến độc lập với nhau và quan sát hệ số R² của các hồi qui phụ
- Thực hiện tính thống kê F

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

k số biến độc lập trong hồi qui phụ

Nếu $F > F^*$ thì chúng ta có thể kết luận rằng R² khác không theo ý nghĩa thống kê và điều này có nghĩa là có đa cộng tuyến trong mô hình.

(4) **Thừa số tăng phương sai (Variance inflation factor-VIF)**

$$VIF = \frac{1}{1 - r_{ij}^2}$$

r_{ij} là hệ số tương quan giữa hai biến độc lập trong mô hình.

Khi r_{ij} tăng làm VIF tăng và làm tăng mức độ đa cộng tuyến

Nguyên tắc kinh nghiệm (Rule of thumb) $VIF \leq 10$ □ Có hiện tượng đa cộng tuyến giữa hai biến độc lập trong mô hình

4.5. CÁC GIẢI PHÁP KHẮC PHỤC ĐA CỘNG TUYẾN

- (1) Bỏ qua đa cộng tuyến nếu $t > 2$
- (2) Bỏ qua đa cộng tuyến nếu R² của mô hình cao hơn R² của mô hình hồi qui phụ.
- (3) Bỏ qua đa cộng tuyến nếu mục tiêu xây dựng mô hình sử dụng để dự báo chứ không phải kiểm định.

(4) Bỏ bớt biến độc lập.

Ví dụ: bỏ biến của cải ra khỏi mô hình hàm tiêu dùng.

Điều này xảy ra với giả định rằng không có mối quan hệ giữa biến phụ thuộc và biến độc lập loại bỏ mô hình.

Nếu lý thuyết khẳng định có mối quan hệ với biến dự định loại bỏ thì việc loại bỏ này sẽ dẫn đến loại bỏ biến quan trọng và chúng ta mắc sai lầm về nhận dạng mô hình (Specification Error).

(5) Bổ sung dữ liệu hoặc tìm dữ liệu mới

Tìm mẫu dữ liệu khác hoặc gia tăng cỡ mẫu. Nếu mẫu lớn hơn mà vẫn còn đa cộng tuyến thì vẫn có giá trị vì mẫu lớn hơn sẽ làm cho phương sai nhỏ hơn và hệ số ước lượng chính xác hơn so với mẫu nhỏ.

(6) Thay đổi dạng mô hình

Mô hình kinh tế lượng có nhiều dạng hàm khác nhau. Thay đổi dạng mô hình cũng có nghĩa là tái cấu trúc mô hình

(7) Sử dụng thông tin hậu nghiệm “priori information”

Sử dụng kết quả của các mô hình kinh tế lượng trước ít có đa cộng tuyến

Ví dụ: Ta có thể biết tác động biên của của cải lên tiêu dùng chỉ bằng 1/10 so với tác động biên của của cải lên tiêu dùng $\beta_3 = 0.10 * \beta_2$

Chạy mô hình với điều kiện tiên nghiệm.

$$Y_i = \beta_1 + \beta_2 X_{2i} + 0.10 * \beta_2 X_{3i} + u_i$$

$$Y_i = \beta_1 + \beta_2 X_i + u_i \text{ trong đó } X_i = X_{2i} + 0.1 X_{3i}$$

Khi ước lượng được β_2 thì suy ra β_3 từ mối quan hệ tiên nghiệm

trên. (8) Sử dụng sai phân cho các biến của mô hình

Sai phân làm cho vấn đề đa cộng tuyến có thể nhẹ đi

Quay trở lại ví dụ hàm tiêu dùng: Thu nhập và của cải có mối quan hệ khá chặt chẽ và do đó không tránh khỏi đa cộng tuyến

Chúng ta muốn ước lượng:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

Ứng với t-1

$$Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + u_{t-1}$$

Lấy sai phân các biến theo thời gian

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2t-1}) + \beta_3 (X_{3t} - X_{3t-1}) + v_t$$

Điều này có thể giải quyết vấn đề đa cộng tuyến vì đa cộng tuyến xảy ra từ bản thân các biến độc lập chứ không xảy ra từ sai phân các biến này.

Tuy nhiên có thể vi phạm giả định chuẩn về sai số ngẫu nhiên.

(9) Kết hợp dữ liệu chéo và dữ liệu chuỗi thời gian

Ví dụ: Nghiên cứu cầu xe hơi và chỉ có dữ liệu chuỗi thời gian.

$$\ln Y_t = \beta_1 + \beta_2 \ln PRICE_t + \beta_3 \ln INCOME_t + u_t$$

Trong đó : Y_t số xe hơi bán ra trong thời đoạn t.

Thông thường giá và thu nhập tương quan mạnh với nhau theo thời gian nên chắc chắn mô hình có đa cộng tuyến khi sử dụng chuỗi thời gian

Giả sử chúng ta có dữ liệu chéo, chúng ta có thể ước lượng độ co giãn theo thu nhập khi sử dụng dữ liệu chéo. Còn độ co giãn theo giá chúng ta phải tìm từ chuỗi dữ liệu theo thời gian

Ước lượng hàm hồi qui theo thời gian

$$Y_t = \beta_1 + \beta_2 \ln P_t + u_t$$

Khi đó $Y_t = \ln Y_t - \beta_3 \ln INCOME_t$

Y đại diện cho số xe hơi bán ra sau khi loại trừ tác động của thu nhập

Căn cứ vào β_3 cho trước chúng ta ước lượng được độ co giãn cầu xe hơi theo giá nhưng không có hiện tượng Đa cộng tuyến

Tuy nhiên chúng ta phải giả định rằng, độ co giãn từ chuỗi thời gian và từ dữ liệu chéo là đồng nhất.

Chương V

DẠNG HÀM

Giả sử bạn có một mô hình kinh tế tiên đoán mối quan hệ giữa một biến phụ thuộc Y và các biến độc lập X. Trong nhiều trường hợp, mô hình này sẽ không cho bạn biết dạng hàm mà mối quan hệ này có trong dữ liệu, mặc dù mô hình này sẽ thường cho bạn một số ý niệm về dạng có thể có của mối quan hệ. Giải pháp thông thường là quyết định xem dạng hàm nào có khả năng mô tả tốt dữ liệu nhất, điều này phụ thuộc vào suy luận kinh tế hoặc phụ thuộc vào việc khảo sát dữ liệu. Sau đó, chúng ta thử xây dựng một số dạng hàm khác nhau và xem chúng có cho ra các kết quả tương tự hay không, và nếu không, thì phải xem dạng hàm nào cho ra các kết quả hợp lý nhất.

Chương này sẽ trình bày một số dạng hàm được sử dụng phổ biến nhất, cho biết chúng biểu hiện như thế nào, mô tả các tính chất của chúng, và cho bạn một số ý tưởng về cách chọn lựa giữa các dạng hàm này.

5.1. HÀM TUYẾN TÍNH:

Dạng hàm tổng quát:

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

Ý nghĩa: $\beta_k = \frac{\partial Y_i}{\partial X_{ki}}$: Tác động biên riêng phần của X_{ki} lên Y

Giữ các yếu tố khác cố định, khi X_k tăng lên một đơn vị thì Y tăng lên β_k đơn vị, và điều này đúng bất kể các giá trị của X và Y là bao nhiêu.

Đây là dạng hàm đơn giản nhất, tuy nhiên, do tính đơn giản này nên khả năng mô tả phù hợp dữ liệu của dạng hàm này thường hạn chế.

Ví dụ: Đường biểu diễn chi phí có dạng $C_i = \beta_1 + \beta_2 Q_i + u_i$ ám chỉ là khi Q tăng thêm một đơn vị thì chi phí C tăng thêm β_2 đơn vị. Điều này chỉ có thể đúng trong trường hợp chi phí biên không đổi; nó không thể đúng trong trường hợp chi phí biên tăng dần (hay giảm dần). Nếu bạn nghĩ rằng chi phí biên tăng dần, bạn sẽ không muốn sử dụng dạng hàm tuyến tính.

5.2. HÀM ĐA THỨC.

Dạng hàm này cho phép giải thích tác động của X lên Y phụ thuộc vào giá trị hiện hành của X.

Dạng hàm tổng quát:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_k X_i^k + u_i$$

$+u_i$

Tác dụng:

🚩 Ước lượng chi phí biên (Tác động biên của Q lên C): $MC = \frac{\partial C_i}{\partial Q_i} = \alpha_1 + 2\alpha_2 Q_i$

Nghĩa là: Chi phí biên phụ thuộc vào Q. Tại điểm Q nào đó, khi Q tăng lên một đơn vị thì C tăng lên $\alpha_1 + 2\alpha_2 Q_i$ đơn vị.

Tác động biên gồm 02 thành phần :

- Thành phần cố định theo Q: α_1
- Thành phần thay đổi theo Q : $2\alpha_2 Q_i$

Lý thuyết gợi ý rằng ta thường có MC tăng dần hoặc không đổi, do đó ta có thể thực hiện kiểm định : $\alpha_2 = 0$ (MC tăng dần) hoặc $\alpha_2 = 0$ (MC không đổi)

🚩 Điểm cực trị (cực đại hoặc cực tiểu): $MC = 0 \Rightarrow Q = \frac{\alpha_1}{2\alpha_2}$

Ví dụ :

Đường Chi phí Trung bình Dài hạn (LRAC) là một đường hình chữ U thể hiện bằng một hàm bậc hai (đa thức bậc hai) :

Ví dụ: $LRAC_i = \alpha_0 + \alpha_1 Q_i + \alpha_2 Q_i^2 + u_i$

Sử dụng dữ liệu 86 S&Ls cho năm 1975. Sản lượng Q được đo lường như là tổng tài sản cố. LRAC được đo lường như là chi phí hoạt động trung bình tính theo % của tổng tài sản cố.

SRF : $LRAC = 2,38 - 0,615Q + 0,054 Q^2 + u_i$

Sản lượng cho LRAC tối thiểu của hàm này khi tổng tài sản cố Q đạt 569 tỷ đô la :

$$\frac{\partial LRAC_i}{\partial Q_i} = -0.615 + 2(0.054) Q = 0 \Rightarrow Q = 569$$

5.3. HÀM LOG KÉP:

Khảo sát hàm sản xuất Cobb-Douglas: $Y_i = \alpha_1 K_i^{\alpha_2} L_i^{\alpha_3} e^{u_i}$

- Trong đó: Y = sản lượng
 K = nhập lượng vốn
 L = nhập lượng lao động

Đây là mối quan hệ phi tuyến, nhưng chúng ta có thể biến đổi quan hệ này như sau:

$$\ln Y_i = \ln \alpha_1 + \alpha_2 \ln K_i + \alpha_3 \ln L_i + u_i$$

Đây là mô hình tuyến tính trong các tham số nhưng không tuyến tính trong các biến số. Mô hình này tuyến tính theo lôgarít của các biến số. Mô hình này được gọi là mô hình lôgarít-lôgarít, lôgarít kép hay tuyến tính lôgarít

Dạng tổng quát:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \dots + \beta_k \ln X_{ki} + u_i$$

Tác dụng:

- Tham số độ dốc của một mô hình log kép đo lường độ co giãn riêng phần của Y theo X.

$$\beta_2 = \frac{\frac{\Delta \ln Y}{\ln Y}}{\frac{\Delta \ln X_k}{\ln X_k}} = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta X_k}{X_k}} : \text{ là độ co giãn riêng phần của Y theo } X_k.$$

Trong hàm Cobb-Douglas: $\beta_2 = \frac{\Delta Y}{\Delta K} \cdot \frac{K}{Y}$: là độ co giãn riêng phần của sản lượng theo vốn. Nghĩa là, giữ lao động không đổi, khi vốn tăng 1% thì sản lượng sẽ tăng $\beta_2\%$.

- Tác động biên thay đổi: $\frac{\Delta Y}{\Delta X_k} = \beta_2 \frac{Y}{X_k}$

- Trong hàm Cobb – Douglas $\beta_2 + \beta_3$ đo lường hiệu quả theo qui mô. Đáp ứng của sản lượng đối với thay đổi tương xứng trong các nhập lượng.

- o Nếu $\beta_2 + \beta_3 = 1$: hiệu quả không đổi. Tăng gấp đôi nhập lượng thì sản lượng sẽ tăng gấp đôi.
- o Nếu $\beta_2 + \beta_3 < 1$: hiệu quả giảm dần
- o Nếu $\beta_2 + \beta_3 > 1$: hiệu quả tăng dần

Ví dụ 1: Dữ liệu về nông nghiệp Đài Loan 1957-72:

$$\ln Y = -3.34 + 0.49 \ln K + 1.50 \ln L + u_i$$

$$t \quad (-1.36) \quad (4.80) \quad (0.54)$$

$$R^2 = 0.89$$

Y GNP tính bằng triệu đô la

K là vốn thực tính bằng triệu đô la

L tính bằng triệu ngày công lao động

- Độ co giãn của sản lượng theo vốn là 0,49 : Giữ nhập lượng lao động không đổi, gia tăng 1% nhập lượng vốn dẫn đến gia tăng 0,49% sản lượng.
- Độ co giãn của sản lượng theo lao động là 1,50 : Giữ nhập lượng vốn không đổi, gia tăng 1% nhập lượng lao động dẫn đến gia tăng 1,5% sản lượng.
- R^2 có nghĩa là 89% biến thiên trong lôgarít của sản lượng được giải thích bởi lôgarít của lao động và vốn.
- Hiệu quả tăng theo qui mô bởi vì:

$$\hat{\beta}_2 + \hat{\beta}_3 = 1,99$$

Thực hiện kiểm định : $\beta_2 + \beta_3 = 0$ (Kiểm định Wald) ta có :

Ví dụ 2: Ta có thể lập hàm cầu như một hàm log kép :

$$\ln Q_i = \ln \beta_1 + \beta_2 \ln P_{\text{coffee}} + \beta_3 \ln P_{\text{tea}} + u_i$$

$$\ln Q = -3.34 + 0.49 \ln P_{\text{coffee}} + 1.50 \ln P_{\text{tea}} + \hat{u}_i$$

Q : là mức tiêu dung cà phê mỗi ngày

P_{coffee} : là giá cà phê mỗi cân Anh

P_{tea} : là giá trà mỗi cân Anh

Kết quả:

$$\ln Q = 0.78 - 0.25 \ln P_{\text{coffee}} + 0.38 \ln P_{\text{tea}} + \hat{u}_i$$

t (51.1) (-5.12) (3.25)

- Độ co giãn theo giá riêng là - 0,25: Giữ các yếu tố khác không đổi, nếu giá gia tăng 1% thì lượng cầu sẽ giảm 0,25%.

Không co giãn - giá trị tuyệt đối nhỏ hơn 1.

- Độ co giãn theo gia-chéo là 0,38. Giữ các yếu tố khác không đổi, nếu giá trà gia tăng 1%, thì lượng cầu cà phê sẽ gia tăng 0,38%

Nếu độ co giãn theo giá - chéo dương, thì cà phê và trà là các sản phẩm thay thế.

Nếu độ co giãn theo gia-chéo âm, thì đó là các sản phẩm bổ trợ.

Hàm Cobb-Douglass Tổng quát

Khảo sát hàm sản xuất Cobb-Douglas: $Y_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3} e^{rt} e^{u_i}$

Trong đó: Y = sản lượng

K = nhập lượng vốn

L = nhập lượng lao động

t = thời đoạn - xu hướng thời gian (năm)

Ta có thể biến đổi quan hệ này như sau:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln K_i + \beta_3 \ln L_i + rt + u_i$$

$$r = \frac{\partial \ln Y}{\partial t} = \frac{\partial Y/Y}{\partial t} = H \frac{\partial Y/Y}{\partial t} : \text{tốc độ thay đổi tương đối hàng năm của } Y.$$

5.4. HÀM BÁN LOG

Sử dụng mô hình này khi chúng ta quan tâm đến tốc độ tăng trưởng của biến nào đó như mối quan hệ giữa tốc độ tăng thu nhập theo sự thay đổi tuyệt đối của số năm học hoặc số năm kinh nghiệm

Ví dụ: Giả sử chúng ta có 100.000.000 VNĐ và chúng ta gởi số tiền này vào ngân hàng với lãi suất $r = 8\%$ / năm.

Sau một năm, số tiền này sẽ tăng lên đến: $Y_1 = 100.000.000(1 + 0.08) = 108.000.000$

Sau hai năm, số tiền này sẽ tăng lên đến: $Y_2 = 108.000.000(1 + 0.08) =$
 $100.000.000(1 + 0.08)^2 = 116.064.000$

Vậy sau t năm, số tiền này sẽ tăng lên $Y_t = 100.000.000(1 + 0.08)^t$

Vậy: công thức lãi kép: $Y_t = Y_0 (1 + r)^t$

Y_0 là giá trị ban đầu của Y

Y_t là giá trị của Y vào thời điểm t

r là tỷ lệ tăng trưởng kép của Y

Lấy lôgarít cơ số e của công thức lãi kép :

$$Y_t = Y_0 (1 + r)^t$$

$$\ln Y_t = \ln Y_0 + t \ln(1 + r)$$

Đặt $\beta_1 = \ln Y_0$, $\beta_2 = \ln(1 + r)$ và viết lại và thêm số hạng sai số vào :

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

Dạng hàm tổng quát:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

$$\beta_2 = \frac{\partial \ln Y}{\partial X} = \frac{\partial Y / Y}{\partial X} \text{ H } \frac{\partial Y / Y}{\partial X} :$$

Hệ số độ dốc đo lường thay đổi tương đối của Y đối với sự thay đổi tuyệt đối cho trước trong giá trị của biến giải thích

Ví dụ:

$$\ln(\text{GDP thực}) = 6,9636 + 0,0269t$$

$$\text{SE} \quad (0,0151) \quad (0,0017)$$

$$R^2 = 0,95$$

- Hệ số độ dốc đo lường tốc độ tăng trưởng: GDP thực tăng trưởng với tốc độ 0,0269 mỗi năm, hay 2,69 phần trăm mỗi năm.
- Lấy đối lôgarít cơ số e của 6,9636 để chỉ ra rằng vào đầu năm 1969, GDP thực ước lượng vào khoảng 1057 tỷ đô la, nghĩa là ở t = 0
- Tính tốc độ tăng trưởng kép r

$$\beta_2 = \ln(1 + r) \quad \square \quad e^{\beta_2} = (1 + r)$$

$$\square \quad r = e^{\beta_2} - 1 = 1,0273 - 1 = 0,0273$$

5.5. CÁC DẠNG HÀM KHÁC:

Chương VI

HIỆN TƯỢNG PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI (Heteroscedasticity)

6.1. GIỚI THIỆU:

Một trong những giả thiết quan trọng của mô hình hồi quy tuyến tính ước lượng theo phương pháp OLS là các số hạng sai số u_i có phân phối giống nhau với trị trung bình bằng không và phương sai không đổi là σ^2 . Điều này có nghĩa là mức độ phân tán của giá trị biến phụ thuộc quan sát được (Y) xung quanh đường hồi qui như nhau cho tất cả các quan sát.

Tuy nhiên, trong nhiều trường hợp thông thường có liên quan đến những dữ liệu chéo, giả thuyết này có thể sai. Hiện tượng như vậy được gọi là phương sai của sai số thay đổi (Heteroscedasticity - HET).

Vậy: $\text{Var}(u_i) = \sigma^2 \rightarrow$ Phương sai của sai số không đổi
 $\text{Var}(u_i) = \sigma_i^2 \neq \sigma^2 \rightarrow$ Phương sai của sai số thay đổi

6.2. HẬU QUẢ:

6.2.1. Tác động lên tính chất của các ước lượng OLS:

Nếu mô hình có hiện tượng phương sai của sai số thay đổi (HET), thì các tính chất: không chệch và nhất quán không bị vi phạm nếu ta sử dụng OLS để ước lượng các hệ số hồi quy.

Nghĩa là: $E(\hat{\beta}_k) = \beta_k$

6.2.2. Tác Động Lên Các Kiểm Định Giả Thuyết

Ta biết phương sai của các ước lượng phụ thuộc vào phương sai của sai số, do nếu mô hình có hiện tượng HET thì phương sai của các tham số ước lượng theo OLS cũng sẽ không còn nhỏ nhất và nhất quán nữa. Điều này sẽ dẫn đến các kiểm định giả thuyết không còn giá trị nữa.

Nghĩa là: $\sigma_i^2 \neq \text{const} \rightarrow S_{\hat{\beta}_k} = f(\sigma_i^2) \rightarrow t_{\hat{\beta}_k}$ không còn ý nghĩa.

6.2.3. Tác Động Lên Việc Dự Báo

Do các ước lượng OLS vẫn không chệch, nên các dự báo dựa trên những giá trị ước lượng này cũng sẽ không thiên lệch.

Nhưng do các ước lượng là không hiệu quả, nên các dự báo cũng sẽ không hiệu quả. Nói cách khác, độ tin cậy của những dự báo này (đo lường bằng phương sai của chúng) sẽ kém.

Tóm lại: Khi mô hình có hiện tượng HET thì mô hình không còn BLUE.

6.3. NGUYÊN NHÂN XẢY RA HIỆN TƯỢNG HET

- Do bản chất của các mối quan hệ kinh tế: Có nhiều mối kinh tế đã chứa đựng hiện tượng này, ví dụ: Thu nhập tăng thì tiết kiệm cũng tăng.
- Do kỹ thuật thu thập dữ liệu.
- Do con người học được hành vi trong quá khứ.

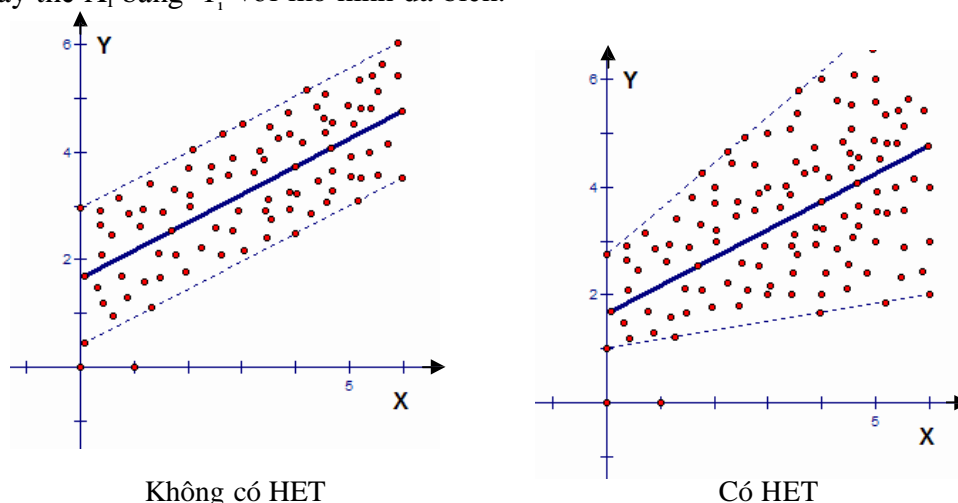
6.4. NHÂN DẠNG HIỆN TƯỢNG HET

6.4.1. Bảng trực giác và kinh nghiệm:

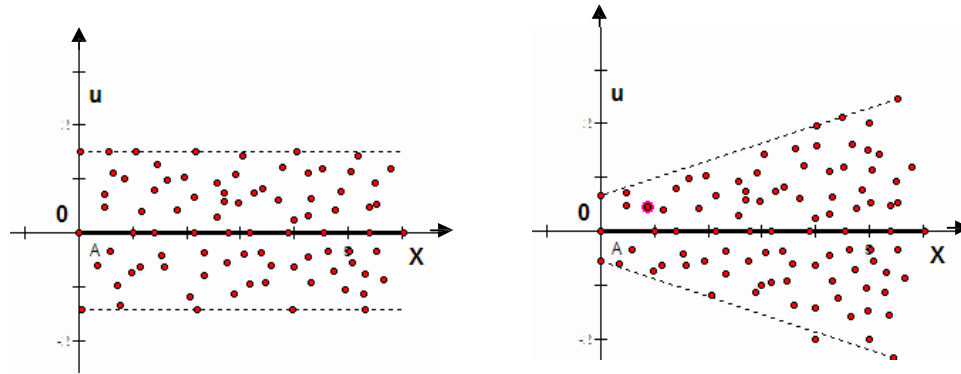
Làm việc thường xuyên với dữ liệu, ta sẽ có một “cảm giác” tốt hơn với dữ liệu, thông thường với dữ liệu chéo (cross-sectional data) khả năng có hiện tượng HET rất cao.

6.4.2. Phân tích bằng đồ thị (Graphical analysis)

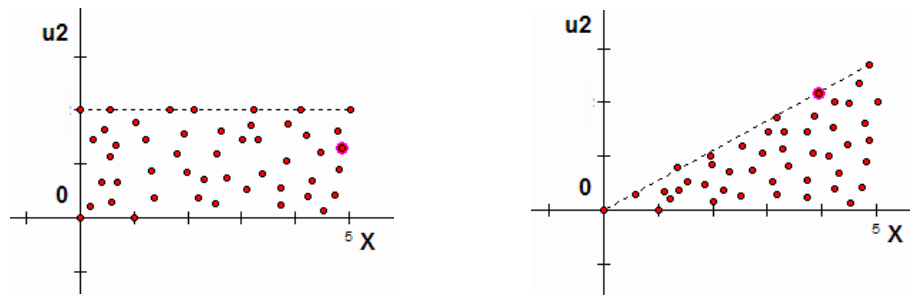
Để phát hiện HET người ta thường dùng các đồ thị phân tán giữa: (X_i, Y_i) ; (X_i, u_i) (X_i, u_i^2) và thay thế X_i bằng \hat{Y}_i với mô hình đa biến.



Hình 6.1: Đồ thị (X_i, Y_i) nhận biết hiện tượng HET



Hình 6.2: Đồ thị $(X_i; u_i)$ nhận biết hiện tượng HET



Không có HET

Có HET

Hình 6.3: Đồ thị $(X_i; u_i^2)$ nhận biết hiện tượng HET

Ví dụ : Data8-2.wf1 của Ramanathan chứa dữ liệu về tổng thu nhập cá nhân và chi tiêu cho đi lại trong nước (1993) đối với 50 tiểu bang và Thủ đô Washington của Mỹ. Các biến trong file này là:

EXPTRAV : Chi tiêu cho đi lại tính bằng tỷ đô la (có giá trị từ 0,708-42,48).

INCOME : Thu nhập cá nhân tính bằng tỷ đô la (có giá trị từ 9,3-683,5)

POP : Dân số tính bằng triệu người (có giá trị từ 0,47-31,217).

Thực hiện mô hình hồi quy đơn giản sau:

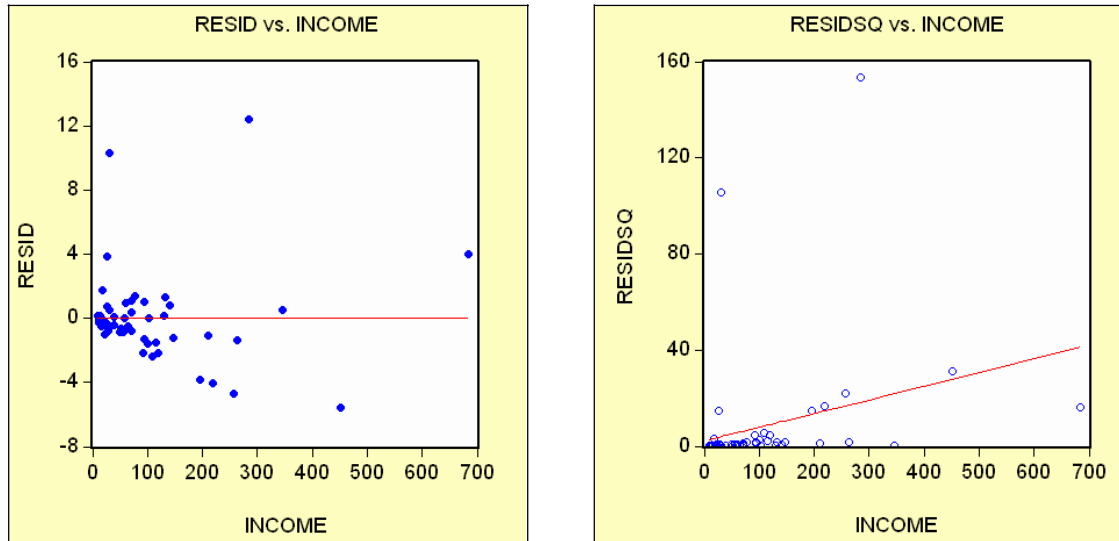
$$EXPTRAV_i = \beta_1 + \beta_2 INCOME_i + u_i$$

Dependent Variable: EXPTRAV				
Method: Least Squares				
Date: 10/27/04 Time: 16:57				
Sample: 1 51				
Included observations: 51				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.498120	0.535515	0.930170	0.3568
INCOME	0.055573	0.003293	16.87558	0.0000
R-squared	0.853199	Mean dependent var	6.340706	
Adjusted R-squared	0.850203	S.D. dependent var	7.538343	
S.E. of regression	2.917611	Akaike info criterion	5.017834	
Sum squared resid	417.1103	Schwarz criterion	5.093591	
Log likelihood	-125.9548	F-statistic	284.7850	
Durbin-Watson stat	2.194928	Prob(F-statistic)	0.000000	

Hình 6.4: Kết quả mô hình hồi quy cơ bản của ví dụ

Vậy: $EXPTRAV_i = 0,49812 + 0,055573 INCOME_i + \hat{u}_i$

Để kiểm tra HET trong mô hình hồi qui này ta sẽ vẽ đồ thị sau:



Đồ thị giữa (X_i, u_i)

Đồ thị giữa (X_i, u_i^2)

Hình 6.5: Đồ thị nhận biết hiện tượng HET

Dựa vào đồ thị nói trên, ta có thể nói có dấu hiệu có hiện tượng HET trong mô hình.

Lưu ý: Bằng trực giác, kinh nghiệm hay đồ thị chỉ cho ta biết dấu hiệu để nhận dạng hiện tượng HET. Để có kết luận chính thức về hiện tượng HET ta phải thực hiện các kiểm định phù hợp.

6.4.3. Kiểm định nhân tử Lagrange (Lagrange Multiplier Test – LM Test):

Phương trình hồi qui tổng thể:

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

Các bước thực hiện:

Bước 1: Thực hiện hồi qui phụ (thay \hat{u}_i bằng u_i) theo một trong những cách sau:

- a) **Glejser:** $\hat{u}_i = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + \epsilon_i$
- b) **Breush-Pagan:** $\hat{u}_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + \epsilon_i$
- c) **God Fray:** $\ln(\hat{u}_i^2) = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + \epsilon_i$
- d) **White:** $\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + \dots + \alpha_p Z_{pi} + \epsilon_i$

Bước 2: Phát biểu giả thiết:

$$H_0: \langle_2 = \langle_3 = \dots = \langle_p = 0 \quad \text{Không có hiện tượng HET.}$$

$$H_1: \text{Có ít nhất 1 số } \langle_j \neq 0 \quad (j = 1, p) \quad \text{Không có hiện tượng HET.}$$

Bước 3: Tính $l_{tt} = nR_{hqp}^2$

Tra bảng $l^* = l^*_{p-1, \langle}$

Nếu: $l_{tt} > l^*$ (hay p-value $> \langle$) \rightarrow Bác bỏ H_0 .

Trong thực hành EVIEWS ta nên thực hiện kiểm định kiểm tra HET bằng phương trình hồi quy phụ White.

Ví dụ: Kiểm tra HET của ví dụ trên bằng kiểm định WHITE với chọn $\langle = 10\%$.

Phương trình hồi quy tổng thể: $EXPTRAV_i = \alpha_1 + \alpha_2 INCOME_i + u_i$

Phương trình hồi quy phụ theo White Test:

$$u_i^2 = \langle_1 + \langle_2 INCOME_t + \langle_3 INCOME^2_t + \langle_t$$

Kết quả mô hình:

White Heteroskedasticity Test:				
F-statistic	2.537633	Probability	0.089614	
Obs*R-squared	4.876820	Probability	0.087300	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 10/27/04 Time: 18:25				
Sample: 1 51				
Included observations: 51				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.689561	5.950486	-0.283937	0.7777
INCOME	0.126986	0.073163	1.735656	0.0890
INCOME^2	-0.000132	0.000127	-1.039498	0.3038
R-squared	0.095624	Mean dependent var	8.178634	
Adjusted R-squared	0.057942	S.D. dependent var	26.00254	
S.E. of regression	25.23798	Akaike info criterion	9.351600	
Sum squared resid	30573.88	Schwarz criterion	9.465237	
Log likelihood	-235.4658	F-statistic	2.537633	
Durbin-Watson stat	2.147294	Prob(F-statistic)	0.089614	

Hình 6.6: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.4

Giả thiết kiểm định: $H_0 : \langle_2 = \langle_3 = 0$
 $H_1 : \langle_2 \neq 0 \text{ hoặc } \langle_3 \neq 0$

Ta có: $l^* = l^*_{p-1, \langle} = l^*_{2, 10\%} = 4.605$

$nR_{hqp}^2 = 4.876820 > l^*$ (hoặc theo p-value = 0.087300 < 10%)

Ⓡ Bác bỏ H_0 : Có hiện tượng phương sai của sai số thay đổi.

6.5. BIÊN PHÁP KHÁC PHỤC

6.5.1. Bình phương tối thiểu tổng quát (hoặc trọng số):

(Generalized Least Squares – GSL or Weighted Least Squares – WLS)

Xét mô hình hồi quy tổng thể:

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

Mô hình có $\text{Var}(u_i) = f_i$ biết trước. Chia các số hạng cho f_i , chúng ta có mô hình hiệu chỉnh như sau:

$$\frac{Y_i}{f_i} = \beta_1 \frac{1}{f_i} + \beta_2 \frac{X_{2i}}{f_i} + \beta_3 \frac{X_{3i}}{f_i} + \dots + \beta_K \frac{X_{Ki}}{f_i} + \frac{u_i}{f_i}$$

Đặt $Y_i^* = \frac{Y_i}{f_i}$; $X_i^* = \frac{1}{f_i}$; $X_{2i}^* = \frac{X_{2i}}{f_i}$; $X_{3i}^* = \frac{X_{3i}}{f_i}$; ... $X_{Ki}^* = \frac{X_{Ki}}{f_i}$; $u_i^* = \frac{u_i}{f_i}$

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + \dots + \beta_K X_{Ki}^* + u_i^* \quad (*)$$

Ta có: $\text{Var}(u_i^*) = \text{Var}\left(\frac{u_i}{f_i}\right) = \frac{\text{Var}(u_i)}{f_i^2} = \frac{f_i}{f_i^2} = 1$.

Mô hình (*) không có số hạng sai số không đổi nên sẽ có tính BLUE.

Thủ tục GLS được áp dụng cho trường hợp phương sai thay đổi thì cũng giống như thủ tục bình phương tối thiểu có trọng số (WLS).

Đặt trọng số $w_i = \frac{1}{f_i}$ và mô hình (*) có thể được viết lại như sau:

$$w_i Y_i = w_i \beta_1 + \beta_2 w_i X_{2i} + \beta_3 w_i X_{3i} + \dots + \beta_K w_i X_{Ki} + w_i u_i$$

6.5.2. Bình Phương Tối Thiểu Tổng Quát Khả Thi (FGLS)

Một cách tổng quát, cấu trúc của phương sai của sai số thay đổi là không biết (nghĩa là f_i không biết trước), vì vậy GLS khó thực hiện. Để cải thiện điều này, trước tiên ta phải tìm cách ước lượng của f_i bằng một số cách và sau đó sử dụng thủ tục WLS.

Để ước lượng f_i ta có thể dùng các phương trình hồi quy phụ của các tác giả Glejser, Breush-Pagan, God Fray, White

Phương pháp này gọi là Bình Phương Tối Thiểu Tổng Quát Khả Thi (FGLS) hay Bình Phương Tối Thiểu có trọng số WLS.

Ví dụ: Thực hiện khắc phục hiện tượng phương sai thay đổi của ví dụ trên:

Thực hiện hồi quy: $u_t^2 = \langle_1 + \langle_2 \text{INCOME}_t + \langle_3 \text{INCOME}_t^2 + \langle_t$

Dependent Variable: USQ				
Method: Least Squares				
Date: 10/27/04 Time: 18:37				
Sample: 1 51				
Included observations: 51				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.689561	5.950486	-0.283937	0.7777
INCOME	0.126986	0.073163	1.735656	0.0890
INCOME^2	-0.000132	0.000127	-1.039498	0.3038
R-squared	0.095624	Mean dependent var	8.178634	
Adjusted R-squared	0.057942	S.D. dependent var	26.00254	
S.E. of regression	25.23798	Akaike info criterion	9.351600	
Sum squared resid	30573.88	Schwarz criterion	9.465237	
Log likelihood	-235.4658	F-statistic	2.537633	
Durbin-Watson stat	2.147294	Prob(F-statistic)	0.089614	

Hình 6.7: Kiểm định White để nhận dạng hiện tượng HET (cách trực tiếp)

Tạo biến: usqf từ forecast

Genr w=1/@sqrt(abs(usqf))

Thực hiện hồi quy mô hình bằng WLS :

Dependent Variable: EXPTRAV				
Method: Least Squares				
Date: 01/11/06 Time: 14:56				
Sample: 1 51				
Included observations: 51				
Weighting series: W				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.851923	0.420225	2.027302	0.0481
INCOME	0.052208	0.004773	10.93874	0.0000
Weighted Statistics				
R-squared	0.535902	Mean dependent var	4.748416	
Adjusted R-squared	0.526431	S.D. dependent var	3.521999	
S.E. of regression	2.423712	Akaike info criterion	4.646903	
Sum squared resid	287.8446	Schwarz criterion	4.722661	
Log likelihood	-116.4960	F-statistic	119.6560	
Durbin-Watson stat	2.187902	Prob(F-statistic)	0.000000	
Unweighted Statistics				
R-squared	0.850070	Mean dependent var	6.340706	
Adjusted R-squared	0.847010	S.D. dependent var	7.538343	
S.E. of regression	2.948538	Sum squared resid	426.0000	
Durbin-Watson stat	2.172427			

Hình 6.8: Thực hiện hồi quy theo WLS theo trọng số từ hồi quy phụ White

Kiểm tra hiện tượng HET trong mô hình 6.9 ta thấy mô hình này đã không còn HET.

White Heteroskedasticity Test:				
F-statistic	0.006038	Probability	0.993981	
Obs*R-squared	0.012827	Probability	0.993607	
Test Equation:				
Dependent Variable: STD_RESID^2				
Method: Least Squares				
Date: 01/11/06 Time: 14:57				
Sample: 1 51				
Included observations: 51				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.409998	5.199220	1.040540	0.3033
INCOME	0.001794	0.063926	0.028059	0.9777
INCOME^2	1.72E-06	0.000111	0.015523	0.9877
R-squared	0.000252	Mean dependent var	5.644011	
Adjusted R-squared	-0.041405	S.D. dependent var	21.60880	
S.E. of regression	22.05161	Akaike info criterion	9.081671	
Sum squared resid	23341.14	Schwarz criterion	9.195308	
Log likelihood	-228.5826	F-statistic	0.006038	
Durbin-Watson stat	2.080884	Prob(F-statistic)	0.993981	

Hình 6.9: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.8

6.5.3. Phương sai của sai số thay đổi với tỷ số biết trước

Giả sử tính phương sai của sai số thay đổi được tính với thông qua một biến Z_i biết trước như sau:

$$\text{Var}(u_i) = f_i^2 = f^2 Z_i^2. \text{ Vậy: } f_i = f Z_i$$

Nói cách khác, độ lệch chuẩn của sai số tỷ lệ với một số biến Z_i biết trước, hằng số của tỷ lệ này là f .

Ta có:

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

$$\frac{Y_i}{Z_i} = \beta_1 \frac{1}{Z_i} + \beta_2 \frac{X_{2i}}{Z_i} + \beta_3 \frac{X_{3i}}{Z_i} + \dots + \beta_K \frac{X_{Ki}}{Z_i} + \frac{u_i}{Z_i}$$

Đặt $Y_i^* = \frac{Y_i}{Z_i}$; $X_1^* = \frac{1}{Z_i}$; $X_{2i}^* = \frac{X_{2i}}{Z_i}$; $X_{3i}^* = \frac{X_{3i}}{Z_i}$; ... $X_{ki}^* = \frac{X_{ki}}{Z_i}$; $u_i^* = \frac{u_i}{Z_i}$

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \beta_3 X_{3i}^* + \dots + \beta_K X_{Ki}^* + u_i^* \quad (*)$$

Ta có: $\text{Var}(u_i^*) = \text{Var} \left(\frac{u_i}{Z_i} \right) = \frac{\text{Var}(u_i)}{\text{Var}(Z_i)} = \frac{f_i^2}{f^2} = f^2 = \text{const}$

Ví dụ : Phương trình hồi quy tổng thể: $\text{EXPTRAV}_i = \beta_1 + \beta_2 \text{INCOME}_i + u_i$.

Ta ta kỳ vọng f_i và POP_i có quan hệ như sau: $f_i = f \text{POP}_i \Rightarrow f = \frac{f_i}{\text{POP}_i}$

Để khử hiện tượng phương sai thay đổi ta chia cả 02 vế cho POP_i :

$$\frac{\text{EXPTRAV}_i}{\text{POP}_i} = \frac{\beta_1}{\text{POP}_i} + \frac{\beta_2 \text{INCOME}_i}{\text{POP}_i} + f \quad (*)$$

Vậy: $\frac{EXPTRAV_i}{POP_i}$: Tổng chi tiêu bình quân đầu người cho di chuyển
 $\frac{INCOME_i}{POP_i}$: Thu nhập bình quân đầu người (tỷ đôla)

Thực hiện ước lượng (*) là thực hiện hồi quy mô hình cơ bản với trọng số $1/POP_i$:

Dependent Variable: EXPTRAV				
Method: Least Squares				
Date: 01/11/06 Time: 14:53				
Sample: 1 51				
Included observations: 51				
Weighting series: 1/INCOME				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.572408	0.293665	1.949190	0.0570
INCOME	0.058816	0.011046	5.324864	0.0000
Weighted Statistics				
R-squared	0.071958	Mean dependent var	2.761030	
Adjusted R-squared	0.053018	S.D. dependent var	2.110383	
S.E. of regression	2.053677	Akaike info criterion	4.315566	
Sum squared resid	206.6618	Schwarz criterion	4.391324	
Log likelihood	-108.0469	F-statistic	28.35417	
Durbin-Watson stat	2.234715	Prob(F-statistic)	0.000003	
Unweighted Statistics				
R-squared	0.847199	Mean dependent var	6.340706	
Adjusted R-squared	0.844081	S.D. dependent var	7.538343	
S.E. of regression	2.976634	Sum squared resid	434.1571	
Durbin-Watson stat	2.090936			

Hình 6.10: Thực hiện hồi quy theo WLS theo trọng số $1/INCOME$

Kiểm tra HET của mô hình 6.10 kết quả:

White Heteroskedasticity Test:				
F-statistic	0.525198	Probability	0.594797	
Obs*R-squared	1.092147	Probability	0.579220	
Test Equation:				
Dependent Variable: STD_RESID^2				
Method: Least Squares				
Date: 01/11/06 Time: 14:55				
Sample: 1 51				
Included observations: 51				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.903660	4.766589	1.658137	0.1038
INCOME	-0.053372	0.058607	-0.910674	0.3670
INCOME^2	6.65E-05	0.000101	0.655747	0.5151
R-squared	0.021415	Mean dependent var	4.052192	
Adjusted R-squared	-0.019360	S.D. dependent var	20.02379	
S.E. of regression	20.21668	Akaike info criterion	8.907916	
Sum squared resid	19618.29	Schwarz criterion	9.021553	
Log likelihood	-224.1519	F-statistic	0.525198	
Durbin-Watson stat	2.115062	Prob(F-statistic)	0.594797	

Hình 6.11: Kiểm định White để nhận dạng hiện tượng HET của mô hình 6.10

Vậy mô hình không còn hiện tượng HET.

6.5.4. Tái cấu trúc mô hình:

Hiện tượng phương sai thay đổi có thể xảy ra trong trường hợp nhận dạng sai dạng hàm của mô hình, trong trường hợp này ta phải xây dựng lại mô hình bằng một dạng hàm phù hợp.

Ví dụ: Xét mô hình sau: $VA_i = \beta_1 + \beta_2 K_i + \beta_3 L_i + u_i$

Trong đó: VA = sản lượng
K = nhập lượng vốn
L = nhập lượng lao động

Mô hình:

Dependent Variable: VA				
Method: Least Squares				
Date: 01/19/06 Time: 16:02				
Sample: 1 27				
Included observations: 27				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
L	2.338136	1.038966	2.250445	0.0339
K	0.471043	0.112439	4.189327	0.0003
C	114.3376	173.4314	0.659267	0.5160
R-squared	0.959805	Mean dependent var	2340.201	
Adjusted R-squared	0.956455	S.D. dependent var	2251.659	
S.E. of regression	469.8642	Akaike info criterion	15.24720	
Sum squared resid	5298536.	Schwarz criterion	15.39119	
Log likelihood	-202.8372	F-statistic	286.5410	
Durbin-Watson stat	2.060297	Prob(F-statistic)	0.000000	

Kiểm tra HET của mô hình :

White Heteroskedasticity Test:			
F-statistic	5.785285	Probability	0.002446
Obs*R-squared	13.84127	Probability	0.007819
Test Equation:			
Dependent Variable: RESID^2			
Method: Least Squares			
Date: 01/19/06 Time: 16:07			
Sample: 1 27			
Included observations: 27			
Variable	Coefficient	Std. Error	Prob.
C	29191.15	207568.1	0.140634
L	-1491.415	1721.318	0.3956
L^2	0.430014	1.730384	0.8060
K	329.9923	137.7174	0.0255
K^2	-0.014946	0.017986	0.4149
R-squared	0.512640	Mean dependent var	196242.1
Adjusted R-squared	0.424029	S.D. dependent var	345587.3
S.E. of regression	262275.8	Akaike info criterion	27.95776
Sum squared resid	1.51E+12	Schwarz criterion	28.19773
Log likelihood	-372.4297	F-statistic	5.785285
Durbin-Watson stat	1.740627	Prob(F-statistic)	0.002446

Kết quả: Có HET ở mức $\alpha = 10\%$.

Thay đổi dạng hàm thành dạng hàm Cobb-Douglas như sau:

$$VA_i = \alpha_0 K_i^{\alpha_1} L_i^{\alpha_2} e^{u_i}$$

Đây là mối quan hệ phi tuyến, nhưng chúng ta có thể biến đổi quan hệ này như sau:

$$\ln VA_i = \ln \alpha_0 + \alpha_1 \ln K_i + \alpha_2 \ln L_i + u_i$$

Dependent Variable: LOG(VA)				
Method: Least Squares				
Date: 01/19/06 Time: 16:03				
Sample: 1 27				
Included observations: 27				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.170644	0.326782	3.582339	0.0015
LOG(K)	0.375710	0.085346	4.402204	0.0002
LOG(L)	0.602999	0.125954	4.787457	0.0001
R-squared	0.943463	Mean dependent var	7.443631	
Adjusted R-squared	0.938751	S.D. dependent var	0.761153	
S.E. of regression	0.188374	Akaike info criterion	-0.396336	
Sum squared resid	0.851634	Schwarz criterion	-0.252355	
Log likelihood	8.350541	F-statistic	200.2489	
Durbin-Watson stat	1.885989	Prob(F-statistic)	0.000000	

Kiểm tra HET của mô hình :

White Heteroskedasticity Test:				
F-statistic	1.380966	Probability	0.272917	
Obs*R-squared	5.418726	Probability	0.246966	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 01/19/06 Time: 16:23				
Sample: 1 27				
Included observations: 27				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.583444	0.856274	-0.681376	0.5027
LOG(K)	-0.112749	0.269050	-0.419065	0.6792
(LOG(K))^2	0.011926	0.019241	0.619808	0.5418
LOG(L)	0.358410	0.457469	0.783463	0.4417
(LOG(L))^2	-0.038152	0.040526	-0.941415	0.3567
R-squared	0.200694	Mean dependent var	0.031542	
Adjusted R-squared	0.055365	S.D. dependent var	0.056811	
S.E. of regression	0.055216	Akaike info criterion	-2.789558	
Sum squared resid	0.067073	Schwarz criterion	-2.549588	
Log likelihood	42.65904	F-statistic	1.380966	
Durbin-Watson stat	1.980465	Prob(F-statistic)	0.272917	

Kết quả: Không có HET ở mức $\alpha = 10\%$.

Chương VII

TƯƠNG QUAN CHUỖI (Auto Regression)

7.1. GIỚI THIỆU:

Một trong những giả thiết quan trọng của mô hình hồi quy tuyến tính ước lượng theo phương pháp OLS là các số hạng sai số u_i không tương quan với nhau.

Tuy nhiên, trong nhiều trường hợp thông thường có liên quan đến những dữ liệu thu thập theo thời gian, giả thuyết này có thể sai. Có nghĩa là số hạng sai số u_i của một mẫu quan sát cụ thể nào đó có quan hệ tuyến tính với một hay nhiều các số hạng sai số của các quan sát khác. Hiện tượng như vậy được gọi là hiện tượng tự tương quan của số hạng sai số (gọi tắt là tương quan chuỗi).

Vậy: $\text{Cov}(u_i, u_j) = 0 \rightarrow$ Không có hiện tượng tương quan chuỗi.

$\text{Cov}(u_i, u_j) \neq 0 \rightarrow$ Có hiện tượng tương quan chuỗi.

Phương trình tổng thể :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + u_t$$

AR(p): tương quan chuỗi bậc p

$$u_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-2} + \dots + \alpha_p u_{t-p} + \epsilon_t$$

7.2. HẬU QUẢ:

7.2.1. Tác động lên tính chất của các ước lượng:

Nếu mô hình có hiện tượng tương quan chuỗi (AR), thì các tính chất: không chệch và nhất quán không bị vi phạm nếu ta sử dụng OLS để ước lượng các hệ số hồi quy.

Nghĩa là: $E(\hat{\beta}_k) = \beta_k$

Tính chất nhất quán sẽ không còn nếu các biến phụ thuộc có hiệu ứng trễ được gộp vào xem như các biến giải thích.

7.2.2. Tác Động Lên Các Kiểm Định Giả Thuyết

Khi có hiện tượng tương quan chuỗi thì các sai số chuẩn ước lượng sẽ khác các sai số thực, và do đó sẽ là ước lượng không chính xác. Điều này làm cho các trị thống kê t và F được ước lượng không đúng. Vì vậy, các kiểm định t và F không còn hợp lệ, các kiểm định giả thuyết không còn giá trị nữa.

7.2.3. Tác Động Lên Việc Dự Báo

Do các ước lượng OLS vẫn không chệch, nên các dự báo dựa trên những giá trị ước lượng này cũng sẽ không thiên lệch tuy nhiên dự báo sẽ không hiệu quả do sai số lớn.

7.3. NGUYÊN NHÂN XẢY RA HIỆN TƯỢNG AR

- Do bản chất của các mối quan hệ kinh tế.
- Do kỹ thuật thu thập dữ liệu, kỹ thuật đo lường, dạng hàm số của mô hình.
- Do con người học được hành vi trong quá khứ.

7.4. NHẬN DẠNG HIỆN TƯỢNG HET

7.4.1. Bằng trực giác và kinh nghiệm:

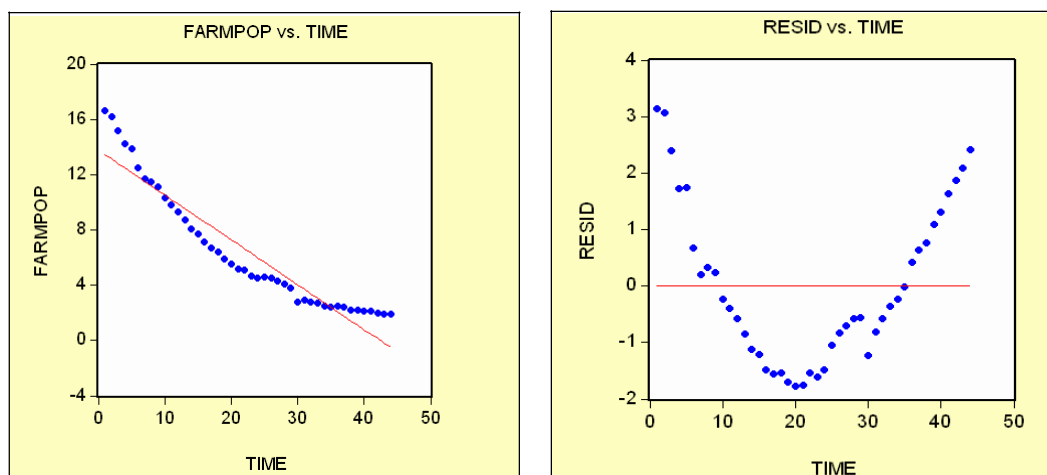
Làm việc thường xuyên với số liệu, ta sẽ có một “cảm giác” tốt hơn với số liệu, thông thường với dữ liệu thời gian (time-series data), rất có khả năng ta có hiện tượng tương quan chuỗi.

7.4.2. Phân tích bằng Biểu đồ (Graphical analysis)

Để phát hiện AR người ta thường dùng các đồ thị phân tán giữa: (X_t, Y_t) (X_t, u_t) hoặc (u_t, u_{t-1}) và thay thế X_t bằng \hat{Y}_t với mô hình đa biến.

Ví dụ: DATA6-6 có dữ liệu hàng năm về dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ năm 1948 đến 1991

Mô hình hồi quy: $FARMPOP = \beta_1 + \beta_2 TIME + u$, trong đó TIME là biến xu thế.



Đồ thị giữa (X_i, Y_i)

Đồ thị giữa (X_i, u_i)

Hình 7.1: Đồ thị nhận biết hiện tượng AR

Theo đồ thị ta thấy các số hạng phần dư có mối quan hệ với nhau theo thời gian \square có dấu hiệu của hiện tượng tương quan chuỗi.

Bằng trực giác, kinh nghiệm hay đồ thị chỉ cho ta biết dấu hiệu để nhận dạng hiện tượng AR. Để có kết luận chính thức về hiện tượng AR ta phải thực hiện các kiểm định phù hợp.

7.4.3. Kiểm định tương quan chuỗi bậc nhất (Durbin - Watson):

Kiểm định Durbin – Watson chỉ nhận dạng được hiện tượng tương quan chuỗi bậc 1

Đôi khi Kiểm định Durbin – Watson không cho kết luận.

Phương trình hồi quy tổng thể :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + u_t$$

$$AR(1): u_t = \rho_1 u_{t-1} + \epsilon_t$$

Các bước kiểm định:

Bước 1: Thực hiện hồi quy phụ:

$$u_t = \rho_1 u_{t-1} + \epsilon_t \quad -1 < \rho_1 < 1$$

Bước 2: Phát biểu giả thiết:

$H_0: \rho_1 = 0$ Không có hiện tượng AR(1).

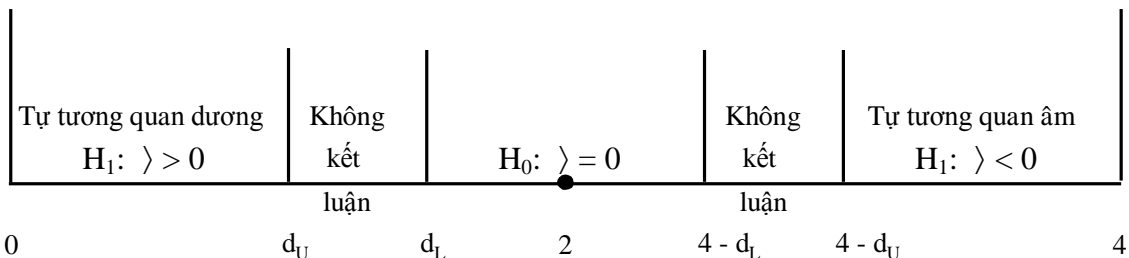
$H_1: \rho_1 \neq 0$ Có hiện tượng AR(1).

Bước 3: Tính toán thống kê Durbin-Watson

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2} \quad \text{Ta có: } \hat{H} = \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2}$$

Do đó: $DW \approx 2(1 - \hat{H})$

Bước 4: Tra bảng tìm giá trị : d_U và d_L . Xác định xem DW nằm trong khoảng nào để có các kết luận phù hợp.



Lưu ý: Trong thực hành bằng EVIEW, trị thống kê Durbin-Watson DW được tính toán sẵn (thể hiện trên bảng kết quả).

Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ năm 1948 đến 1991. Chọn $\alpha = 5\%$.

Dependent Variable: FARMPOP				
Method: Least Squares				
Date: 01/18/06 Time: 11:17				
Sample: 1948 1991				
Included observations: 44				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.77727	0.436669	31.55083	0.0000
TIME	-0.324848	0.016902	-19.22003	0.0000
R-squared	0.897912	Mean dependent var	6.468182	
Adjusted R-squared	0.895481	S.D. dependent var	4.403581	
S.E. of regression	1.423649	Akaike info criterion	3.588713	
Sum squared resid	85.12467	Schwarz criterion	3.669813	
Log likelihood	-76.95169	F-statistic	369.4094	
Durbin-Watson stat	0.055649	Prob(F-statistic)	0.000000	

Hình 7.2: Mô hình hồi quy cơ bản.

Mô hình ước lượng: $FARMPOP_t = 13.77727 - 0.324848TIME_t + \hat{u}_t$

Ta có: $DW = 0.055649$. Số quan sát $n = 44$ và $k' = 1$, $d_L = 1,47$ và $d_U = 1,56$.

Vì $DW < d_L$ ® Bác bỏ H_0 tại mức $\alpha = 5\%$

Kết luận: Có tương quan chuỗi dương trong các phần dư có ý nghĩa tại $\alpha = 5\%$.

7.4.4. **Kiểm định nhân tử Lagrang :**

Kiểm định LM trong việc nhận dạng tương quan chuỗi không chỉ với bậc nhất mà cũng cho cả các bậc cao hơn:

Ta có :

$$(R) : Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + u_t$$

$$(U) : Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + \gamma_1 u_{t-1} + \dots + \gamma_p u_{t-p} + \epsilon_t$$

Nếu chọn mô hình (U) : nghĩa là mô hình có AR(p).

Các bước kiểm định :

Bước 1: Chạy mô hình hồi quy (R).

Bước 2: Phát biểu giả thuyết kiểm định :

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

Không có hiện tượng AR(p).

$$H_1: \text{Có ít nhất 1 số } \gamma_j \neq 0 \text{ (j = 1, p)}$$

Có hiện tượng AR(p).

Bước 3: Tính $|t| = (n - p)R^2_{hqp}$

Tra bảng $|t|^* = |t|_{p, \alpha}$

Nếu: $|t| > |t|^*$ (hay p-value $< \alpha$) \rightarrow Bác bỏ $H_0 \rightarrow$ Có AR(1)

Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ năm 1948 đến 1991. Chọn $\alpha = 5\%$.

Sau khi chạy mô hình (R) : $FARMPOP_t = 13.77727 - 0.324848TIME_t + \hat{u}_t$

Ta tiến hành thực hiện kiểm định AR(1) như sau :

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	209.9702	Probability	0.000000	
Obs*R-squared	36.81190	Probability	0.000000	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 01/18/06 Time: 11:34				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.111732	0.178801	-0.624895	0.5355
TIME	0.007287	0.006932	1.051126	0.2994
RESID(-1)	0.951049	0.065633	14.49035	0.0000
R-squared	0.836634	Mean dependent var	1.03E-15	
Adjusted R-squared	0.828665	S.D. dependent var	1.406998	
S.E. of regression	0.582394	Akaike info criterion	1.822406	
Sum squared resid	13.90648	Schwarz criterion	1.944055	
Log likelihood	-37.09293	F-statistic	104.9851	
Durbin-Watson stat	1.058810	Prob(F-statistic)	0.000000	

Hình 7.3: Kiểm định LM để nhận dạng AR(1).

Giả thuyết kiểm định:

$H_0: \gamma_1 = 0$ Không có hiện tượng AR(1).

$H_1: \gamma_1 \neq 0$ Có hiện tượng AR(1).

Ta có : p-value $H_0 < \alpha \rightarrow$ Bác bỏ H_0 (Hay có tương quan chuỗi bậc nhất)

Lưu ý: Do giả thuyết H_1 của kiểm định là chỉ cần tồn tại ít nhất một số $\gamma_j \neq 0$, nên nếu mô hình có AR(1) thì các kiểm định bậc cao hơn sẽ luôn chọn H_1 và như vậy ta sẽ không xác định được bậc của AR.

Do đó, trong thực hành ta nên thực hiện kiểm định kiểm tra AR(1), nếu có AR(1) thì khắc phục AR(1) và quay trở lại bước kiểm định nhận dạng.

7.4.5. Kiểm định Correlogram:

Một phương pháp khác giúp nhận dạng AR là kiểm định Q. Để thực hiện kiểm định này chúng ta cần xem xét một khái niệm “tự tương quan” (AutoCorrellation - AC)

$$AC_k = \rho(u_t, u_{t-k}): \text{Hệ số tương quan giữa } u_t \text{ và } u_{t-k}$$

Phát biểu giả thuyết kiểm định :

$$H_0: AC_1 = AC_2 = \dots = AC_p = 0 \quad \text{Không có hiện tượng AR}(p).$$

$$H_1: \text{Có ít nhất 1 số } AC_j \neq 0 \quad (j = \overline{1, p}) \quad \text{Có hiện tượng AR}(p).$$

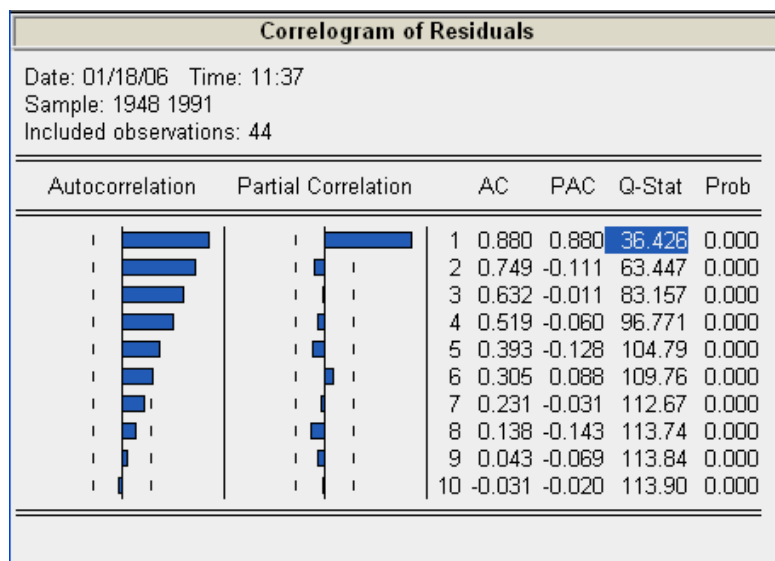
Trị số thống kê kiểm định (Box-Lung):

$$Q_{BL} = \frac{\sum_{k=1}^p AC_k^2}{n}$$

Nếu: $Q_{BL} > \chi^2_{\alpha, df}$ (hay p-value $< \alpha$) \rightarrow Bác bỏ $H_0 \rightarrow$ Có AR(p)

Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ năm 1948 đến 1991. Chọn $\alpha = 5\%$.

Sau khi chạy mô hình cơ bản (hình 7.2), ta thực hiện kiểm định AR(1) như sau :



Hình 7.4: Kiểm định LM để nhận dạng AR(1).

$$H_0: \rho_1 = 0 \quad \text{Không có hiện tượng AR}(1).$$

$$H_1: \rho_1 \neq 0 \quad \text{Có hiện tượng AR}(1).$$

Ta có : Q-stat = 36,426 hay p-value $H_0 < \alpha \rightarrow$ Bác bỏ H_0 hay có AR(1)

7.5. BIÊN PHÁP KHÁC PHỤC

7.5.1. Thay đổi dạng hàm:

Hiện tượng tương quan chuỗi có thể xảy ra trong trường hợp nhận dạng sai dạng hàm của mô hình, trong trường hợp này ta phải xây dựng lại mô hình bằng một dạng hàm đúng.

Ví dụ: Giả sử rằng đáng ra ta hồi qui Y theo X và X^2 . Nếu X tăng hoặc giảm có hệ thống theo thời gian, thì hồi qui của Y chỉ theo X sẽ thể hiện sự tương quan chuỗi. Không có thủ tục ước lượng nào có thể hiệu chỉnh vấn đề mà nó thực sự do đặc trưng sai trong phần xác định hơn là trong số hạng sai số. Một giải pháp ở đây là thiết lập lại mô hình có tính đến số hạng bậc hai sao cho không có tương quan chuỗi xuất hiện.

Các giải pháp hay dùng trong trường hợp này:

- Tìm các dạng phù hợp khác.
- Đưa biến độ trễ vào mô hình.
- Lập mô hình các sai phân bậc nhất.

7.5.2. Thủ tục ước lượng:

🔧 Ước lượng phương trình cơ bản bằng OLS và tính toán phần dư của nó \hat{u}_t .

🔧 Ước lượng hệ số tương quan chuỗi bậc nhất (còn gọi là $\hat{\rho}$).

🔧 Biến đổi các biến như sau:

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1}, X_{2t}^* = X_{2t} - \hat{\rho} X_{2(t-1)} \quad \dots \text{v.v.}$$

Lưu ý rằng các biến có dấu (*) được xác định chỉ với t nhận giá trị từ 2 đến n vì có t-1 số hạng xuất hiện.

🔧 Hồi quy Y_t^* theo $X_{2t}^*, X_{3t}^*, \dots$ và xác định các tham số ước lượng này.

🔧 Sử dụng những tham số ước lượng vừa tính toán thay vào các giá trị tham số ước lượng trong phương trình cơ bản ta sẽ tính toán được được một tập mới các giá trị ước lượng \hat{u}_t mới.

🔧 Sau đó, quay tính lặp bước 2 với những giá trị mới này cho đến khi có thể áp dụng được quy tắc dừng sau.

Quy tắc dừng: Thủ tục tính lặp trên đây có thể dừng lại khi hiệu số giá trị ước lượng của ρ từ hai kết quả liên tiếp tính được không lớn hơn giá trị chọn trước nào đó, như 0,001 chẳng hạn.

Lưu ý: Trong EVIEW, các bước lặp đã được thực hiện và ra kết quả sau cùng.

Ví dụ: Theo ví dụ dân số nông trại theo phần trăm tổng dân số tại Mỹ FARMPOP từ năm 1948 đến 1991. Chọn $\alpha = 5\%$.

Sau khi dùng các kiểm định nhận dạng ta kết luận có hiện tượng AR(1) trong mô hình. Để khắc phục hiện tượng này, ta tiến hành thủ tục ước lượng thêm AR(1) vào mô hình cơ bản (mô hình R) để được mô hình U như sau:

Dependent Variable: FARMPOP				
Method: Least Squares				
Date: 01/19/06 Time: 15:23				
Sample(adjusted): 1949 1991				
Included observations: 43 after adjusting endpoints				
Convergence achieved after 4 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.110588	17.86673	-0.286039	0.7763
TIME	0.087428	0.267420	0.326933	0.7454
AR(1)	0.956023	0.027084	35.29805	0.0000
R-squared	0.996830	Mean dependent var	6.232558	
Adjusted R-squared	0.996672	S.D. dependent var	4.165603	
S.E. of regression	0.240311	Akaike info criterion	0.053448	
Sum squared resid	2.309974	Schwarz criterion	0.176323	
Log likelihood	1.850863	F-statistic	6289.979	
Durbin-Watson stat	2.290753	Prob(F-statistic)	0.000000	
Inverted AR Roots	.96			

Hình 7.5: Thực hiện thủ tục ước lượng để khắc phục AR(1).

Mô hình ước lượng:

$$FARMPOP_t = -5.110588 + 0.087428TIME_t - 0.956023 \hat{u}_{t-1} + \hat{\epsilon}_t$$

Sau đó ta lại tiến hành kiểm định nhận dạng AR cho mô hình hình 7.5 như sau:

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	1.545178	Probability	0.221270	
Obs*R-squared	1.638732	Probability	0.200500	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 01/19/06 Time: 15:25				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.500721	17.96828	-0.194828	0.8465
TIME	0.051139	0.268783	0.190261	0.8501
AR(1)	0.005438	0.027255	0.199513	0.8429
RESID(-1)	-0.197927	0.159227	-1.243052	0.2213
R-squared	0.038110	Mean dependent var	-9.32E-14	
Adjusted R-squared	-0.035882	S.D. dependent var	0.234519	
S.E. of regression	0.238690	Akaike info criterion	0.061105	
Sum squared resid	2.221941	Schwarz criterion	0.224937	
Log likelihood	2.686250	F-statistic	0.515059	
Durbin-Watson stat	2.028883	Prob(F-statistic)	0.674328	

Giả thuyết kiểm định:

$H_0: \rho_1 = 0$ Không có hiện tượng AR(1) trong mô hình 7.5.

$H_1: \rho_1 \neq 0$ Có hiện tượng AR(1) trong mô hình 7.5.

Ta có : $p\text{-value} = 0.200500 > \alpha \rightarrow$ Không bác bỏ H_0 . (Không có hiện tượng AR(1) trong mô hình 7.5 ở mức ý nghĩa $\alpha = 5\%$)