

Phương sai thay đổi và tự tương quan (Heteroskedasticity and Autocorrelation)

Lê Việt Phú
Trường Chính sách Công và Quản lý Fulbright

Ngày 4 tháng 1 năm 2018

Bổ sung: Kiểm định Lagrange Multiplier (LM) trong hồi quy bội

- ▶ Dùng trong kiểm định với mẫu lớn, ngay cả khi điều kiện phân phối chuẩn không được thỏa, do đó kiểm định này còn được gọi là kiểm định tiệm cận.
- ▶ Giả sử ta muốn kiểm định $H_0 : \beta_1 = \beta_2 = 0$ trong mô hình:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

- ▶ Kiểm định đơn giản nhất là kiểm định F, với điều kiện mô hình thỏa CLRM.
- ▶ Với cỡ mẫu lớn, đặc tính tiệm cận (asymptotic) cho phép nối lỏng điều kiện phân phối chuẩn của $u \Rightarrow$ Kiểm định LM.

Các bước tiến hành kiểm định LM

- ▶ Hồi quy y với mô hình ràng buộc (thiếu các biến x_1, x_2). Ước lượng phần dư \hat{u} .
- ▶ Hồi quy phần dư \hat{u} đối với tất cả các biến giải thích (gọi là hồi quy phụ - auxiliary regression):

$$\hat{u} = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + v$$

- ▶ Trị kiểm định được tính từ R^2 của mô hình hồi quy phụ:
 $LM = nR^2 \sim \chi_q^2$ với q là số ràng buộc.
- ▶ Bác bỏ $H_0 \Rightarrow$ mô hình ban đầu bị **thiếu biến quan trọng** (x_1 và x_2).

Ví dụ: so sánh kiểm định F và kiểm định LM

Giả sử ta muốn kiểm định $H_0 : \beta_2 = \beta_3 = 0$

$$\log(\text{income}) = \beta_0 + \beta_1 \text{yoeduc} + \beta_2 \text{yoexper} + \beta_3 \text{yoexpersq} + \beta_4 \text{married} \\ + \beta_5 \text{school} + \beta_6 \text{public} + \beta_7 \text{foreign} + \beta_8 \text{official} + u$$

- ▶ $F = 310.83$, p-value = 0.
- ▶ $LM = 575.004$, p-value = 0.

Các giả định của mô hình MLR

- ▶ $E(u|X) = 0 \Rightarrow$ Ước lượng OLS là không chệch
- ▶ $Var(u|X) = \sigma^2$ (homoskedasticity) \Rightarrow Ước lượng OLS là BLUE
- ▶ Sai số u độc lập với các biến giải thích, có phân phối chuẩn với giá trị trung bình là 0 và phương sai σ^2 (independent, identically distributed - *iid*):

$$u \sim N(0, \sigma^2)$$

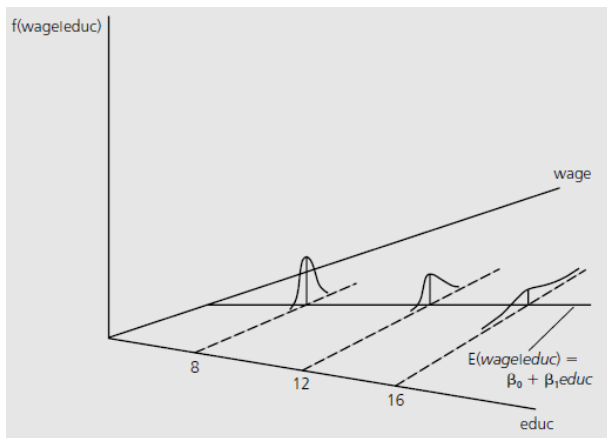
\Rightarrow

$$\hat{\beta} \sim N(\beta, Var(\beta))$$

Phương sai của sai số thay đổi (heteroskedasticity)

- ▶ Vi phạm điều kiện *iid*: $Var(u|X) \neq \sigma^2$
- ▶ Ước lượng OLS vẫn là không chệch, nhưng không còn là hiệu quả nhất do sai số của $\hat{\beta}$ không còn là nhỏ nhất.
- ▶ Các kiểm định dựa trên phân phối $\hat{\beta}$ sai do sai số của $\hat{\beta}$ bị sai (t-test, F-test).

Phương sai thay đổi xảy ra khi nào?



- ▶ Phương sai của sai số tương quan với biến khác, ví dụ số năm học nhiều thì mức độ dao động của thu nhập càng lớn.
- ▶ Do tương quan chuỗi hoặc tương quan không gian.

- ▶ Tương quan chuỗi (auto-correlation): Xảy ra với các chuỗi dữ liệu theo thời gian. Ví dụ với chuỗi thời gian tương quan bậc nhất (Autoregressive AR(1)):

$$y_t = \beta_0 + \beta_1 x_t^1 + \beta_2 x_t^2 + \dots + \epsilon_t$$
$$\epsilon_t = \rho \epsilon_{t-1} + v_t$$

- ▶ Tương quan không gian (spatial correlation): Xảy ra với dữ liệu địa lý. Các quan sát có vị trí cận kề thường có tương quan lẫn nhau.

$$y_j = \rho \sum_{i=1}^n g(d_i) y_i + \beta x + \epsilon$$

với $g(d)$ là hàm khoảng cách địa lý của quan sát j với các còn lại. $g(d)$ giảm khi khoảng cách d tăng lên.

Kiểm định hiện tượng phương sai thay đổi

- ▶ Kiểm định Bresch-Pagan về phụ thuộc tuyến tính giữa phương sai của sai số và các biến giải thích.
- ▶ Kiểm định White trong trường hợp tổng quát.

Kiểm định Breusch-Pagan

Ước lượng mô hình với giả định $E(u|X) = 0$ và $cov(u, X) = 0$ thỏa. Ước lượng OLS vẫn không chệch và nhất quán. Chúng ta muốn kiểm định nếu phương sai của sai số không đổi.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad (1)$$

$$H_0 : Var(u|X) = E(u^2) = \sigma^2$$

Do $E(u|X) = 0$ nên $Var(u|X) = E(u^2) - [E(u)]^2 = E(u^2)$

Các bước thực hiện kiểm định Breusch-Pagan (BP)

1. Ước lượng mô hình (1) thông thường, tính giá trị của phần dư \hat{u}
2. Tạo biến phụ thuộc là bình phương của phần dư, \hat{u}^2
3. Hồi quy \hat{u}^2 theo tất cả các biến giải thích trong mô hình hồi quy phụ (auxiliary regression):

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k + v \quad (2)$$

4. Kiểm định nếu $\delta_1, \dots, \delta_k$ đồng thời bằng 0 trong mô hình (2) bằng F-test hoặc LM-test. Trị kiểm định được tính từ R_a^2 của mô hình phụ:

$$F = \frac{R_a^2/k}{(1 - R_a^2)/(n - k - 1)} \sim F_{k, n-k-1}$$

$$LM = nR_a^2 \sim \chi_k^2$$

5. Nếu bác bỏ H_0 chứng tỏ mô hình có hiện tượng phương sai thay đổi.

Ví dụ kiểm định BP

- ▶ Ước lượng lại mô hình tỷ suất thu nhập từ bộ dữ liệu VHLSS 2010.
- ▶ Kiểm định phương sai thay đổi thủ công thông qua F và LM statistic.
- ▶ Thực hiện tự động bằng Stata.

Kiểm định White trong trường hợp tổng quát

3. Tương tự như kiểm định Breusch-Pagan ở bước 1-2, nhưng ở bước 3 giả định cấu trúc hàm của phần dư linh hoạt hơn bằng cách thêm bình phương và tương tác giữa các biến giải thích:

$$\begin{aligned}\hat{u}^2 = & \delta_0 + \delta_1 x_1 + \dots + \delta_k x_k \\ & + \sum_{i=1}^K \delta_i x_i^2 + \sum_{i=1}^K \sum_{j=1}^K \delta_{ij} x_i x_j + v\end{aligned}$$

4. Kiểm định bằng F-test hoặc LM-test nếu tất cả các tham số δ (loại trừ tung độ gốc δ_0) trong hồi quy phụ bằng 0.

Kiểm định White - thực hiện đơn giản

5. Cách [4] sẽ làm giảm số bậc tự do trong mô hình, ví dụ mô hình có 3 biến tự do sẽ có tổng cộng là 9 ràng buộc. Một hình thức khác có thể kiểm định bằng cách:
- ▶ Tính \hat{u} và \hat{y} từ bước 1.
 - ▶ Hồi quy \hat{u}^2 lên biến \hat{y} và \hat{y}^2 trong mô hình phụ:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v$$

- ▶ Kiểm định $\delta_1 = \delta_2 = 0$ bằng F-test hoặc LM test với 2 ràng buộc.

Ví dụ kiểm định White

- ▶ Ước lượng lại mô hình tỷ suất thu nhập từ bộ dữ liệu VHLSS 2010.
- ▶ Kiểm định phương sai thay đổi thủ công thông qua F và LM statistic.
- ▶ Thực hiện tự động bằng Stata.

Ước lượng mô hình với phương sai thay đổi

Sai số chuẩn của tham số được điều chỉnh:

$$\widehat{Var}(\hat{\beta}) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{SSR_j^2}$$

Được gọi là:

- ▶ Heteroskedasticity-robust standard errors.
- ▶ White-Huber standard errors.
- ▶ Robust standard errors.

So sánh mô hình γ suất thu nhập có và không điều chỉnh phương sai thay đổi

Regression Results

	Homo. b/se	Robust b/se
yoeduc	0.0926*** (0.0027)	0.0926*** (0.0031)
yoexper	0.0617*** (0.0025)	0.0617*** (0.0032)
yoexpersq	-0.0012*** (0.0000)	-0.0012*** (0.0001)
married	0.0352 (0.0221)	0.0352 (0.0217)
publicSchool	-0.1146** (0.0424)	-0.1146* (0.0465)
public	-0.1043** (0.0329)	-0.1043* (0.0423)
foreign	0.4499*** (0.0364)	0.4499*** (0.0328)
official	0.2705*** (0.0359)	0.2705*** (0.0430)
Constant	8.4936*** (0.0475)	8.4936*** (0.0539)
Obs	7552.0000	7552.0000
R2	0.3026	0.3026
R2-adj	0.3019	0.3019
df(r)	7543.0000	7543.0000
SSR	4040.8653	4040.8653

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Kiểm định giả thuyết bội đối với trường hợp phương sai thay đổi

- ▶ Heteroskedasticity-robust F-test: Stata.
- ▶ Heteroskedasticity-robust LM test.

Giả sử ta muốn kiểm định $H_0 : \beta_1 = \beta_2 = 0$ trong mô hình sau:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u$$

Do phương sai thay đổi, kiểm định LM-robust phức tạp hơn so với kiểm định LM thông thường.

Các bước thực hiện LM-robust test

1. Hồi quy mô hình ràng buộc (loại trừ biến x_1, x_2), ước lượng phần dư \hat{u} .
2. Hồi quy từng biến loại trừ (x_1, x_2) lên tất cả các biến độc lập sử dụng tại bước 1, và ước lượng phần dư \hat{r}_1, \hat{r}_2 .
3. Tạo biến tương tác giữa \hat{r}_1, \hat{r}_2 với \hat{u} .
4. Ước lượng mô hình hồi quy không tung độ gốc của biến phụ thuộc $z = 1$ lên các biến giải thích là các biến tương tác $\hat{r}_1\hat{u}, \hat{r}_2\hat{u}$.
5. Tính trị kiểm định từ SSR^* của hồi quy ở bước 4,

$$LM = n - SSR^* \sim \chi_2^2$$

6. Nếu bác bỏ H_0 có nghĩa là $\beta_1 \neq 0$ hoặc $\beta_2 \neq 0$ hoặc đồng thời cả hai.

Ví dụ kiểm định F- và LM-test khi có phương sai thay đổi

Kiểm định nếu số năm kinh nghiệm và số năm kinh nghiệm bình phương đồng thời bằng không.

$$\log(\text{income}) = \beta_0 + \beta_1 \text{yoeduc} + \beta_2 \text{yoexper} + \beta_3 \text{yoexpersq} + \beta_4 \text{married} \\ + \beta_5 \text{school} + \beta_6 \text{public} + \beta_7 \text{foreign} + \beta_8 \text{official} + u$$

- ▶ Do phương sai thay đổi, kỳ vọng F- và LM-test sẽ khác so với kết quả ở slide 4.
- ▶ $F = 185.97$, $p\text{-value} = 0$.
- ▶ $LM = 295.850$, $p\text{-value} = 0$.

Phương sai biết cấu trúc hàm - Phương pháp WLS

WLS - Weighted Least Square: Hồi quy bình phương tối thiểu có trọng số.

- ▶ Giả định phương sai của sai số là một hàm số của x :

$$\text{Var}(u|x) = \sigma^2 h(x)$$

- ▶ Có thể thực hiện chuyển đổi dữ liệu trước khi ước lượng:

$$\frac{y}{\sqrt{h(x)}} = \beta_0 + \beta_1 \frac{x_1}{\sqrt{h(x)}} + \beta_2 \frac{x_2}{\sqrt{h(x)}} + \dots + \frac{u}{\sqrt{h(x)}}$$

- ▶ Ước lượng mô hình hồi quy dựa trên các biến đã chuyển đổi là BLUE.

Phương sai không biết cấu trúc hàm - FGLS

FGLS: Feasible Generalized Least Square - Bình phương tối thiểu tổng quát khả thi.

- ▶ Thông thường giả định phương sai của sai số là hàm mũ của x , nhưng không biết cấu trúc hàm:

$$\text{Var}(u|x) = \sigma^2 e^{\delta_0 + \delta_1 x_1 + \dots + \delta_k x_k}$$

- ▶ Phương pháp FGLS sẽ ước lượng hàm của $\text{Var}(u|x)$ để làm trọng số trong phương pháp WLS.

Các bước thực hiện FGLS

1. Hồi quy y theo các biến giải thích, và ước lượng phần dư \hat{u} .
2. Tạo biến $\log(\hat{u}^2)$.
3. Ước lượng hồi quy $\log(\hat{u}^2)$ lên các biến giải thích, và ước lượng giá trị dự báo (fitted value), $\widehat{\log(\hat{u}^2)}$.
4. Lấy lũy thừa cơ số e của giá trị dự báo ở bước 3, $\widehat{h(x)} = e^{\widehat{\log(\hat{u}^2)}}$.
5. Ước lượng lại mô hình ban đầu bằng WLS, với trọng số là $1/\widehat{h(x)}$.

Ví dụ hồi quy WLS và FGLS

Sử dụng lại mô hình tỷ suất thu nhập với bộ dữ liệu VHLSS 2010.

- ▶ Ước lượng WLS nếu biết phương sai của sai số tuân theo:

$$\text{Var}(u|x) = \sigma^2 inc$$

- ▶ Ước lượng FGLS cho trường hợp phương sai thay đổi và không biết cấu trúc hàm.
- ▶ So sánh các ước lượng sử dụng OLS, Heteroskedasticity-robust standard errors, WLS, và FGLS.

Regression Results

	Homo b/se	Robust b/se	WLS b/se	FGLS b/se
yoeduc	0.0926*** (0.0027)	0.0926*** (0.0031)	0.0909*** (0.0043)	0.0993*** (0.0026)
yoexper	0.0617*** (0.0025)	0.0617*** (0.0032)	0.1088*** (0.0029)	0.0681*** (0.0028)
yoexpersq	-0.0012*** (0.0000)	-0.0012*** (0.0001)	-0.0019*** (0.0000)	-0.0013*** (0.0001)
married	0.0352 (0.0221)	0.0352 (0.0217)	0.0966** (0.0339)	0.0073 (0.0206)
publicSchool	-0.1146** (0.0424)	-0.1146* (0.0465)	-0.0153 (0.0530)	-0.1262** (0.0435)
public	-0.1043** (0.0329)	-0.1043* (0.0423)	-0.3122*** (0.0489)	-0.0938* (0.0472)
foreign	0.4499*** (0.0364)	0.4499*** (0.0328)	0.9413*** (0.0735)	0.4529*** (0.0286)
official	0.2705*** (0.0359)	0.2705*** (0.0430)	0.8263*** (0.0614)	0.2296*** (0.0475)
Constant	8.4936*** (0.0475)	8.4936*** (0.0539)	6.9670*** (0.0567)	8.4044*** (0.0492)
Obs	7552.0000	7552.0000	7552.0000	7552.0000
R2	0.3026	0.3026	0.3249	0.3460
R2-adj	0.3019	0.3019	0.3242	0.3453
df(r)	7543.0000	7543.0000	7543.0000	7543.0000
SSR	4040.8653	4040.8653	9608.9369	3408.4380

* p<0.05, ** p<0.01, *** p<0.001