

CHƯƠNG 4.

HỒI QUI VỚI BIẾN GIẢ

I.1. Bản chất của biến giả

- Biến định tính thể hiện sự có hay không một tính chất, trạng thái của đối tượng quan sát.
- Các biến định tính đã được lượng hoá mang các giá trị 0 và 1 được gọi là biến giả. Biến giả thường được kí hiệu là D (dummy).

I.1. Bản chất của biến giả

- Nguyên tắc xây dựng biến giả
 - ☞ Nếu biến định tính có m thuộc tính thì số biến giả cần xây dựng là $(m - 1)$
(Nếu ta đưa vào mô hình m biến giả tương ứng với m thuộc tính thì mô hình sẽ có hiện tượng đa cộng tuyến hoàn hảo (chương V))
 - ☞ Thuộc tính không được đại diện bằng biến giả riêng biệt gọi là thuộc tính cơ sở. Thuộc tính này xuất hiện khi tất cả các biến giả thể hiện biến định tính bằng 0

I.1. Bản chất của biến giả

Ví dụ:

Biển miền gồm 3 thuộc tính Bắc, Trung, Nam \Rightarrow xây dựng 2 biến giả:

$$D_{1i} = \begin{cases} 1 \text{ nếu quan sát thứ } i \text{ thuộc miền Bắc} \\ 0 \text{ nếu quan sát thứ } i \text{ không thuộc miền Bắc} \end{cases}$$

$$D_{2i} = \begin{cases} 1 \text{ nếu quan sát thứ } i \text{ thuộc miền Trung} \\ 0 \text{ nếu quan sát thứ } i \text{ không thuộc miền Trung} \end{cases}$$

Vậy, miền Nam được gọi là thuộc tính cơ sở và nó xảy ra khi $D_{1i} = D_{2i} = 0$

I.1. Bản chất của biến giả

Xây dựng các biến giả để đưa vào mô hình biến định tính “trình trạng hôn nhân”



I.2. Hồi qui với biến giả

Xét mô hình các yếu tố quyết định thu nhập
Ta có 100 quan sát về mức thu nhập cá nhân (Y), số năm kinh nghiệm (X), vùng (R), giới tính (S)

I.2. Hồi qui với biến giả

a. Mô hình hồi qui trong đó biến giả thích chỉ gồm biến định tính

☆ Biến định tính có 2 thuộc tính

Ta bắt đầu với mô hình đơn giản: tiền lương phụ thuộc vào vùng địa lý

I.2. Hồi qui với biến giả

⇒ Các biến giả sẽ được định nghĩa như sau:

$$R_{1i} = \begin{cases} 1 & \text{nếu quan sát thứ } i \text{ ở miền Bắc} \\ 0 & \text{nếu quan sát thứ } i \text{ không ở miền Bắc} \end{cases}$$

$$R_{2i} = \begin{cases} 1 & \text{nếu quan sát thứ } i \text{ ở miền Trung} \\ 0 & \text{nếu quan sát thứ } i \text{ không ở miền Trung} \end{cases}$$

⇒ Thuộc tính cơ sở ?

I.2. Hồi qui với biến giả

Mô hình: $Y_i = \beta_1 + \beta_2 R_{1i} + \beta_3 R_{2i} + U_i$

- Hệ số chặn là β_1 nếu $R_{1i} = R_{2i} = 0$: cho biết tiền lương trung bình của lao động ở miền Nam
- Hệ số chặn là $\beta_1 + \beta_2$ nếu $R_{1i} = 1$ và $R_{2i} = 0$: cho biết tiền lương trung bình lao động ở miền Bắc
- Hệ số chặn là $\beta_1 + \beta_3$ nếu $R_{1i} = 0$ và $R_{2i} = 1$: cho biết tiền lương trung bình lao động ở miền Trung

I.2. Hồi qui với biến giả

- Vậy, β_1 cho biết giá trị trung bình của biến phụ thuộc ứng với thuộc tính cơ sở.
- β_2, β_3 được gọi là “hệ số chặn chênh lệch” (differential intercept coefficient) cho biết giá trị của hệ số chặn khi biến giả nhận giá trị 1 chênh lệch bao nhiêu so với hệ số chặn khi biến giả nhận giá trị 0.

I.2. Hồi qui với biến giả

***b. Mô hình hồi qui trong đó biến giả
thích gồm biến định tính và biến định
lượng***

☆ Trong mô hình chỉ có 1 biến định tính làm biến giải thích

Ví dụ, xét trường hợp tiền lương của lao động phụ thuộc vào biến vùng và số năm kinh nghiệm.

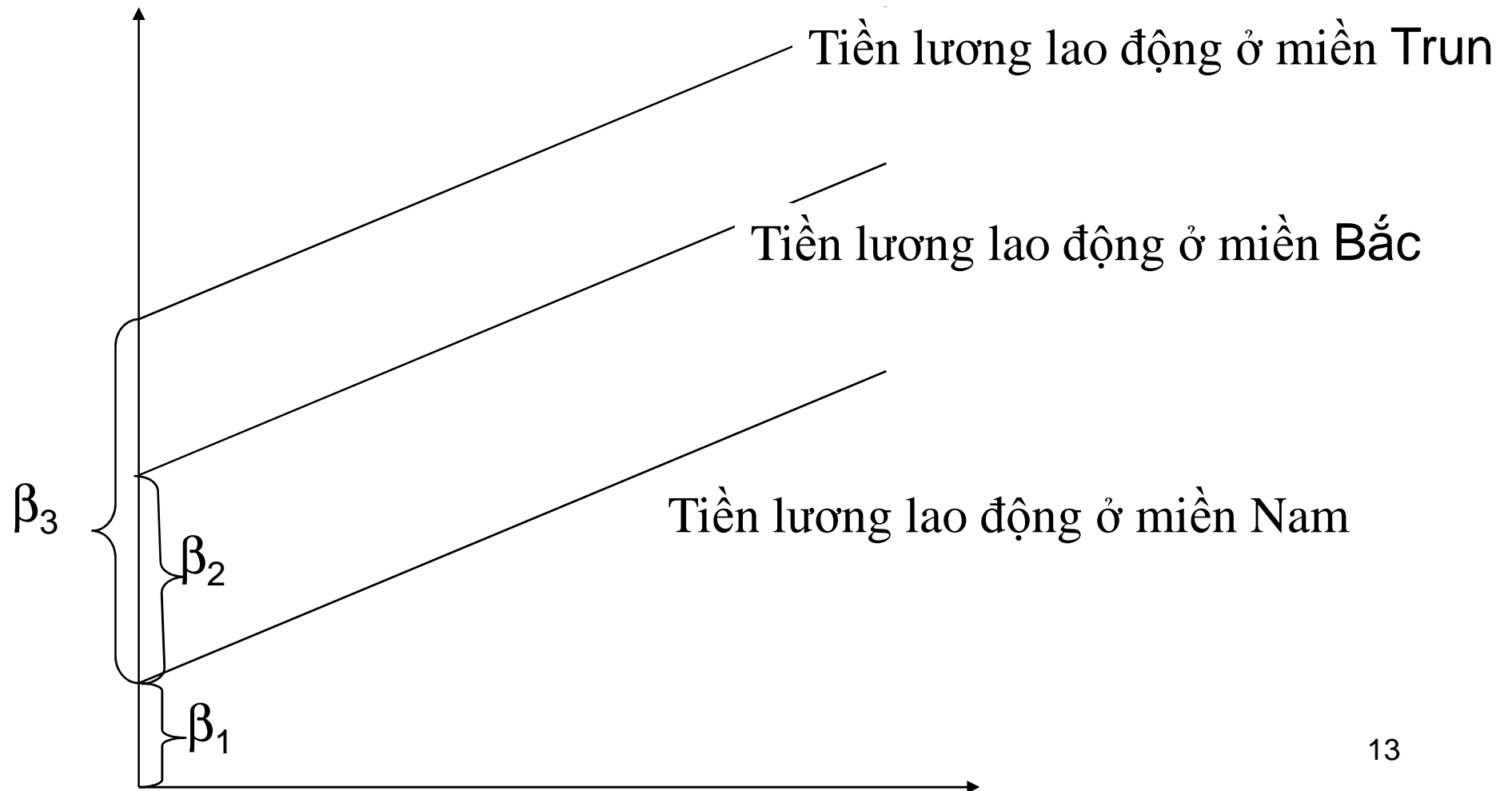
I.2. Hồi qui với biến giả

Mô hình: $Y_i = \beta_1 + \beta_2 R_{1i} + \beta_3 R_{2i} + \beta_4 X_i + u_i$

- Tiền lương trung bình của lao động ở miền Nam:
- Tiền lương trung bình của lao động ở miền Bắc:
- Tiền lương trung bình của lao động ở miền Trung:



I.2. Hồi qui với biến giả



I.2. Hồi qui với biến giả

☆ Trong mô hình có nhiều hơn 1 biến định tính làm biến giải thích

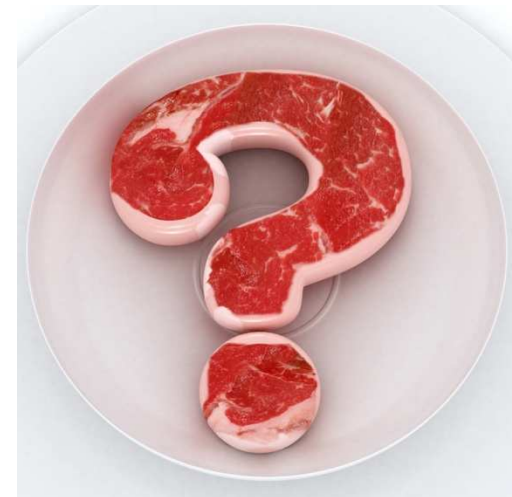
Ví dụ, xét trường hợp tiền lương của lao động phụ thuộc vào biến vùng và số năm kinh nghiệm và giới tính

$$S_i = \begin{cases} 1 & \text{nếu người lao động là nữ} \\ 0 & \text{nếu người lao động là nam} \end{cases}$$

I.2. Hồi qui với biến giả

$$Y_i = \beta_1 + \beta_2 R_{1i} + \beta_3 R_{2i} + \beta_4 X_i + \beta_5 S_i + u_i$$

- Biểu diễn tiền lương trung bình của lao động:
 - nam ở miền Bắc?
 - nam ở miền Trung?
 - nam ở miền Nam?
 - nữ ở miền Bắc?
 - nữ ở miền Trung?
 - nữ ở miền Nam?



I.2. Hồi qui với biến giả

★ Trường hợp biến định tính có ảnh hưởng đến hệ số góc.

Ví dụ, tác động của giới tính và số năm kinh nghiệm tới tiền lương.

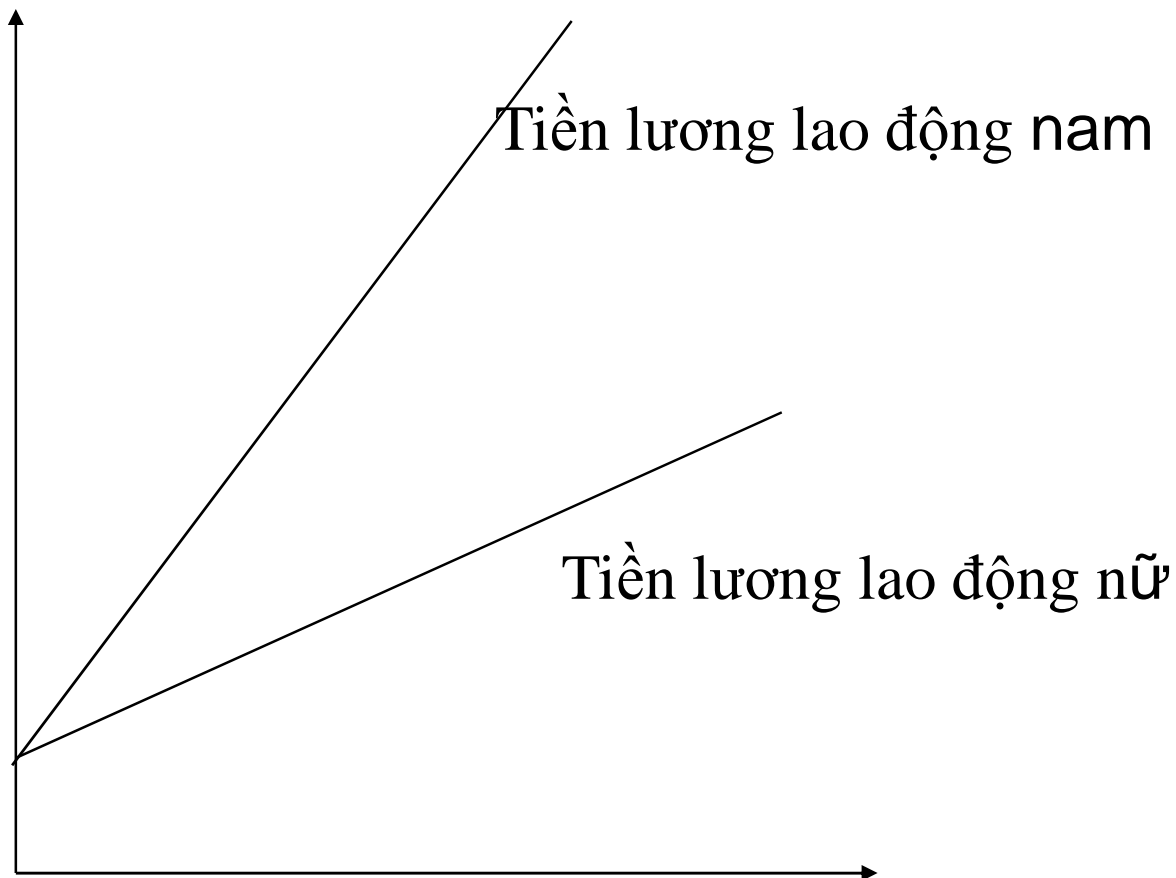
I.2. Hồi qui với biến giả

Mô hình: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_i X_i + u_i$

- $E(Y/X_i, S_i = 0) = \beta_1 + \beta_2 X_i$: tiền lương trung bình của lao động nam
- $E(Y/X_i, S_i = 1) = \beta_1 + (\beta_2 + \beta_3) X_i$: tiền lương trung bình của lao động nữ

I.2. Hồi qui với biến giả

- Giải sử $\beta_3 < 0$ ta có đồ thị sau:





1.2. Hồi qui với biến giả

- Xây dựng mô hình kinh tế lượng mà ở đó tiền lương chịu ảnh hưởng của số năm kinh nghiệm và giới tính (Giới tính có tác động tới cả hệ số chặn và hệ số góc).
- Tìm biểu thức biểu diễn:
 - tiền lương trung bình của lao động nữ
 - tiền lương trung bình của lao động nam
- Vẽ đồ thị minh họa 2 trường hợp trên

II. Một số ứng dụng của biến giả trong phân tích kinh tế

II.1. So sánh hai hồi qui

- Ví dụ, nghiên cứu về mối quan hệ giữa chi tiêu và qui mô hộ gia đình ở thành thị và nông thôn, ta có các hồi qui:

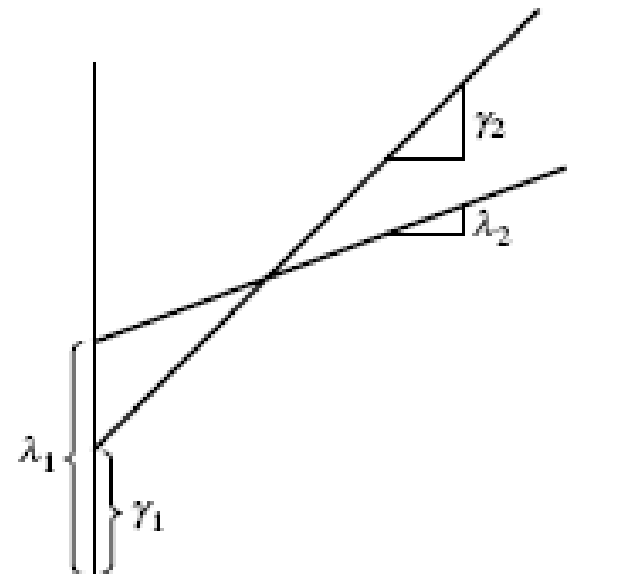
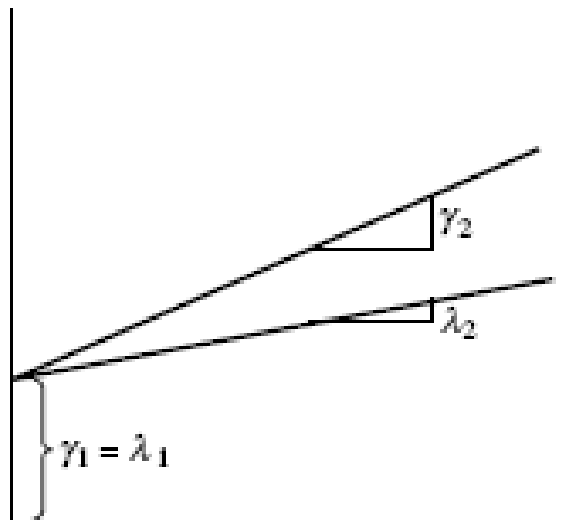
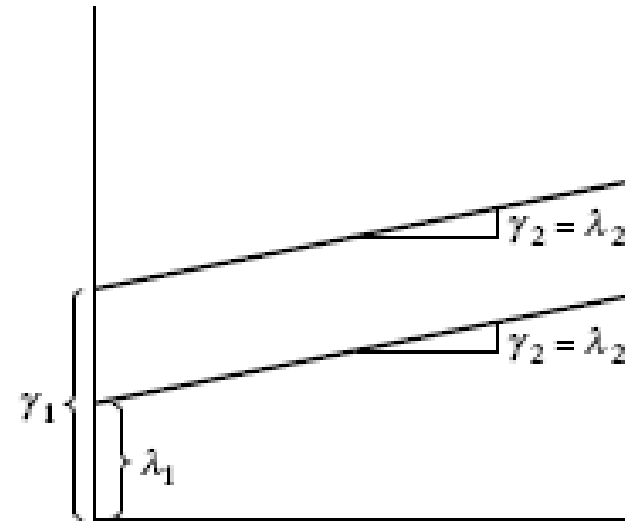
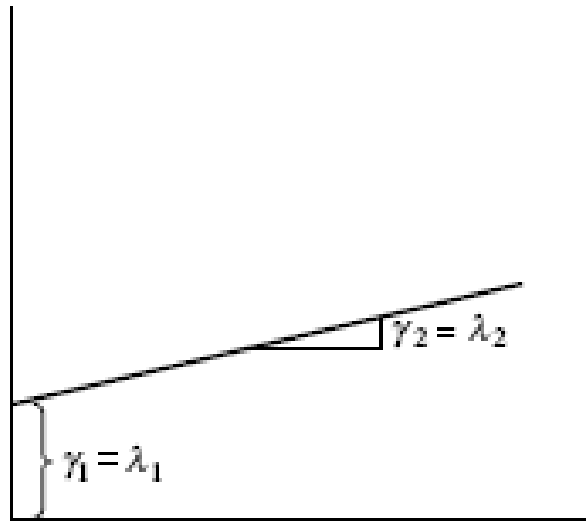
- Ở thành thị:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{1i} \quad (i = 1, n_1) \quad (1)$$

- Ở nông thôn:

$$Y_j = \gamma_1 + \gamma_2 X_j + u_{2j} \quad (j = 1, n_2) \quad (2)$$

Y: chi tiêu, X: qui mô hộ



II.1. So sánh hai hồi qui

a. Kiểm định Chow

Kiểm định Chow dựa trên giả thiết: U_{1i} và U_{2j} có phân phối chuẩn $N(0, \sigma^2)$ và chúng có phân phối độc lập.

- Với cặp giả thiết:

H_0 : hồi qui (1) và (2) là như nhau

H_1 : hồi qui (1) và (2) khác nhau

II.1. So sánh hai hồi qui

a. Kiểm định Chow

Các bước kiểm định như sau:

- *Bước 1*: kết hợp tất cả các quan sát của 2 thời kỳ, $n = n_1 + n_2$;

Hồi qui với n quan sát:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (i=1, n)$$

$$\Rightarrow \text{RSS với df} = n - k \quad (k=2)$$

II.1. So sánh hai hồi qui

a. Kiểm định Chow

- *Bước 2:*

Ước lượng mô hình (1) thu được RSS_1
có $df_1 = n_1 - k$

Ước lượng mô hình (2) thu được RSS_2
có $df_2 = n_2 - k$

Đặt $\overline{RSS} = RSS_1 + RSS_2 \Rightarrow \overline{RSS}$ có số bậc
tự do là $n_1 + n_2 - 2k$

II.1. So sánh hai hồi qui

a. Kiểm định Chow

Bước 3: kiểm định

$$F = \frac{(\text{RSS} - \overline{\text{RSS}}) / k}{\overline{\text{RSS}} / (n - 2k)}$$

II.1. So sánh hai hồi qui

Kiểm định Chow kiểm định sự đồng thời bằng nhau của các cặp hệ số hồi qui trong hai mô hình chứ không cho phép kiểm định riêng rẽ sự bằng nhau của từng cặp hệ số. Kỹ thuật biến giả có thể làm được điều này.

II.1. So sánh hai hồi qui

b. Kỹ thuật biến giả

Gộp 2 bộ số liệu lại, được n quan sát và tiến hành hồi qui sau với n quan sát

$$Y_i = \beta_1 + \beta_2 D_i + \beta_3 X_i + \beta_4 (D_i X_i) + U_i$$

$$D = \begin{cases} 1 & \text{với các quan sát ở thành thị} \\ 0 & \text{với các quan sát ở nông thôn} \end{cases}$$

- $E(Y/X_i, D_i = 1) = (\beta_1 + \beta_2) + (\beta_3 + \beta_4)X_i$
- $E(Y/X_i, D_i = 0) = \beta_1 + \beta_3 X_i$

II.1. So sánh hai hồi qui

b. Kỹ thuật biến giả

- Muốn so sánh 2 hồi qui ta lần lượt kiểm định sự bằng 0 của các hệ số β_2 và β_4 .

- Các giả thiết:

H_0 : Hai hệ số chặn giống nhau $\Leftrightarrow \beta_2 = 0$

H_0 : Hai hệ số góc giống nhau $\Leftrightarrow \beta_4 = 0$

H_0 : Biến giả không cần thiết trong mô hình

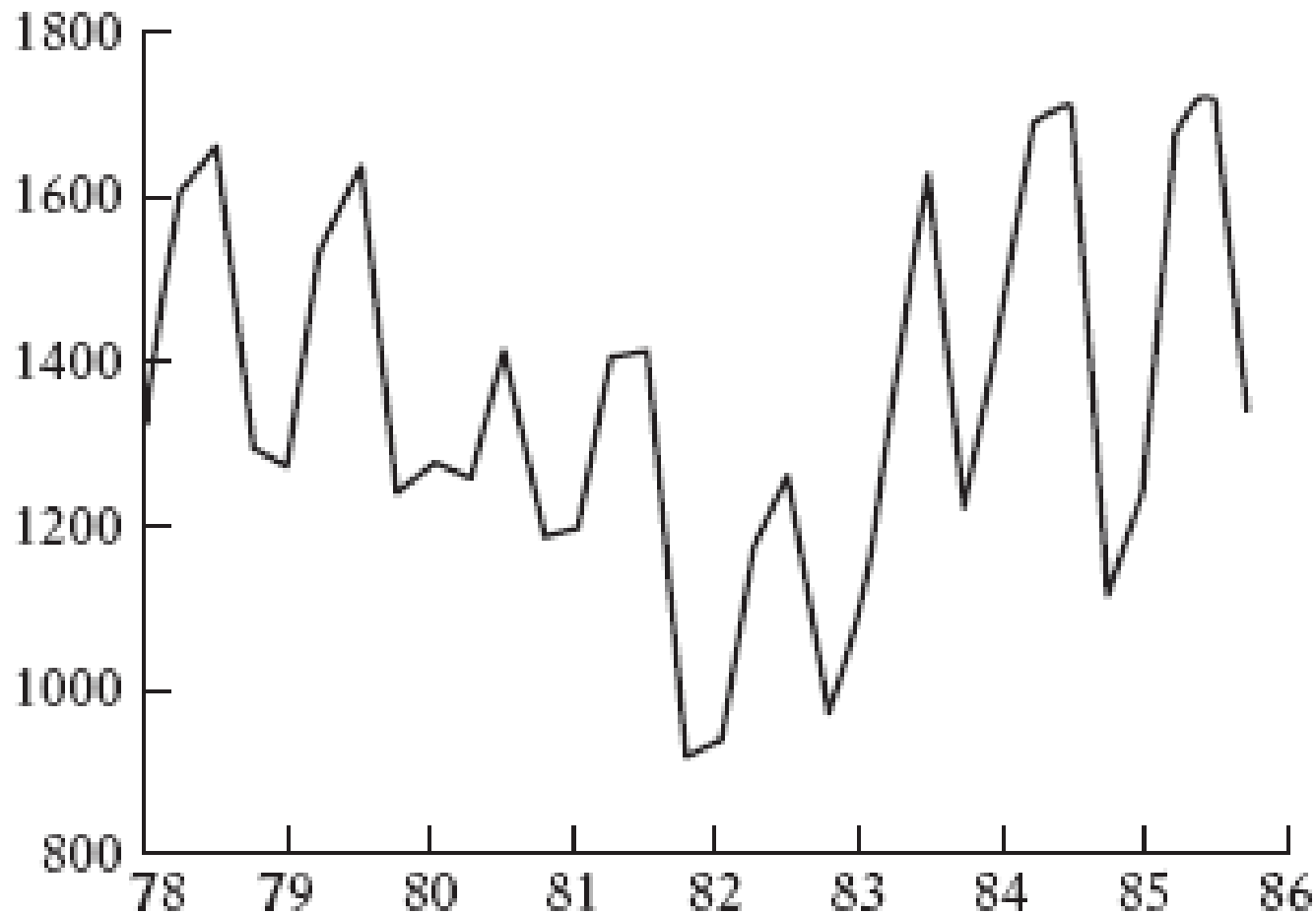
\Leftrightarrow Hai hồi qui giống nhau $\Leftrightarrow \beta_2 = \beta_4 = 0$

II.2. Biến giả trong phân tích thời vụ

Ví dụ:

Nghiên cứu mối quan hệ giữa doanh số tiêu thụ tủ lạnh và mức chi tiêu cho hàng lâu bền theo quý trong giai đoạn 1978 – 1986 của một vùng, ta có đồ thị về doanh số tiêu thụ tủ lạnh như sau:

II.2. Biến giả trong phân tích thời vụ



Hồi qui với biến giả

Nguyễn Thị Minh Hiếu, 2010

II.2. Biến giả trong phân tích thời vụ

Biến định tính quý có thể được biểu diễn bằng 3 biến giả:

$$\begin{aligned} D_2 &= \begin{cases} 1 & \text{nếu tủ lạnh được bán ở quý II} \\ 0 & \text{nếu tủ lạnh không được bán ở quý II} \end{cases} \\ D_3 &= \begin{cases} 1 & \text{nếu tủ lạnh được bán ở quý III} \\ 0 & \text{nếu tủ lạnh không được bán ở quý III} \end{cases} \\ D_4 &= \begin{cases} 1 & \text{nếu tủ lạnh được bán ở quý IV} \\ 0 & \text{nếu tủ lạnh không được bán ở quý IV} \end{cases} \end{aligned}$$

II.2. Biến giả trong phân tích thời vụ

Mô hình:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \beta_5 X_i + u_i$$

Y: số tủ lạnh tiêu thụ

X: chi tiêu cho đồ dùng lâu bền

II.2. Biến giả trong phân tích thời vụ

- Số tử lạnh tiêu thụ trong quý I là:
- Số tử lạnh tiêu thụ trong quý II là:
- Số tử lạnh tiêu thụ trong quý III là:
- Số tử lạnh tiêu thụ trong quý IV là:



II.2. Biến giả trong phân tích thời vụ

Có thể sử dụng mô hình tổng quát hơn để xem xét ảnh hưởng tương tác giữa yếu tố mùa vụ và chỉ tiêu cho đồ dùng lâu bền lên số tử lạnh tiêu thụ:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{4i} + \beta_5 X_i + \beta_6 (D_{2i} X_i) + \beta_7 (D_{3i} X_i) + \beta_8 (D_{4i} X_i) + u_i$$

Biến giả không chỉ loại bỏ yếu tố mùa vụ trong biến Y mà còn loại bỏ yếu tố thời vụ trong biến X (nếu có)