CHƯƠNG 7

HIỆN TƯỢNG PHƯƠNG SAI THAY ĐỔI (HETEROSCEDASTICITY)

BIẾN GIẢ

MỤC TIÊU

1. Hiểu bản chất và hậu quả của phương sai sai số thay đổi

2. Biết cách phát hiện phương sai sai số thay đổi và biện pháp khắc phục

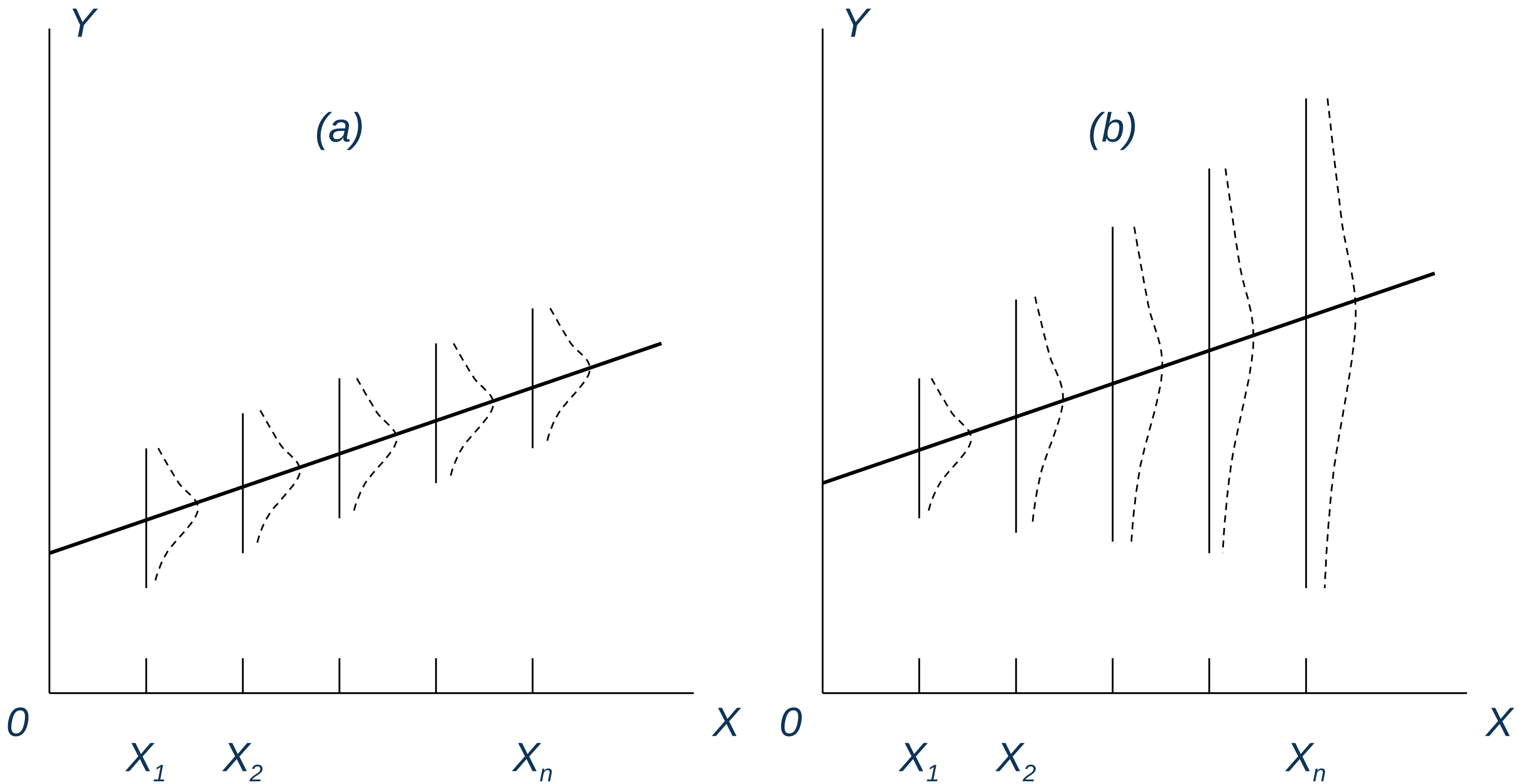
NỘI DUNG

1. Bản chất hiện tượng phương sai sai số thay đổi
2. Hậu quả
3. Cách phát hiện phương sai sai số thay đổi
4. Cách khắc phục phương sai sai số thay đổi

7.1 Bản chất

- Xét ví dụ mô hình hồi qui 2 biến trong đó biến phụ thuộc *Y* là tiết kiệm của hộ gia đình và biến giải thích *X* là thu nhập khả dụng của hộ gia đình

7.1 Bản chất



Hình 7.1: (a) Phương sai của sai số không đổi và (b) Phương sai của sai số thay đổi

7.1 Bản chất

- Hình 7.1a cho thấy tiết kiệm trung bình có khuynh hướng tăng theo thu nhập. Tuy nhiên mức độ dao động giữa tiết kiệm của từng hộ gia đình so với mức tiết kiệm trung bình không thay đổi tại mọi mức thu nhập.
- Đây là trường hợp của phương sai sai số (nhiễu) không đổi, hay phương sai bằng nhau.

$$E(u_i^2) = \sigma^2$$

7.1 Bản chất

- Trong hình 7.1b, mức độ dao động giữa tiết kiệm của từng hộ gia đình so với mức tiết kiệm trung bình thay đổi theo thu nhập. Đây là trường hợp phương sai của sai số thay đổi.

$$E(u_i^2) = \sigma_i^2$$

7.1 Nguyên nhân của phương sai thay đổi

- Do tích lũy kinh nghiệm mà sai số theo thời gian ngày càng giảm
- Do bản chất của hiện tượng kinh tế
- Công cụ về thu thập xử lý số liệu cải thiện dẫn đến sai số đo lường và tính toán giảm

7.1 Nguyên nhân của phương sai thay đổi

- Trong mẫu có các outlier (giá trị rất nhỏ hoặc rất lớn so với các giá trị quan sát khác)
- Mô hình hồi quy không đúng (dạng hàm sai, thiếu biến quan trọng)
- Hiện tượng phương sai thay đổi thường gặp khi thu thập số liệu chéo (theo không gian)

7.1 Hậu quả của phương sai thay đổi

1. Ước lượng OLS vẫn tuyến tính, không chệch nhưng không phải là ước lượng hiệu quả (vì phương sai không nhỏ nhất)
2. Ước lượng phương sai của ước lượng OLS, nhìn chung, sẽ bị chệch.

3. Các khoảng tin cậy và kiểm định giả thuyết thông thường dựa trên phân phối *t* và *F* sẽ không còn đáng tin cậy nữa.

Chẳng hạn thống kê t

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

Do sử dụng ước lượng của $SE(\beta_i)$ là $SE(\hat{\beta}_i)$ nên không đảm bảo t tuân theo quy luật phân phối t-student =>kết quả kiểm định không còn tin cậy

4. Kết quả dự báo không còn hiệu quả nữa khi sử dụng các ước lượng OLS có phương sai không nhỏ nhất.

7.2 Phương pháp phát hiện phương sai thay đổi

Phương pháp định tính

1. Dựa vào bản chất vấn đề nghiên cứu
2. Xem xét đồ thị của phần dư

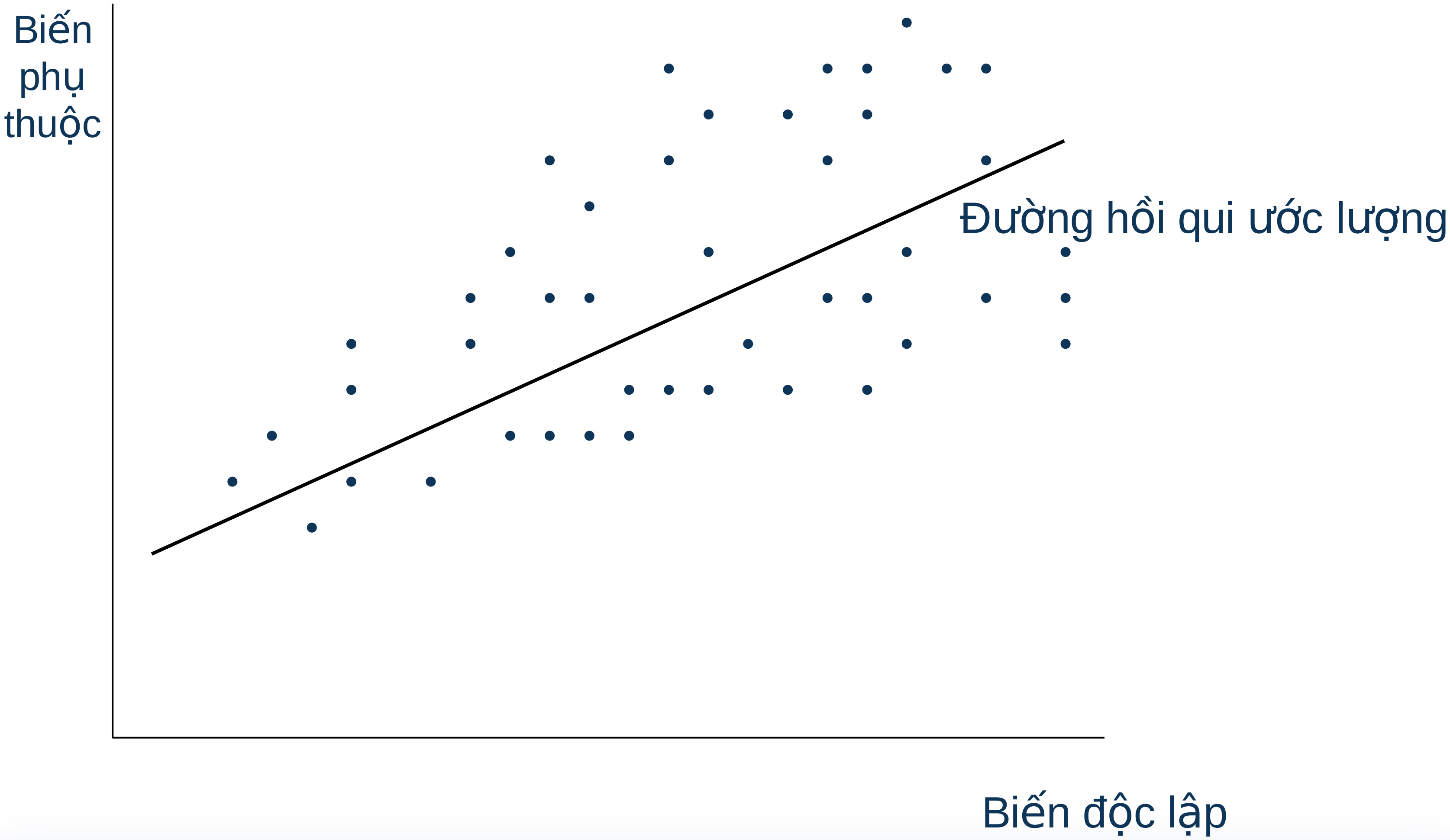
Phương pháp định lượng

1. Kiểm định Park
2. Kiểm định Glejser
3. Kiểm định Goldfeld – Quandt
4. Kiểm định White

1. Dựa vào bản chất vấn đề nghiên cứu

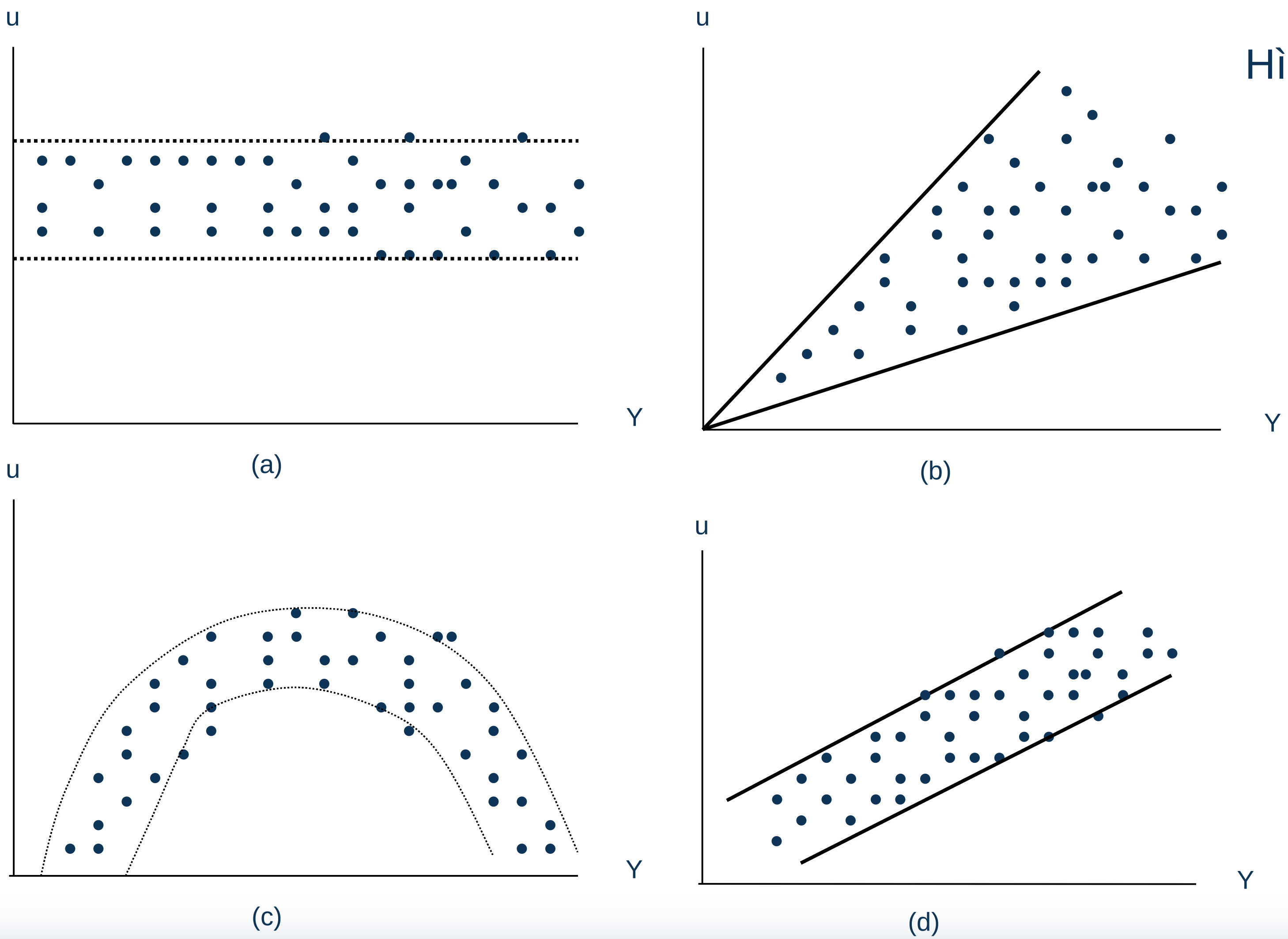
VD: nghiên cứu quan hệ giữa chi tiêu tiêu dùng so với thu nhập, phương sai phần dư của chi tiêu tiêu dùng có xu hướng tăng theo thu nhập. Do đó đối với các mẫu điều tra tương tự, người ta có khuynh hướng giả định phương sai của nhiễu thay đổi

2. Xem xét đồ thị của phần dư



2. Xem xét đồ thị của phần dư

Hình a cho thấy biến đổi của các e_i^2 không có tính hệ thống



Hình b,c,d cho thấy các e_i^2 thay đổi khi Y tăng

3. Kiểm định Park

- Park cho rằng σ_i^2 là một hàm số nào đó của biến giải thích X

$$\sigma_i^2 = B_1 + B_2 \ln|X_i| + v_i$$ trong đó v_i là phần sai số ngẫu nhiên.

- Vì σ_i^2 chưa biết, Park đề nghị sử dụng $\ln e_i^2$ thay cho σ_i^2 và chạy mô hình hồi qui sau

$$\ln e_i^2 = B_1 + B_2 \ln|X_i| + v_i \quad (*)$$

e_i^2 được thu thập từ mô hình hồi qui gốc

3. Kiểm định Park

- Các bước của kiểm định Park:

1)Chạy hàm hồi qui gốc $Y_i = \beta_1 + \beta_2 X_i + U_i$

2) Từ hàm hồi qui, tính $\hat{Y}_i$, phần dư e_i và $\ln e_i^2$

3. Chạy hàm hồi qui (*), sử dụng biến giải thích của hàm hồi qui ban đầu. Nếu có nhiều biến giải thích, chạy hồi qui cho từng biến giải thích đó. Hay, chạy hồi qui mô hình với biến giải thích $\hat{Y}_i$ là

3. Kiểm định Park

4) Kiểm định giả thuyết H_0: $\beta_2 = 0$,tức, không có phương sai của sai số thay đổi. Nếu giả thuyết H_0 bị bác bỏ, mô hình gốc có phương sai của sai số thay đổi.

5) Nếu giả thuyết H_0 được chấp nhận, B_1 trong mô hình (*) có thể được xem là giá trị chung của phương sai của sai số không đổi, σ^2.

4. Kiểm định Glejser

- Tương tự như kiểm định Park: Sau khi thu thập được phần dư từ mô hình hồi qui gốc, Glejser đề nghị chạy hồi qui | e_i | theo biến X nào mà có quan hệ chặt chẽ với σ_i^2.
- Glejser đề xuất một số dạng hàm hồi qui sau:

$$|e_i| = B_1 + B_2 X_i + v_i$$

$$|e_i| = B_1 + B_2\sqrt{X_i} + v_i$$

$$|e_i| = B_1 + B_2\frac{1}{X_i} + v_i$$

4. Kiểm định Glejser

$$|e_i| = B_1 + B_2 \frac{1}{\sqrt{X_i}} + v_i$$

$$|e_i| = \sqrt{B_1 + B_2 X_i} + v_i$$

$$|e_i| = \sqrt{B_1 + B_2 X_i^2} + v_i$$

- Nếu giả thuyết H_0: $\beta_2 = 0$ bị bác bỏ thì có thể có hiện tượng phương sai sai số thay đổi.

4. Kiểm định Glejser

- Kiểm định Glejser có một số vấn đề như kiểm định Park như sai số v_i trong các mô hình hồi qui có giá trị kỳ vọng khác không, nó có tương quan chuỗi.
  - 4 mô hình đầu cho kết quả tốt khi sử dụng OLS
  - 2 mô hình sau (phi tuyến tính tham số) không sử dụng OLS được
- Do vậy, kiểm định Glejser được dùng để chẩn đoán đối với những mẫu lớn.

5. Kiểm định Goldfeld - Quandt

- Xét mô hình hồi qui sau:

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

Giả sử σ_i^2 có quan hệ dương với biến X theo cách sau:

$$\sigma_i^2 = \sigma^2 X_i^2 \text{ trong đó } \sigma^2 \text{ là hằng số.}$$

- Các bước thực hiện kiểm định Goldfeld - Quandt như sau:

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

5. Kiểm định Goldfeld - Quandt

2. Bỏ qua quan sát ở giữa theo cách sau:

Đối với mô hình 2 biến:

$c = 4$ nếu cỡ mẫu khoảng $n = 30$;

$c = 10$ nếu cỡ mẫu khoảng $n = 60$.

và chia số quan sát còn lại thành 2 nhóm, trong đó mỗi nhóm có $(n – c)/2$ quan sát.

5. Kiểm định Goldfeld - Quandt

3. Sử dụng phương pháp OLS để ước lượng tham số của các hàm hồi qui đối với (n – c)/2 quan sát đầu và cuối; tính RSS_1 và RSS_2 tương ứng.

   Bậc tự do tương ứng là $\frac{n-c}{2}-k$ (k là các tham số được ước lượng kể cả hệ số chặn).

5. Kiểm định Goldfeld - Quandt

4. Tính tỷ số

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

λ tuân theo phân phối *F* với bậc tự do ở tử số và mẫu số là $\frac{n-c-2k}{2}$

Nếu λ > F ở mức ý nghĩa α thì bác bỏ giả thuyết H_0, nghĩa là phương sai của sai số thay đổi.

6. Kiểm định White

- White đã đề nghị một phương pháp không cần đòi hỏi *u* có phân phối chuẩn.
- Xét mô hình hồi qui sau:

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

Bước 1: Ước lượng mô hình trên bằng OLS, thu được các phần dư e_i.

Bước 2: Ước lượng một trong các mô hình sau

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + v_{2i} \quad (1)$$

6. Kiểm định White

hay

$$e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + V_{2i} \qquad (2)$$

(1) và (2) có thể có số mũ cao hơn và nhất thiết phải có hệ số chặn bất kể mô hình gốc có hay không.

R^2 là hệ số xác định bội, thu được từ (1) với mô hình không có số hạng chéo hay (2) với mô hình có số hạng chéo.

6. Kiểm định White

- *Bước 3*

  Đặt GT Ho: $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$ (1)

  $$\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0 \ (2)$$

Tương đương H_0: phương sai của sai số không đổi.

- nR^2 có phân phối xấp xỉ χ^2(df), với df bằng số hệ số của mô hình (1) và (2) không kể hệ số chặn.

6. Kiểm định White

- *Bước 4* Quy tắc quyết định
- $nR^2 < \chi^2(df)$: chấp nhận Ho
- $nR^2 > \chi^2(df)$: bác bỏ Ho, hay có hiện tượng phương sai sai số thay đổi.

1. Trường hợp đã biết σ_i^2

Có mô hình hồi qui tổng thể 2 biến:

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

giả sử rằng phương sai sai số σ_i^2 đã biết; nghĩa là phương sai sai số của mỗi quan sát đã biết, chia hai vế của mô hình cho σ_i đã biết.

$$\frac{Y_i}{\sigma_i} = \alpha_1\left(\frac{1}{\sigma_i}\right) + \alpha_2\left(\frac{X_i}{\sigma_i}\right) + \frac{u_i}{\sigma_i}$$

1. Trường hợp đã biết σ_i^2

Khi đó

$$Var\left(\frac{u_i}{\sigma_i}\right) = \frac{Var(u_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1, \forall i$$

Trong thực tế, chia mỗi quan sát Y_i và X_i cho σ_i đã biết và chạy hồi qui OLS cho dữ liệu đã được chuyển đổi này.

Ước lượng OLS của α_1 và α_2 được tính theo cách này được gọi là ước lượng bình phương bé nhất có trọng số (WLS); mỗi quan sát Y và X được chia cho trọng số (độ lệch chuẩn) của riêng nó, σ_i.

2. Trường hợp chưa biết σ_i^2

Trường hợp 1: ***Phương sai sai số tỷ lệ với biến giải thích***.

$$Var(u_i) = E(u_i^2) = \sigma^2 X_i$$

Chia hai vế của mô hình cho căn bậc hai của X_i, với $\sqrt{X_i} > 0$

$$\frac{Y_i}{\sqrt{X_i}} = \alpha_1 \frac{1}{\sqrt{X_i}} + \alpha_2 \frac{X_i}{\sqrt{X_i}} + \frac{u_i}{\sqrt{X_i}}$$

$$= \alpha_1 \frac{1}{\sqrt{X_i}} + \alpha_2 \sqrt{X_i} + v_i$$

2. Trường hợp chưa biết σ_i^2

- Khi đó

$$Var\left(\frac{u_i}{\sqrt{X_i}}\right)=\frac{Var(u_i)}{X_i}=\sigma^2,\forall i$$

- *Lưu ý là để ước lượng mô hình trên, phải sử dụng mô hình hồi qui qua gốc.*

2. Trường hợp chưa biết σ_i^2

Trường hợp 2: ***Phương sai sai số tỷ lệ với bình phương của biến giải thích***

$$Var(u_i) = E(u_i^2) = \sigma^2 X_i^2$$

Chia hai vế của mô hình cho X_i với $X_i \neq 0$

$$\frac{Y_i}{X_i} = \alpha_1\left(\frac{1}{X_i}\right) + \alpha_2 + \frac{u_i}{X_i} = \alpha_1\left(\frac{1}{X_i}\right) + \alpha_2 + v_i$$

Khi đó:

$$Var\left(\frac{u_i}{X_i}\right) = \frac{Var(u_i)}{X_i^2} = \sigma^2, \forall i$$

2. Trường hợp chưa biết σ_i^2

Trường hợp 3: ***Phương sai sai số tỷ lệ với bình phương của giá trị kỳ vọng của Y***

$$Var(u_i) = E(u_i^2) = \sigma^2[E(Y_i)]^2.$$

Chia hai vế của mô hình cho $E(Y_i)$ với

$$E(Y_i) = \hat{Y}_i = \hat{\alpha}_1 + \hat{\alpha}_2 X_i$$

2. Trường hợp chưa biết σ_i^2

Bước 1: Ước lượng mô hình hồi qui bằng phương pháp OLS:

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

và tính $\hat{Y}_i$

Biến đổi mô hình gốc về dạng như sau:

$$\frac{Y_i}{\hat{Y}_i} = \alpha_1 \frac{1}{\hat{Y}_i} + \alpha_2 \frac{X_i}{\hat{Y}_i} + v_i$$

2. Trường hợp chưa biết σ_i^2

Bước 2: Ước lượng hồi qui trên dù $\hat{Y}_i$ không chính xác là $E(Y_i \backslash X_i)$, nhưng chúng là ước lượng vững, nghĩa là khi cỡ mẫu tăng lên vô hạn thì chúng hội tụ về $E(Y_i|X_i)$. Do vậy, phép biến đổi trên có thể dùng được khi cỡ mẫu tương đối lớn.

Khi đó

$$Var\left(\frac{u_i}{\hat{Y}_i}\right) = \frac{Var(u_i)}{\hat{Y}_i^2} = \frac{\sigma^2.\left[E\left(Y_i\right)\right]^2}{\hat{Y}_i^2} \approx \sigma^2, \forall i$$

2. Trường hợp chưa biết σ_i^2

Trường hợp 4: Định dạng lại mô hình.

Thay vì ước lượng mô hình hồi qui gốc, ước lượng mô hình hồi qui:

$$\ln Y_i = \alpha_1 + \alpha_2 \ln X_i + u_i$$

Tình trạng phương sai sai số không đồng nhất sẽ bớt nghiêm trọng hơn so với mô hình gốc bởi vì khi được logarit hóa, độ lớn các biến bị 'nén lại'.

Lưu ý

Khi nghiên cứu mô hình có nhiều biến giải thích thì việc chọn biến nào để biến đổi cần phải được xem xét cẩn thận.

- Phép biến đổi logarit không dùng được khi các giá trị của các biến âm.
- Khi σ_i^2 chưa biết, nó sẽ được ước lượng từ một trong các cách biến đổi trên. Các kiểm định *t*, *F* mà chúng ta sử dụng chỉ đáng tin cậy khi cỡ mẫu lớn, do đó chúng ta phải cẩn thận khi giải thích các kết quả dựa trên các phép biến đổi khác nhau trong các mẫu nhỏ.