



Chương VI – Phương sai sai số thay đổi (*Heteroskedasticity*)



Chương VI – Phương sai sai số thay đổi



- 1. Bản chất hiện tượng phương sai sai số thay đổi**
- 2. Hậu quả**
- 3. Phát hiện**
- 4. Khắc phục**



Chương VI – Phương sai sai số thay đổi



1. Bản chất của phương sai sai số thay đổi

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

Giả thiết OLS: Phương sai của sai số ngẫu nhiên U_i là thuần nhất.

$$\text{var}(U_i) = E(U_i^2) = \sigma^2 (\forall i)$$

Giả thiết bị vi phạm → khuyết tật phương sai sai số thay đổi

$$\text{var}(U_i) = E(U_i^2) = \sigma_i^2$$

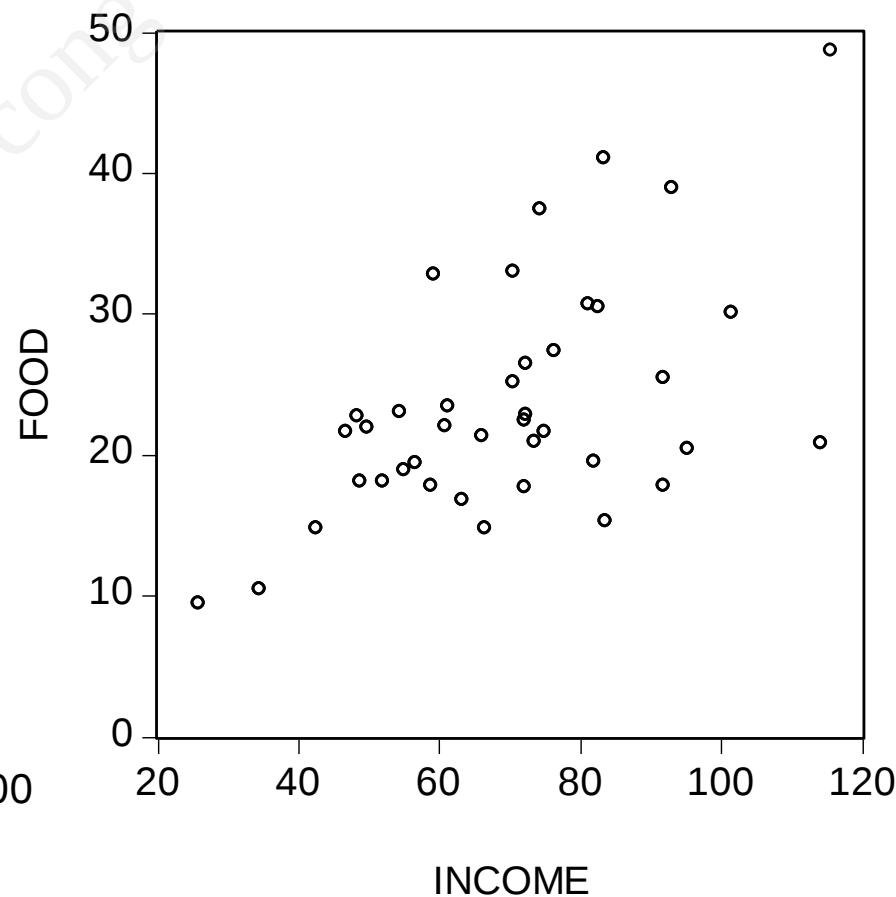
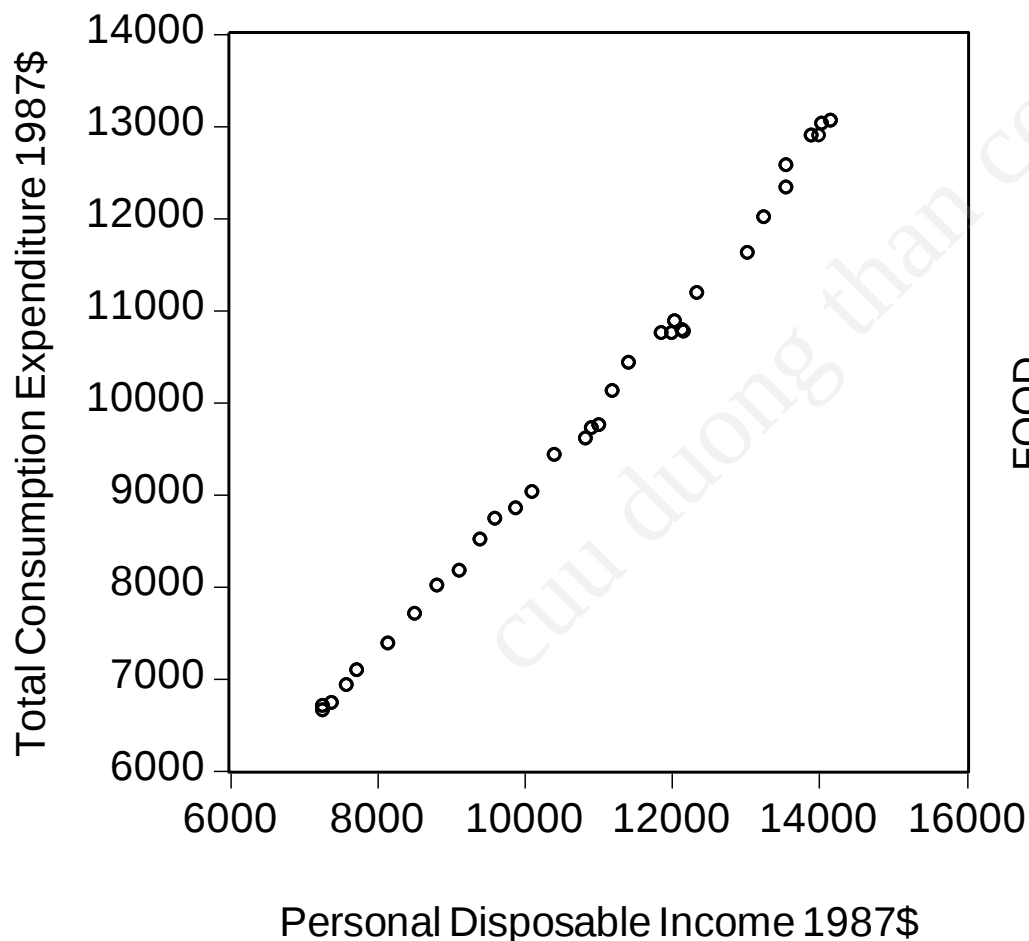
(*) Nguyên nhân:

- Do bản chất quan hệ kinh tế (**chỉ tiêu** phụ thuộc **thu nhập**)
- Do con người điều chỉnh hành vi theo thời gian
- Do kỹ thuật điều tra được cải thiện
- Do xác định dạng hàm sai
- Do sự xuất hiện của các điểm ngoại lai (**out-liers**)



Chương VI – Phương sai sai số thay đổi

1. Bản chất của phương sai sai số thay đổi





Chương VI – Phương sai sai số thay đổi



1. Bản chất của phương sai sai số thay đổi

(*) Các ước lượng OLS khi có phương sai sai số thay đổi

Xét hàm hồi qui 2 biến:

$$\text{var}(\beta_2) = \frac{\sum_{i=1}^n x_i^2 \cdot \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

sẽ khác biệt với $\text{var}(\beta_2) = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}$ trong tình huống phương sai sai số đồng đều

B mẫu là ước lượng không chệch
varBmu là ước lượng chệch



Chương VI – Phương sai sai số thay đổi



1. Bản chất của phương sai sai số thay đổi

(*) Các ước lượng OLS khi có phương sai sai số thay đổi

Xét hàm hồi qui k biến:

$$\text{cov}(\hat{\beta}) = \sigma_i^2 (X^T X)^{-1}$$

sẽ khác biệt với $\text{cov}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$ trong tình huống phương sai sai số đồng đều.

Như vậy các ước lượng OLS, $\hat{\beta}$ vẫn có thể là ước lượng tuyến tính và không chệch, nhưng không còn là ước lượng hiệu quả nữa (do chúng không còn là ước lượng tốt nhất nữa)



Chương VI – Phương sai sai số thay đổi



1. Bản chất của phương sai sai số thay đổi

(*) Cấu trúc của hiện tượng phương sai sai số thay đổi:

$$\text{var}(U_i) = E(U_i^2) = \sigma_i^2 = f(X_i)$$

(*) Phương pháp bình phương nhỏ nhất tổng quát (GLS):

Chính là phương pháp OLS áp dụng với mô hình đã đổi dạng các biến số để thỏa mãn giả thiết phương sai thuần nhất

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \frac{Y_i}{\sigma_i} = \beta_1^* \frac{1}{\sigma_i} + \beta_2^* \frac{X_i}{\sigma_i} + \frac{U_i}{\sigma_i}$$

Tiêu chuẩn ước lượng của GLS: $\sum_{i=1}^n w_i e_i^2 \rightarrow \min$

Với $w_i = \frac{1}{\sigma_i^2}$ (còn được gọi là phương pháp Weighted LS)



Chương VI – Phương sai sai số thay đổi



2. Hậu quả (*Consequences*)

$\text{Var}(\mathbf{B}_{\text{mu}}); \text{se}(\mathbf{B}_{\text{mu}})$

- Phương sai và sai số chuẩn của các ước lượng OLS khi có hiện tượng phương sai sai số thay đổi bị ước lượng chệch
- Các ước lượng OLS không còn là ước lượng tốt nhất nữa
- Điều này dẫn tới các kiểm định t và F mất ý nghĩa
- Cũng như không thể xây dựng chính xác khoảng tin cậy cho các β .
- R^2 thấp hơn thực tế
- Dự báo kém chính xác

Kết luận (*Conclusion*): nếu chúng ta vẫn muốn sử dụng các công cụ phân tích hồi qui thông thường khi mô hình có hiện tượng phương sai sai số thay đổi thì bất cứ kết luận nào được rút ra đều có thể là sai lầm.



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

(*) Các phương pháp phi chính thống (*Informal methods*)

- Dựa trên bản chất số liệu:
Phương sai sai số thay đổi $\Leftrightarrow \sigma^2$ và X_i tồn tại quan hệ
Phương sai sai số đồng đều $\Leftrightarrow \sigma^2$ và X_i không tồn tại quan hệ

Thông thường, số liệu không gian là nguyên nhân gây ra hiện tượng phương sai sai số thay đổi (do sự không đồng nhất về qui mô của các đơn vị được điều tra). Nên trong các hồi qui sử dụng số liệu chéo, khuyết tật này mang tính qui luật. Có thể sử dụng thông tin tiên nghiệm để suy đoán sự tồn tại của khuyết tật trong hồi qui.

- Phương pháp đồ thị (*Graphical method*):

Với mẫu đủ lớn, có thể vẽ đồ thị của e_i^2 theo $\hat{Y}_i$ hoặc X_i

Thay cho σ^2 –
của tổng thể

<https://sh.com/tailorsh.com>
của mẫu

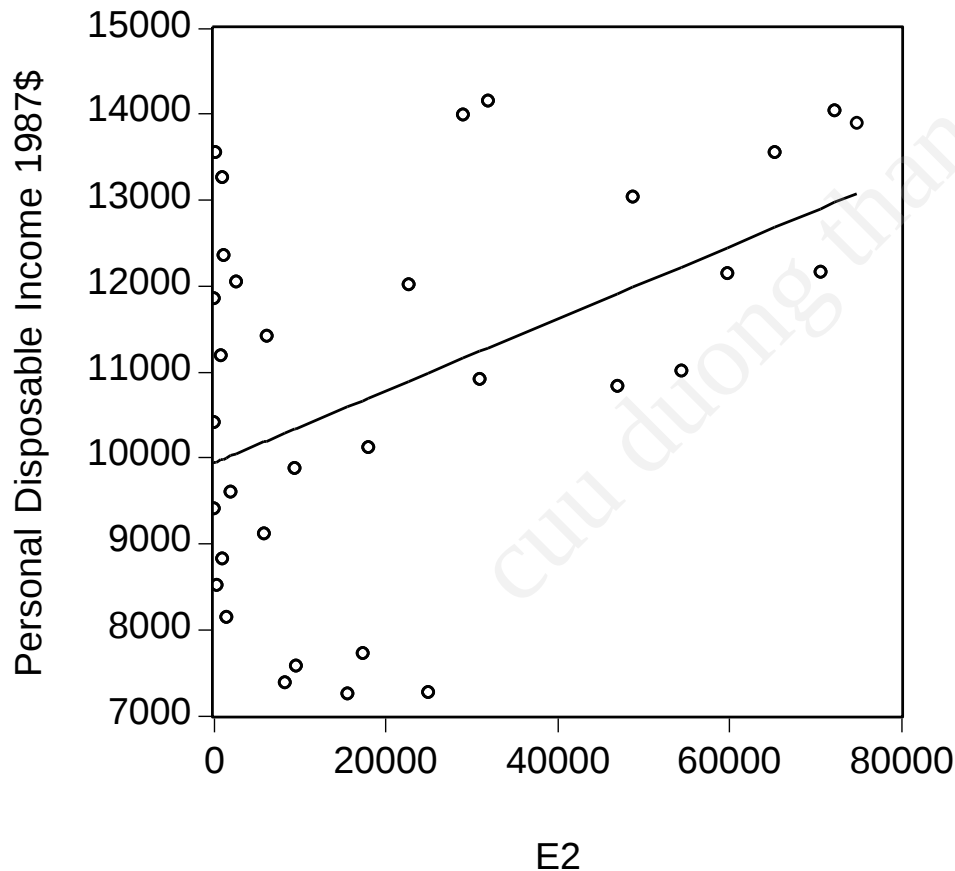


Chương VI – Phương sai sai số thay đổi

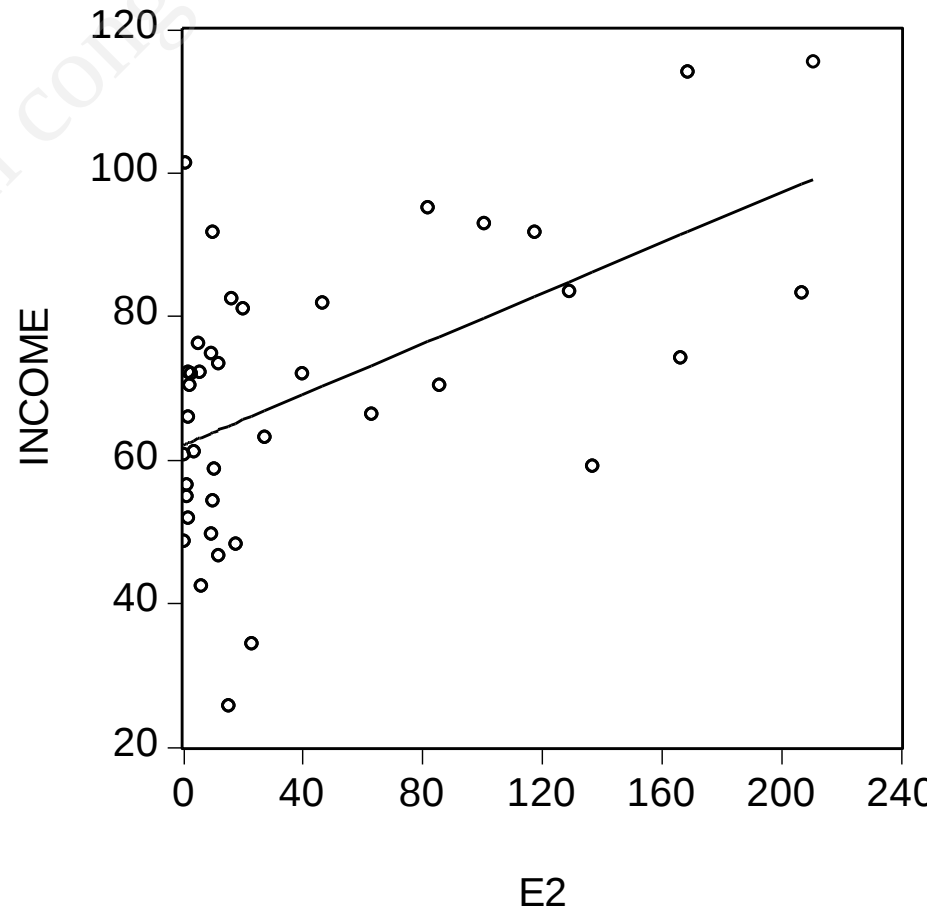
3. Phát hiện (*Detection*)



Personal Disposable Income 1987\$ vs. E2



INCOME vs. E2





Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

(*) Các phương pháp chính thống (*Formal methods*)

3.1. Kiểm định Park (R. E. Park - 1966)

$$\sigma_i^2 = \sigma^2 X_i^{m_2} e^{v_i}$$

$$\ln(\sigma_i^2) = \ln(\sigma^2) + m_2 \ln(X_i) + V_i$$

Sử dụng bình phương phần dư e_i^2 thay thế cho σ_i^2

$$\ln(e_i^2) = m_1 + m_2 \ln(X_i) + V_i \quad (P)$$

$$\begin{cases} H_0: R_P^2 = 0 \\ H_1: R_P^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0: m_2 = 0 \\ H_1: m_2 \neq 0 \end{cases} \rightarrow \begin{cases} H_0: \text{phương sai sai số đồng đều} \\ H_1: \text{phương sai sai số thay đổi} \end{cases}$$



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.1. Kiểm định Park (R. E. Park – 1966)

Dependent Variable: LOG(E2)				
Method: Least Squares				
Sample: 1 40				
Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.117862	3.819827	-1.601607	0.1175
LOG(INCOME)	2.064950	0.906545	2.277825	0.0284
R-squared	0.120136	Mean dependent var		2.560378
Adjusted R-squared	0.096982	S.D. dependent var		1.833497
Log likelihood	-77.94044	F-statistic		5.188488
Durbin-Watson stat	2.067663	Prob(F-statistic)		0.028449



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.2. Kiểm định Glejser (H. Glejser - 1969)

$$|e_i| = m_1 + m_2 X_i + V_i \quad (G)$$

$$|e_i| = m_1 + m_2 \sqrt{X_i} + V_i \quad (G)$$

$$|e_i| = m_1 + m_2 \frac{1}{X_i} + V_i \quad (G)$$

$$|e_i| = m_1 + m_2 \frac{1}{\sqrt{X_i}} + V_i \quad (G)$$

$$\begin{cases} H_0: R_G^2 = 0 \\ H_1: R_G^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0: m_2 = 0 \\ H_1: m_2 \neq 0 \end{cases} \rightarrow \begin{cases} H_0: \text{phương sai sai số đồng đều} \\ H_1: \text{phương sai sai số thay đổi} \end{cases}$$



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.2. Kiểm định Glejser (H. Glejser - 1969)

Dependent Variable: ABS(RESID)				
Method: Least Squares				
Sample: 1959 1991				
Included observations: 33				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-46.32291	73.95218	-0.626390	0.5356
Y	0.015355	0.006672	2.301376	0.0283
R-squared	0.145919	Mean dependent var		120.5058
Log likelihood	-192.0246	F-statistic		5.296331
Durbin-Watson stat	1.500807	Prob(F-statistic)		0.028255



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.3. Kiểm định White

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i \quad (1)$$

B1: Hồi qui (1) được $e_i \rightarrow e_i^2$

B2: Hồi qui:

$$(2) : e_i^2 = m_1 + m_2 X_{2i} + m_3 X_{3i} + m_4 (X_2 X_3)_i + m_5 X_{2i}^2 + m_6 X_{3i}^2 + V_i$$

$$(3) : e_i^2 = m_1 + m_2 X_{2i} + m_3 X_{3i} + m_4 X_{2i}^2 + m_5 X_{3i}^2 + V_i$$

$$\begin{cases} H_0 : R_i^2 = 0 \\ H_1 : R_i^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} \mathbf{H_0: phương sai sai số đồng đều} \\ \mathbf{H_1: phương sai sai số thay đổi} \end{cases}$$

Với $i = \overline{2,3}$



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.3. Kiểm định White:

Tiêu chuẩn kiểm định:

$$F_{qs} = \frac{R_i^2 / (m - 1)}{(1 - R_i^2) / (n - m)} = F \text{ - statistic}$$

$$W_{\alpha} = \left\{ F : F > F_{\alpha}^{(m-1, n-m)} \right\}$$

hoặc:

$$\chi_{qs}^2 = n \times R_i^2 = \text{obs} \times R \text{ - squared}$$

$$W_{\alpha} = \left\{ \chi^2 : \chi^2 > \chi_{\alpha}^2 (m - 1) \right\}$$

Với m là số hệ số của mô hình (i)



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.3. Kiểm định White:

The screenshot shows the EViews software interface. A window titled 'Equati...' is open, displaying a list of options for residual tests. The 'Residual Tests' option is selected, and a sub-menu is displayed showing various tests. The 'White Heteroskedasticity (no cross terms)' option is highlighted.

Menu Item	Sub-menu Item
Representations	
Estimation Output	
Actual, Fitted, Residual	
Gradients and Derivatives	
Covariance Matrix	
Coefficient Tests	
Residual Tests	Correlogram - Q-statistics
Stability Tests	Correlogram Squared Residuals
Label	Histogram - Normality Test
	Serial Correlation LM Test...
	ARCH LM Test...
	White Heteroskedasticity (no cross terms)
	White Heteroskedasticity (cross terms)

White Heteroskedasticity Test: no cross terms

F-statistic	6.834680	Probability	0.002437
Obs*R-squared	12.91429	Probability	0.011702

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample: 1 20

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	137157.5	81153.06	1.690108	0.1117
P	-34008.00	32297.88	-1.052948	0.3090
P^2	3168.705	3244.459	0.976651	0.3442
AD	-22455.42	4911.876	-4.571658	0.0004
AD^2	2770.563	640.0002	4.329003	0.0006
R-squared	0.645714	Mean dependent var		5959.482
Adjusted R-squared	0.551238	S.D. dependent var		6917.901
Sum squared resid	3.22E+08	Schwarz criterion		20.18160
Durbin-Watson stat	2.075225	Prob(F-statistic)		0.002437



Chương VI – Phương sai sai số thay đổi



3. Phát hiện (*Detection*)

3.4. Kiểm định dựa trên biến phụ thuộc:

$$Y_i = \beta_1 + \beta_2 X_i + U_i \quad (1)$$

B1: Hồi qui (1) thu được $e_i \rightarrow e_i^2$ **và** $\hat{Y}_i \rightarrow \hat{Y}_i^2$

B2: Hồi qui phụ:

$$(2) : e_i^2 = m_1 + m_2 \hat{Y}_i^2 + V_i$$

$$\begin{cases} H_0 : R_2^2 = 0 \\ H_1 : R_2^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0 : \text{phương sai sai số đồng đều} \\ H_1 : \text{phương sai sai số thay đổi} \end{cases}$$

$$F_{qs} = \frac{\frac{R_2^2}{(2-1)}}{\frac{(1-R_2^2)}{(n-2)}}$$

$$W_\alpha = F_{\alpha}^{(1, n-2)} : F > F_{\alpha}^{(1, n-2)}$$

$$\chi_{qs}^2 = n \times R_2^2$$

$$W_\alpha = \chi_{\alpha}^2 : \chi^2 > \chi_{\alpha}^2(1)$$



Chương VI – Phương sai sai số thay đổi



4. Khắc phục (*Remedial methods*)

4.1. Đã biết σ_i^2 :

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \frac{Y_i}{\sigma_i} = \beta_1^* \frac{1}{\sigma_i} + \beta_2^* \frac{X_i}{\sigma_i} + \frac{U_i}{\sigma_i}$$

4.2. Chưa biết σ_i^2 :

Các giả thiết về cấu trúc của σ_i^2

Giả thiết 1: $\sigma_i^2 = \sigma^2 \times X_i^2$ (kiểm định Park, White)

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \frac{Y_i}{X_i} = \beta_1^* \frac{1}{X_i} + \beta_2^* + \frac{U_i}{X_i}$$



Chương VI – Phương sai sai số thay đổi



4. Khắc phục (*Remedial methods*)

4.2. Chưa biết σ_i^2 :

Giả thiết 2: $\sigma_i^2 = \sigma^2 \times X_i$ (kiểm định Park, Glejser)

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \frac{Y_i}{\sqrt{X_i}} = \beta_1^* \frac{1}{\sqrt{X_i}} + \beta_2^* \sqrt{X_i} + \frac{U_i}{\sqrt{X_i}}$$

Giả thiết 3: $\sigma_i^2 = \sigma^2 \times [E(Y_i)]^2$ (kiểm định dựa trên biến phụ thuộc)

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \frac{Y_i}{\hat{Y}_i} = \beta_1^* \frac{1}{\hat{Y}_i} + \beta_2^* \frac{X_i}{\hat{Y}_i} + \frac{U_i}{\hat{Y}_i}$$

Giả thiết 4: Dạng hàm hồi qui bị xác định sai

$$(1) : Y_i = \beta_1 + \beta_2 X_i + U_i \rightarrow \ln(Y_i) = \beta_1^* + \beta_2^* \ln(X_i) + V_i$$