

CÁC KHUYẾT TẬT CỦA MÔ HÌNH

Việc vi phạm một trong những giả thuyết của mô hình sẽ dẫn đến những khuyết tật sau:

Giả thiết 3. Các yếu tố ngẫu nhiên có phương sai bằng nhau (phương sai thuần nhất): $Var(U_i) = Var(U|X_{2i}, X_{3i}, \dots, X_{ki}) = \sigma^2 \forall i$

-> Vi phạm: Phương sai của sai số thay đổi

Giả thiết 4. Các U_i không tương quan : $Cov(U_i, U_j) = 0 \forall i \neq j$

-> Vi phạm: Tự tương quan

Giả thiết 5. Các X_j ($j=2..k$) độc lập tuyến tính với nhau $\leftrightarrow$ $rank(X)=k$ -> Vi phạm: Đa cộng tuyến

PHƯƠNG SAI CỦA SAI SỐ THAY ĐỔI



I. NGUYÊN NHÂN VÀ HẬU QUẢ

I.1

Hiện tượng phương sai của sai số thay đổi và nguyên nhân

I.2

Hậu quả của hiện tượng phương sai của sai số thay đổi

I.1

Hiện tượng phương sai của sai số thay đổi và nguyên nhân

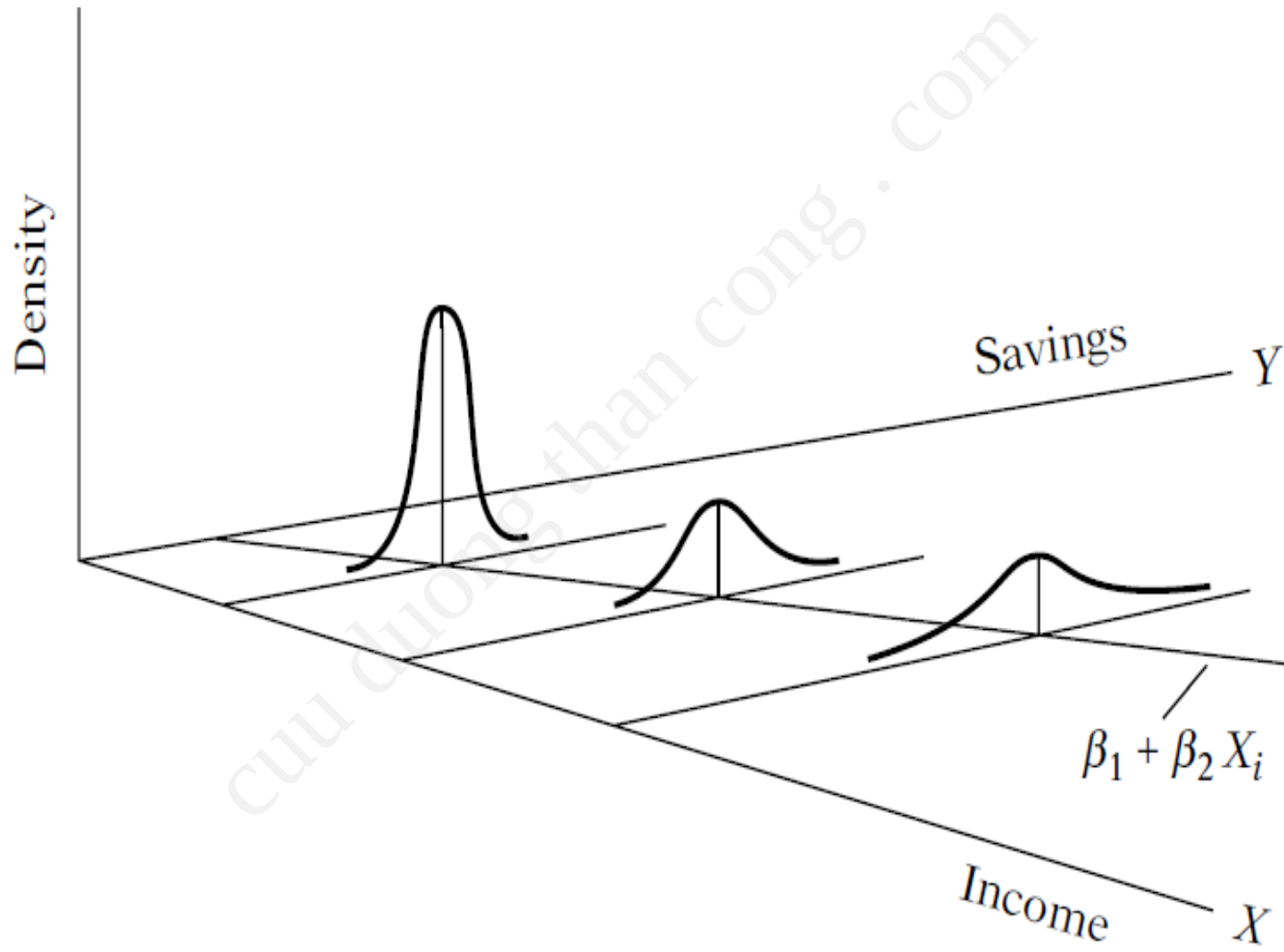
Xảy ra khi giả thuyết sau bị vi phạm:

$$\text{Var}(U_i) = \text{Var}(U|X_{2i}, X_{3i}, \dots, X_{ki}) = \sigma^2 \quad \forall i$$

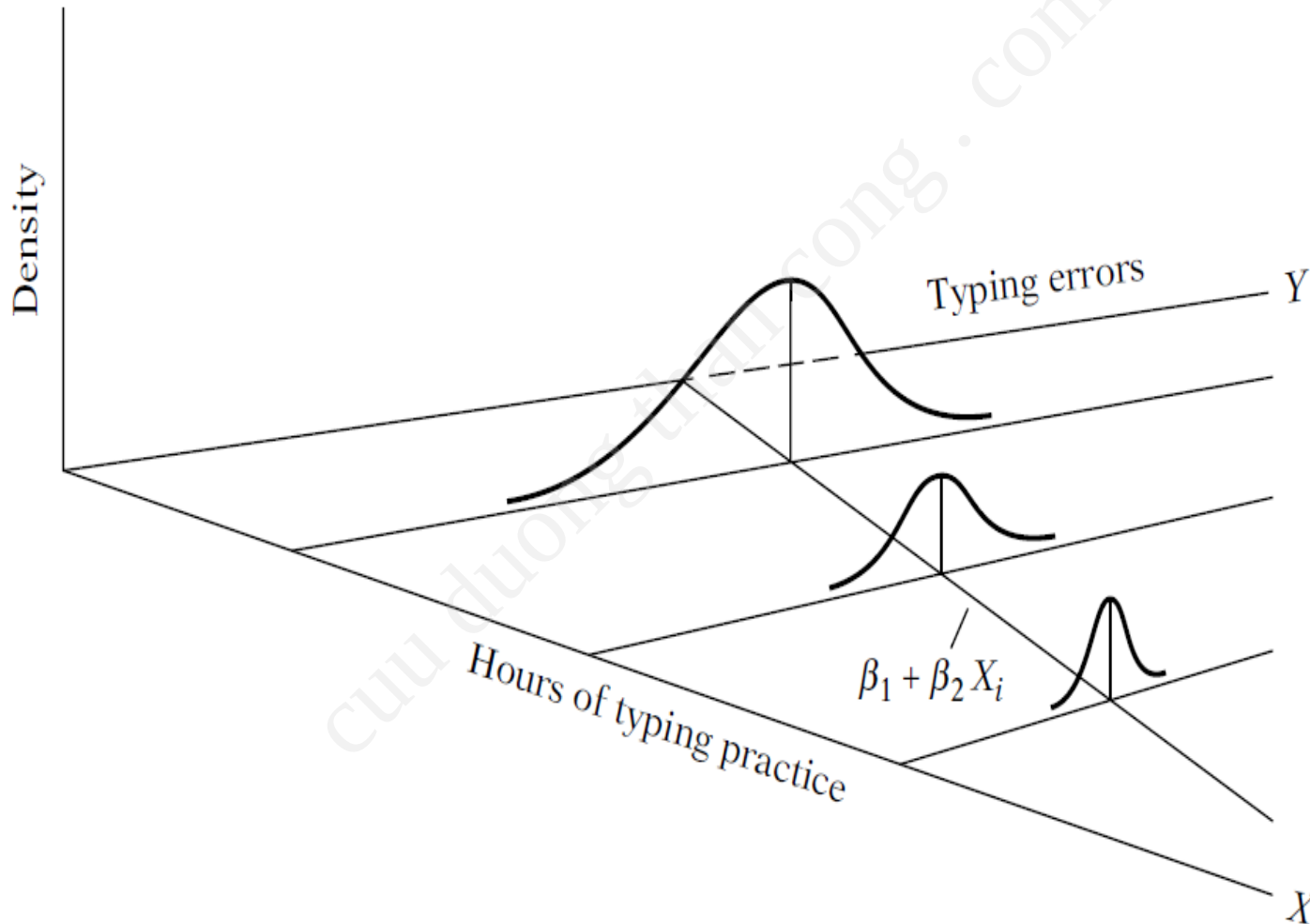
Việc vi phạm này có thể do một số nguyên nhân sau:

- Do bản chất của các mối liên hệ giữa các đại lượng.
- Do kỹ thuật thu thập và xử lý số liệu
- Do con người học được các hành vi trong quá khứ.
- Do xuất hiện các quan sát khác biệt rất nhiều (quá nhỏ hoặc quá lớn) so với các quan sát khác trong mẫu.
- Do dạng hàm sai

Mối quan hệ giữa thu nhập và tiết kiệm



Mối quan hệ giữa thời gian thực hành và lỗi đánh máy mắc phải



I.2

Hậu quả của hiện tượng phương sai của sai số thay đổi

Các ước lượng BPNN vẫn là các ước lượng tuyến tính, không chệch nhưng không còn là hiệu quả.

Các ước lượng của các phương sai sẽ là các ước lượng chệch, thống kê T và F không còn có ý nghĩa. Do đó khoảng tin cậy và các kiểm định dựa trên thống kê T và F không còn đáng tin cậy nữa.



II. PHÁT HIỆN KHUYẾT TẬT

II.1 Phương pháp đồ thị

II.2 Kiểm định Park

II.3 Kiểm định Glejser

II.4 Kiểm định Goldfield – Quant (GQ)

II. PHÁT HIỆN KHUYẾT TẬT

II.5 Kiểm định Breusch – Pagan – Godfrey ([BPG](#))

II.6 Kiểm định [White](#)

II.7 Kiểm định tương quan hạng [Spearman](#)

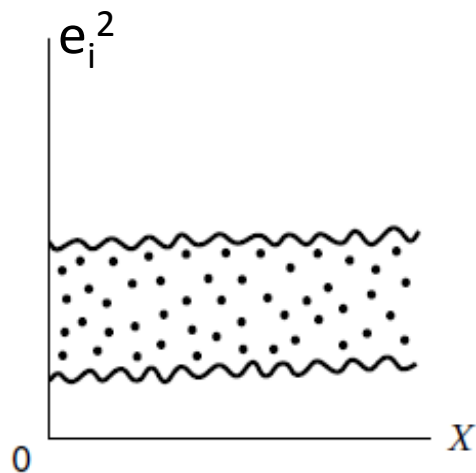
II.8 Kiểm định dựa trên [biến phụ thuộc](#)

II.1

Phương pháp đồ thị

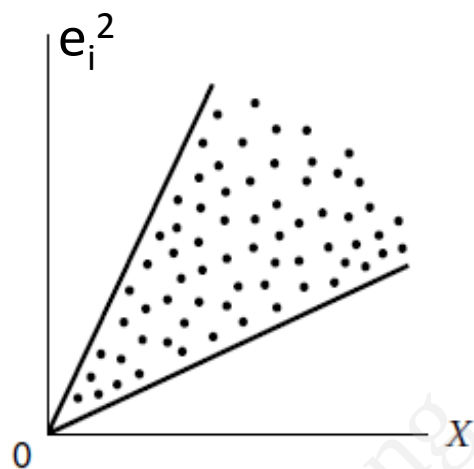
Vì phần dư e_i của hàm hồi quy mẫu chính là ước lượng của sai số ngẫu nhiên U_i nên dựa vào đồ thị e_i (hoặc e_i^2) đối với biến giải thích X ta có kết luận:

Nếu độ rộng của e_i (hay e_i^2) tăng hay giảm khi X tăng thì có thể nghi ngờ phương sai của sai số thay đổi. Trong trường hợp nhiều hơn 1 biến giải thích, có thể dùng đồ thị e_i (hoặc e_i^2) đối với $\hat{Y}_i$



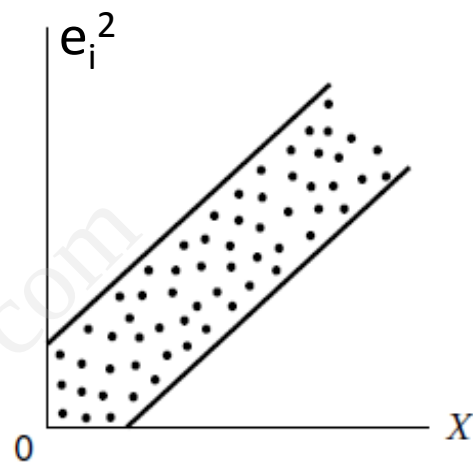
(a)

Không



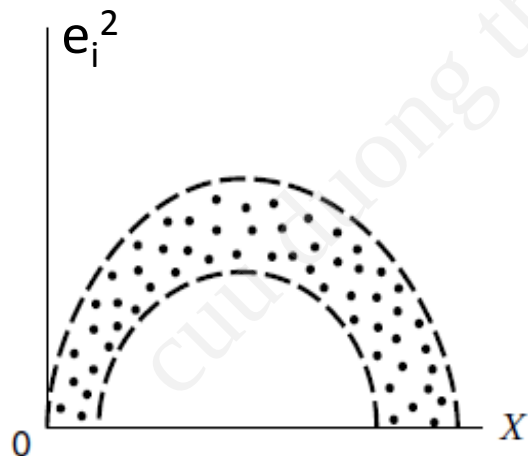
(b)

Có



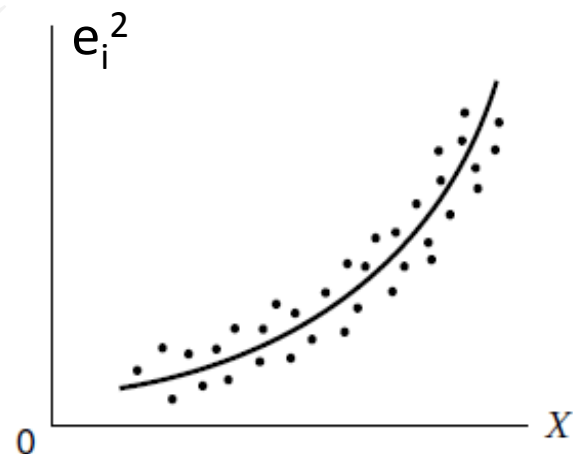
(c)

Có



(d)

Có



(e)

Có



II.2

Kiểm định Park

Giả thiết: $\sigma_i^2 = \sigma^2 X_i^{\beta_2} e^{v_i} \Leftrightarrow \ln \sigma_i^2 = \ln \sigma^2 + \beta_2 \ln X_i + v_i$

Vì thường σ_i^2 chưa biết nên thay thế bởi ước lượng của nó là e_i^2 . Khi đó: $\ln e_i^2 = \ln \sigma^2 + \beta_2 \ln X_i + v_i$

Bước 1: Ước lượng hồi quy gốc để thu được các phần dư e_i

Bước 2: Ước lượng hồi quy $\ln e_i^2 = \beta_1 + \beta_2 \ln X_i + v_i$

Nếu có nhiều biến giải thích thì ước lượng hồi quy này với từng biến giải thích hoặc với $\hat{Y}_i$

Bước 3: Kiểm định giả thuyết $H_0: \beta_2 = 0$. Nếu H_0 bị bác bỏ thì kết luận phương sai của sai số thay đổi.



II.3 Kiểm định Glejser

Giống cách làm của Park, tuy nhiên Glejser thay đổi một số dạng hàm liên hệ giữa biến giải thích X_i và $|e_i|$ như sau:

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

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

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

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

Nếu giả thuyết $H_0: \beta_2 = 0$ bị bác bỏ thì kết luận có phương sai của sai số thay đổi.

Điều kiện: cỡ mẫu lớn.



II.4 Kiểm định Goldfield – Quant ($G - Q$)

Bước 1. Sắp xếp các giá trị quan sát theo chiều tăng của biến X_i

Bước 2. Bỏ c quan sát ở giữa theo quy tắc:

Nếu $n \approx 30$: lấy $c = 4$ hoặc 6 Nếu $n \approx 60$: lấy $c = 10$

Các quan sát còn lại chia 2 nhóm, mỗi nhóm có $(n-c)/2$ quan sát

Bước 3: Ước lượng mô hình với $(n-c)/2$ quan sát đầu và cuối thu được RSS_1 và RSS_2 tương ứng với bậc tự do là:

$$d = \frac{n - c}{2} - k = \frac{n - c - 2k}{2}$$

Bước 4: Xây dựng TCKĐ: $F = \frac{RSS_2}{d} / \frac{RSS_1}{d}$

Nếu giả thuyết H_0 : phương sai của sai số ngẫu nhiên không đổi được thỏa mãn thì $F \sim F(d, d)$.

Miền bác bỏ H_0 : $F > F_\alpha(d, d)$

Chú ý: Trường hợp có nhiều hơn 1 biến giải thích, nếu không chắc biến nào thực sự thích hợp để sắp xếp ở bước 1, chúng ta có thể tiến hành kiểm định GQ với từng biến giải thích rồi chọn biến phù hợp nhất (hệ số $\neq 0$ và có r^2 lớn hơn).

Điều kiện: U_i có phân phối chuẩn.



II.5

Kiểm định Breusch – Pagan – Godfrey (BPG)

Bước 1: Ước lượng hồi quy gốc để thu được các phần dư e_i

Bước 2: Tính $\tilde{\sigma}^2 = \sum e_i^2 / n$

Bước 3: Xây dựng biến $p_i = e_i^2 / \tilde{\sigma}^2$

Bước 4: Ước lượng hồi quy $p_i = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi} + v_i$
trong đó v_i là nhiễu ngẫu nhiên; $Z_{2i}, \dots, Z_{mi}$ là một hoặc một số (thậm chí tất cả) biến X_i có ảnh hưởng đến σ_i^2

Bước 5: Thu được ESS của bước 4 và tính $\theta = ESS/2$

Bước 6: Kiểm định giả thuyết $H_0: \alpha_2 = \alpha_3 = \dots = \alpha_m = 0$. Nếu $\theta > \chi^2(m-1)$ thì có hiện tượng phương sai của sai số thay đổi.

Điều kiện: U_i có phân phối chuẩn và mẫu lớn.



II.6 Kiểm định White

Xét mô hình 3 biến (trường hợp k biến tương tự):

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

Bước 1: Ước lượng hồi quy trên để thu được các phần dư e_i

Bước 2: Ước lượng mô hình:

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

Mô hình này có thể có số mũ cao hơn (trường hợp này số mũ = 2), có hệ số xác định bội R^2 và có số hệ số hồi quy riêng kí hiệu là df (trường hợp này df = 5).

Bước 3: Kiểm định giả thuyết H_0 : phương sai thuần nhất.

Nếu $nR^2 > \chi^2_{\alpha}(df)$ thì bác bỏ H_0 , nghĩa là phương sai của sai số thay đổi.

Ưu điểm: không đòi hỏi U_i có phân phối chuẩn.

Nhược điểm: dùng để phát hiện phương sai của sai số thay đổi hoặc phát hiện sai lầm định dạng, từ kết quả kiểm định, không thể nói cụ thể hiện tượng nào đã xảy ra.

Eview: White Heteroskedasticity

no cross terms: không có hạng tử $X_i X_j$ ($i \neq j$)

cross terms: có hạng tử $X_i X_j$ ($i \neq j$)



II.7

Kiểm định tương quan hạng Spearman

Ứng với một biến nào đó, người ta tiến hành sắp xếp theo chiều giảm dần các biểu hiện của nó (trên các quan sát). Hạng của một quan sát ứng với biến đó chính là số thứ tự của quan sát.

Kí hiệu d_i là chênh lệch về hạng của quan sát thứ i ứng với 2 biến giải thích nào đó.

Hệ số tương quan hạng Spearman r_s được xác định:

$$r_s = 1 - 6 \frac{\sum d_i^2}{n(n^2 - 1)}$$

Bước 1: Ước lượng hồi quy gốc để thu được các phần dư e_i

Bước 2: Xếp hạng $|e_i|$ và X_i (nào đó) theo chiều giảm dần, tính $d_i = \text{hạng } |e_i| - \text{hạng } X_i$ từ đó suy ra r_s

Bước 3: Kiểm định giả thuyết H_0 : phương sai thuần nhất.

Tiêu chuẩn kiểm định: $t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$

Miền bác bỏ H_0 : $t > t_{\alpha}(n-2)$: phương sai của sai số thay đổi

Nếu có nhiều biến giải thích thì ta tiến hành các bước kiểm định trên với từng biến giải thích hoặc với $\hat{Y}_i$.



II.8

Kiểm định dựa trên biến phụ thuộc

Ý tưởng: nếu phương sai của sai số thay đổi:

$$\sigma_i^2 = \alpha_1 + \alpha_2 (E(Y_i))^2$$

Vì cả σ_i^2 và $E(Y_i)$ đều chưa biết nên ta dùng các ước lượng của nó là $e_i^2; \hat{Y}_i$

Bước 1: Ước lượng hồi quy gốc thu được $e_i^2; \hat{Y}_i$

Bước 2: Ước lượng hồi quy: $e_i^2 = \alpha_1 + \alpha_2 \hat{Y}_i^2 + U_i$ thu được R^2 .

Bước 3: Kiểm định giả thuyết H_0 : phương sai thuần nhất.

- Kiểm định χ^2 : nếu $nR^2 > \chi_\alpha^2(1)$ thì bác bỏ H_0

- Kiểm định F: tiêu chuẩn kiểm định: $F = \left(\frac{\hat{\alpha}_2}{se(\hat{\alpha}_2)} \right)^2$

Nếu $F > F_\alpha(1, n-2)$ thì bác bỏ H_0



III. BIỆN PHÁP KHẮC PHỤC

Ý tưởng: thay vì thực hiện hồi quy trên mô hình gốc thỏa mãn các giả thuyết của phương pháp OLS trừ giả thuyết phương sai thuần nhất, ta biến đổi mô hình gốc về mô hình thỏa mãn tất cả các giả thuyết của phương pháp OLS

III.1 Khi $E(U_i^2)$ đã biết

III.2 Khi $E(U_i^2)$ chưa biết

III.1 Khi $E(U_i^2)$ đã biết

Phép biến đổi: chia cả 2 vế của mô hình gốc cho $\sigma_i = \sqrt{E(U_i^2)}$

Thay vì đi ước lượng mô hình thông thường:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + U_i$$

Ta ước lượng mô hình: $\frac{Y_i}{\sigma_i} = \beta_1 \cdot \frac{1}{\sigma_i} + \beta_2 \cdot \frac{X_{2i}}{\sigma_i} + \dots + \beta_k \cdot \frac{X_{ki}}{\sigma_i} + \frac{U_i}{\sigma_i}$

$$\Leftrightarrow Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \dots + \beta_k X_{ki}^* + U_i^*$$

Khi đó:

$$E(U_i^*) = 0; \text{Var}(U_i^*) = \text{Var}\left(\frac{U_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2} \text{Var}(U_i) = \frac{1}{\sigma_i^2} \cdot \sigma_i^2 = 1$$

-> Mô hình mới thỏa mãn các điều kiện của OLS



III.2 Khi $E(U_i^2)$ chưa biết

Xét mô hình hồi quy gốc: $Y_i = \beta_1 + \beta_2 X_i + U_i$



Phương sai của sai số tỉ lệ với bình phương biến giải thích $E(U_i^2) = \sigma^2 X_i^2$



Phương sai của sai số tỉ lệ với biến giải thích $E(U_i^2) = \sigma^2 X_i$



Phương sai của sai số tỉ lệ với bình phương kỳ vọng của biến phụ thuộc $E(U_i^2) = \sigma^2 (E(Y_i))^2$



Biến đổi Logarit



$$E(U_i^2) = \sigma^2 X_i^2$$

* Cách nhận biết mối liên hệ:

- Phương pháp đồ thị: đồ thị giữa e_i^2 và X_i dạng Parabol
- Hồi quy e_i^2 theo X_i^2 (không có hệ số chặn). Nếu kết quả kiểm định hệ số của X_i^2 bằng 0 bị bác bỏ thì tồn tại mối liên hệ.

* Phép biến đổi: chia 2 vế của mô hình gốc cho $X_i (\neq 0)$

Ước lượng mô hình: $\frac{Y_i}{X_i} = \beta_1 \cdot \frac{1}{X_i} + \beta_2 + \frac{U_i}{X_i} \Leftrightarrow Y_i^* = \beta_2 + \beta_1 X_i^* + U_i^*$

Rõ ràng:

$$E(U_i^*) = E\left(\frac{U_i}{X_i}\right) = \frac{E(U_i)}{X_i} = 0;$$

$$Var(U_i^*) = Var\left(\frac{U_i}{X_i}\right) = \frac{1}{X_i^2} Var(U_i) = \frac{1}{X_i^2} \cdot \sigma^2 X_i^2 = \sigma^2$$

-> Mô hình mới
thỏa mãn các điều
kiện của OLS





$$E(U_i^2) = \sigma^2 X_i$$

* Cách nhận biết mối liên hệ:

- Phương pháp đồ thị: đồ thị giữa e_i^2 và X_i dạng tuyến tính
- Hồi quy e_i^2 theo X_i (không có hệ số chặn). Nếu kết quả kiểm định hệ số của X_i bằng 0 bị bác bỏ thì tồn tại mối liên hệ.

* Phép biến đổi: chia 2 vế của mô hình gốc cho $\sqrt{X_i}$ ($X_i > 0$)

Ước lượng mô hình: $\frac{Y_i}{\sqrt{X_i}} = \beta_1 \cdot \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + \frac{U_i}{\sqrt{X_i}} \Leftrightarrow Y_i^* = \beta_1 X_{0i}^* + \beta_2 X_i^* + U_i^*$

Rõ ràng: $E(U_i^*) = E\left(\frac{U_i}{\sqrt{X_i}}\right) = \frac{E(U_i)}{\sqrt{X_i}} = 0;$ $\rightarrow$ Mô hình mới thỏa mãn các điều kiện của OLS

$$Var(U_i^*) = Var\left(\frac{U_i}{\sqrt{X_i}}\right) = \frac{1}{X_i} Var(U_i) = \frac{1}{X_i} \cdot \sigma^2 X_i = \sigma^2$$





$$E(U_i^2) = \sigma^2 (E(Y_i))^2$$

* Cách nhận biết mối liên hệ:

- Phương pháp đồ thị: đồ thị giữa e_i^2 và $\hat{Y}_i$ dạng tuyến tính
- Hồi quy e_i^2 theo $\hat{Y}_i^2$ (không có hệ số chặn). Nếu kết quả kiểm định hệ số của $\hat{Y}_i^2$ bằng 0 bị bác bỏ thì tồn tại mối liên hệ.

* Phép biến đổi: chia 2 vế của mô hình gốc cho $\hat{Y}_i$

Ước lượng mô hình:
$$\frac{Y_i}{\hat{Y}_i} = \beta_1 \cdot \frac{1}{\hat{Y}_i} + \beta_2 + \frac{U_i}{\hat{Y}_i} \Leftrightarrow Y_i^* = \beta_2 + \beta_1 X_i^* + U_i^*$$

Rõ ràng:

$$E(U_i^*) = E\left(\frac{U_i}{E(Y_i)}\right) = \frac{E(U_i)}{E(Y_i)} = 0;$$

-> Mô hình mới thỏa mãn các điều kiện của OLS

$$Var(U_i^*) = Var\left(\frac{U_i}{E(Y_i)}\right) = \frac{1}{(E(Y_i))^2} Var(U_i) = \frac{1}{(E(Y_i))^2} \cdot \sigma^2 (E(Y_i))^2 = \sigma^2$$

Nhận xét

Cả 4 trường hợp trên đều là các trường hợp riêng của trường hợp tổng quát: $E(U_i^2) = \sigma^2 Z_i^2$ với Z_i là một biến nào đó.

Phép đổi biến: chia cả 2 vế của mô hình gốc cho Z_i

$$\frac{Y_i}{Z_i} = \beta_1 \cdot \frac{1}{Z_i} + \beta_2 \frac{X_i}{Z_i} + \frac{U_i}{Z_i} \Leftrightarrow Y_i^* = \beta_1 Z_i^* + \beta_2 X_i^* + U_i^*$$

Khi đó mô hình mới thỏa mãn các điều kiện OLS. Các ước lượng OLS thu được bằng cách cực tiểu hàm:

$$\sum e_i^{*2} = \sum (Y_i^* - \hat{\beta}_1^* Z_i^* - \hat{\beta}_2^* X_i^*)^2 = \sum \frac{1}{Z_i^2} (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2$$

-> Phương pháp bình phương nhỏ nhất có trọng số

Khi $Z_i = 1$, ta trở lại các ước lượng OLS thông thường. Khi đó:

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





Biến đổi Logarit

Đôi khi thay vì ước lượng hồi quy gốc, người ta ước lượng hồi quy: $\ln Y_i = \beta_1 + \beta_2 \ln X_i + V_i$

Phép biến đổi này làm giảm bớt hiện tượng phương sai của sai số thay đổi so với hồi quy gốc $Y_i = \beta_1 + \beta_2 X_i + U_i$

Chú ý

- Lựa chọn kĩ biến dùng để biến đổi (mô hình nhiều biến)
- Phép biến đổi log không dùng khi X hoặc Y không dương
- Có thể xảy ra tình trạng bản thân các biến trong mô hình gốc không tương quan nhưng các biến trong mô hình đã biến đổi lại tương quan.
- Khi $\sigma_i^2 = E(U_i^2)$ chưa biết, nó được ước lượng từ một trong các cách biến đổi trên, mọi kiểm định t, F mà chúng ta sử dụng chỉ có hiệu lực trên mẫu lớn. Cần cẩn thận khi giải thích các kết quả dựa trên các phép biến đổi khác nhau trong các mẫu nhỏ.