

# Chương 5

## Chẩn đoán hồi quy: Phương sai thay đổi

(Gujarati: Econometrics by example, 2011)<sup>1</sup>.

Người dịch và diễn giải: Phùng Thanh Bình

<http://vnp.edu.vn/>



Một trong những vấn đề thường gặp trong dữ liệu chéo là phương sai thay đổi (phương sai không bằng nhau) trong thành phần hạng nhiễu. Có nhiều lý do của phương sai thay đổi, như hiện diện của các quan sát bất thường trong dữ liệu (outliers), sai dạng hàm của mô hình hồi quy, hoặc chuyển đổi dữ liệu không đúng, hoặc hỗn hợp các quan sát với các thước đo quy mô khác nhau (như hỗn hợp các gia đình thu nhập cao với các gia đình thu nhập thấp), vân vân.

### 5.1 Hậu quả của phương sai thay đổi<sup>2</sup>

Mô hình hồi quy tuyến tính cổ điển giả định rằng hạng nhiễu  $u_i$  trong mô hình hồi quy có phương sai không đổi (phương sai bằng nhau) qua các quan sát, được ký hiệu là  $\sigma^2$ . Ví dụ, khi nghiên cứu chi tiêu cho tiêu dùng trong mỗi quan hệ với thu nhập, giả định này hàm ý rằng các hộ gia đình thu nhập thấp và thu nhập cao có cùng phương sai của hạng nhiễu mặc dù mức chi tiêu cho tiêu dùng trung bình là khác nhau.

Tuy nhiên, nếu giả định phương sai không đổi, hoặc phương sai bằng nhau, không thỏa mãn, thì chúng ta gặp vấn đề phương sai không đồng nhất, được ký hiệu là  $\sigma_i^2$  (lưu ý chỉ số dưới  $i$ ). Vì thế, so sánh với các hộ gia đình thu nhập thấp, thì các hộ gia đình thu nhập cao không chỉ có mức chi tiêu cho tiêu dùng trung bình cao hơn, mà mà khả năng biến thiên lớn hơn trong chi tiêu cho tiêu dùng của họ. Kết quả là, trong mô hình hồi quy chi tiêu cho tiêu dùng theo thu nhập gia đình chúng ta có thể gặp phải vấn đề phương sai thay đổi.

Phương sai thay đổi có các hậu quả sau đây:

1. Phương sai thay đổi không làm thay đổi các tính chất không chệch (unbiasedness) và nhất quán (consistency) của các ước lượng OLS.
2. Nhưng các ước lượng OLS không còn hiệu quả, hoặc không có phương sai bé nhất nữa. Nghĩa là, chúng không còn là các ước lượng tuyến tính không chệch tốt nhất (BLUE); chúng chỉ đơn giản là các ước lượng tuyến tính không chệch (LUE).

---

<sup>1</sup> Hiện nay đã có ấn bản mới (lần 2, năm 2015). Dữ liệu của phiên bản 2011:

<https://www.macmillanihe.com/companion/Gujarati-Econometrics-By-Example/student-zone/>

<sup>2</sup> Để biết chi tiết, xem Gujarati/Porter, Chương 11.

3. Kết quả là, các kiểm định t và F dựa trên các giả định chuẩn của mô hình hồi quy tuyến tính cổ điển không thể tin cậy, dẫn đến các kết luận sai lầm về ý nghĩa thống kê của các hệ số hồi quy được ước lượng.
4. Khi có hiện tượng phương sai thay đổi, các ước lượng BLUE được cung cấp bởi phương pháp bình phương bé nhất có trọng số (WLS, weighted least squares).

Bởi vì những hậu quả này, nên điều quan trọng là chúng ta phải kiểm tra phương sai thay đổi, thường gặp ở dữ liệu chéo. Trước khi làm điều này, chúng ta hãy xem xét một ví dụ cụ thể.

## 5.2 Tỷ lệ phá thai ở Mỹ

Đâu là các yếu tố quyết định tỷ lệ phá thai giữa 50 bang ở Mỹ? Để nghiên cứu vấn đề này, chúng ta thu thập dữ liệu như trong **Table 5.1**, bạn có thể tìm thấy tập dữ liệu này trên trang web của cuốn sách<sup>3</sup>.

Các biến được sử dụng trong phần tích này như sau:

*State* = tên của bang (50 bang của Mỹ)

*ABR* = tỷ lệ phá thai, số ca phá thai trên 1000 phụ nữ tuổi từ 15 đến 44 năm 1992.

*Religion* = phần trăm dân số của bang theo đạo Cơ đốc giáo, đạo Báp tit miền nam, đạo Tin lành, hoặc đạo Mặc môn.

*Price* = giá trung bình năm 1993 cho các thiết bị y tế ngoài bệnh viện tính cho một ca phá thai giai đoạn 10 tuần có gây mê tại chỗ (tính trọng số bởi số ca phá thai thực hiện năm 1992).

*Laws* = một biến nhận giá trị bằng 1 nếu bang ban hành luật hạn chế phá thai, bằng 0 nếu không có luật này.

*Funds* = một biến nhận giá trị bằng 1 nếu quỹ của bang sẵn có cho việc sử dụng thanh toán một lần phá thai cho hầu hết các trường hợp, bằng 0 nếu không có sẵn quỹ.

*Educ* = phần trăm dân số của bang mà dân số từ 25 tuổi trở lên có bằng phổ thông trung học hoặc tương đương, 1990.

*Income* = thu nhập khả dụng bình quân đầu người, 1992.

*Picket* = phần trăm người trả lời khai báo có tham gia biểu tình ngăn chặn bệnh nhân.

## Mô hình

Như một điểm khởi đầu, chúng ta xem xét mô hình hồi quy tuyến tính sau đây:

$$\begin{aligned} ABR_i = & B_1 + B_2 Rel_i + B_3 Price_i + B_4 Laws_i + B_5 Funds_i \\ & + B_6 Educ_i + B_7 Income_i + B_8 Picket_i + u_i \end{aligned} \quad (5.1)$$

$$i = 1, 2, \dots, 50$$

Chúng ta kỳ vọng ABR có mối quan hệ âm với tôn religion, price, laws, picket, educ, và có mối quan hệ dương với fund và income. Chúng ta giả định hạng nhiễu thỏa mãn các giả định cổ điển chuẩn, bao gồm giả định phương sai không đổi. Dĩ nhiên, chúng ta sẽ

<sup>3</sup> Dữ liệu được thu thập từ trang web của Leo H. Kahane, <http://www.cbe.csueastbay.edu/~kahane>.

thực hiện phân tích hậu ước lượng, để biết liệu giả định này có thỏa trong tình huống hiện tại hay không.

Sử dụng Eviews 6, chúng ta có kết quả như trong Bảng 5.2, đây là một mẫu kết quả chuẩn của phần mềm Eviews.

**Bảng 5.2:** Ước lượng OLS của hàm tỷ lệ phá thai.

Dependent Variable: ABORTION Method: Least Squares Sample: 1 50 Included observations: 50				
	Coefficient	Std. Error	t-Statistic	Prob.
C	14.28396	15.07763	0.947361	0.3489
RELIGION	0.020071	0.086381	0.232355	0.8174
PRICE	-0.042363	0.022223	-1.906255	0.0635
LAWS	-0.873102	2.376566	-0.367380	0.7152
FUNDS	2.820003	2.783475	1.013123	0.3168
EDUC	-0.287255	0.199555	-1.439483	0.1574
INCOME	0.002401	0.000455	5.274041	0.0000
PICKET	-0.116871	0.042180	-2.770782	0.0083
R-squared	0.577426	Mean dependent var	20.57800	
Adjusted R-squared	0.506997	S.D. dependent var	10.05863	
S.E. of regression	7.062581	Akaike info criterion	6.893145	
Sum squared resid	2094.962	Schwarz criterion	7.199069	
Log likelihood	-164.3286	Durbin-Watson stat	2.159124	
F-statistic	8.198706	Prob(F-statistic)	0.000003	

Như các kết quả này cho thấy, trên cơ sở thống kê t, thì price, income, và picket có ý nghĩa thống kê ở mức ý nghĩa 10%, trong khi đó các biến khác không có ý nghĩa thống kê, mặc dù một số (như laws và educ) có dấu đúng. *Nhưng nhớ rằng nếu có phương sai thay đổi, thì các giá trị t ước lượng có thể không tin cậy.*

Giá trị  $R^2$  cho thấy 58% biến thiên trong tỷ lệ phá thai được giải thích bởi mô hình. Thống kê F, kiểm định giả thuyết cho rằng tất cả các hệ số độ dốc đồng thời bằng 0, rõ ràng bác bỏ giả thuyết này, vì giá trị 8.199 có ý nghĩa thống kê rất cao, và giá trị xác suất p của nó thực tế bằng 0. *Một lần nữa, hãy nhớ rằng thống kê F có thể không tin cậy nếu có hiện tượng phương sai thay đổi.*

Lưu ý rằng F có ý nghĩa thống kê thì không có nghĩa là mỗi biến giải thích có ý nghĩa thống kê, vì thống kê t cho thấy chỉ có vài biến giải thích có ý nghĩa thống kê.

### Phân tích kết quả

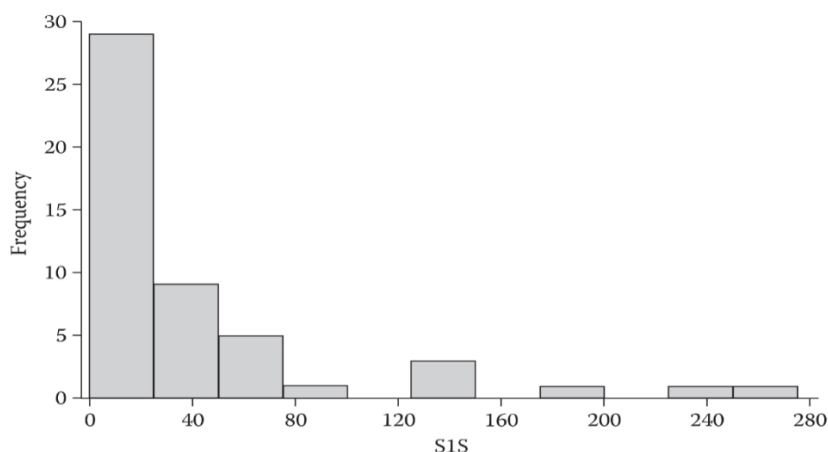
Nhưng đã lưu ý, một vấn đề thường gặp trong dữ liệu chéo là vấn đề phương sai thay đổi. Trong ví dụ của chúng ta, bởi vì sự đa dạng của các bang nên chúng ta hoài nghi có phương sai thay đổi.

Như một kiểm định đơn giản để phát hiện phương sai thay đổi, chúng ta vẽ đồ thị phần dư bình phương (S1S) từ hồi quy trong Bảng 5.2, xem Hình 5.1.

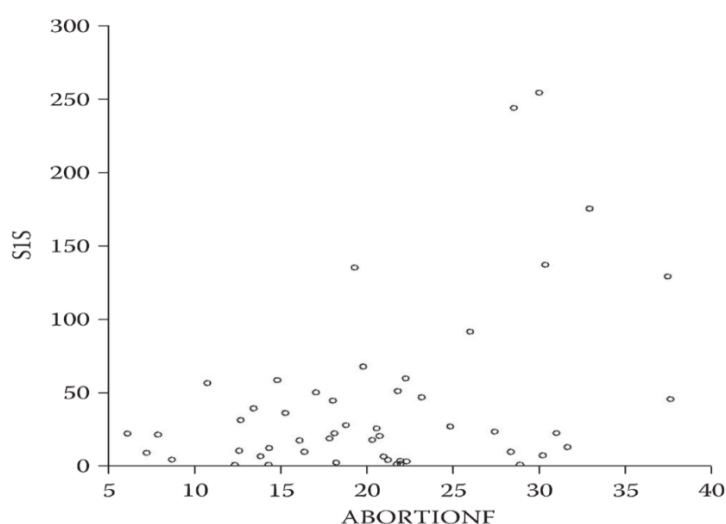
Rõ ràng từ hình này chúng ta thấy rằng, phần dư bình phương – một đại diện của hạng nhiễu bình phương, không cho thấy rằng hạng nhiễu có phương sai không đổi<sup>4</sup>.

Chúng ta có cái nhìn rõ hơn về phương sai thay đổi nếu chúng ta vẽ đồ thị phần dư bình phương (S1S) theo tỷ lệ phá thai ước lượng từ mô hình hồi quy, tức  $\widehat{ABR}$  (Hình 5.2).

**Hình 5.1:** Đồ thị tần suất của phần dư bình phương.



**Hình 5.2:** Phần dư bình phương theo tỷ lệ phá thai ước lượng.



[Diễn giải: Trên Eviews, sau khi vừa hồi quy Bảng 5.2, chúng ta tạo biến  $S1S = \text{resid}^2$  và  $ABR\_hat = ABR - \text{resid}$ . Với Stata, sau khi hồi quy Bảng 5.2, chúng ta sử dụng lệnh `predict S1S, resid;` và `predict ABR_hat`].

### 5.3 Phát hiện phương sai thay đổi

Bên cạnh các phương pháp đồ thị được mô tả ở phần trước, chúng ta có thể sử dụng hai kiểm định phương sai thay đổi được sử dụng phổ biến, đó là kiểm định Breusch-Pagan và kiểm định White<sup>5</sup>.

<sup>4</sup> Nhớ lại rằng ước lượng OLS của phương sai hạng nhiễu được cho bằng:  $\hat{\sigma}^2 = \sum e_i^2 / (n - k)$ , nghĩa là tổng bình phương phần dư chia cho bậc tự do.

<sup>5</sup> Chi tiết về hai kiểm định này và những kiểm định khác có thể thấy trong Gujarati/Porter, Chương 11.

## Kiểm định Breusch-Pagan (BP)

Kiểm định này bao gồm các bước sau đây:

1. Ước lượng hồi quy OLS, như ở Bảng 5.2, và lưu phần dư OLS, tạo biến phần dư bình phương,  $e_i^2$ , thử hồi quy này.
2. Hồi quy  $e_i^2$  theo  $k$  biến giải thích trong mô hình; ý tưởng ở đây là xem phần dư bình phương (như một đại diện của hạng nhiễu bình phương) có liên quan với một hoặc nhiều biến  $X^6$ . Bạn cũng có thể chọn các biến giải thích khác mà bạn cho là có liên quan đến phương sai hạng nhiễu. Bây giờ chạy phương trình hồi quy sau đây:

$$e_i^2 = A_1 + A_2 Rel_i + A_3 Price_i + A_4 Laws_i + A_5 Funds_i + A_6 Educ_i + A_7 Income_i + A_8 Picket_i + v_i \quad (5.2)$$

Trong đó  $v_i$  là hạng nhiễu.

Lưu  $R^2$  từ hồi quy (5.2); gọi là  $R^2_{aux}$  (tức  $R^2$  của hồi quy phụ), vì phương trình (5.2) là hồi quy phụ của phương trình hồi quy chính (5.1) (xem Bảng 5.3). Ý tưởng đằng sau phương trình (5.2) là tìm hiểu xem liệu phần dư bình phương có quan hệ gì với một hoặc nhiều biến giải thích, nếu có thì đó là dấu hiệu cho chúng ta biết có lẽ có phương sai thay đổi hiện diện trong dữ liệu.

**Bảng 5.3:** Kiểm định BP về phương sai thay đổi.

Heteroskedasticity Test: Breusch–Pagan–Godfrey				
F-statistic	2.823820	Prob. F(7,42)	0.0167	
Obs*R-squared	16.00112	Prob. Chi-Square(7)	0.0251	
Scaled explained SS	10.57563	Prob. Chi-Square(7)	0.1582	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 10/05/09 Time: 13:14				
Sample: 1 50				
Included observations: 50				
	Coefficient	Std. Error	t-Statistic	Prob.
C	16.68558	110.1532	0.151476	0.8803
RELIGION	−0.134865	0.631073	−0.213707	0.8318
PRICE	0.286153	0.162357	1.762492	0.0853
LAWS	−8.566472	17.36257	−0.493387	0.6243
FUNDS	24.30981	20.33533	1.195447	0.2386
EDUC	−1.590385	1.457893	−1.090879	0.2815
INCOME	0.004710	0.003325	1.416266	0.1641
PICKET	−0.576745	0.308155	−1.871606	0.0682
R-squared	0.320022	Mean dependent var	41.89925	
Adjusted R-squared	0.206693	S.D. dependent var	57.93043	
S.E. of regression	51.59736	Akaike info criterion	10.87046	
Sum squared resid	111816.1	Schwarz criterion	11.17639	
Log likelihood	−263.7616	Durbin–Watson stat	2.060808	
F-statistic	2.823820	Prob(F-statistic)	0.016662	

<sup>6</sup> Mặc dù  $e_i^2$  không giống  $u_i^2$ , nhưng trong các mẫu lớn thì phần dư bình phương là một đại diện tốt cho phương sai hạng nhiễu.

3. Giả thuyết không ( $H_0$ ) ở đây là phương sai của hạng nhiễu là đồng nhất, nghĩa là, tất cả các hệ số độ dốc trong phương trình (5.2) đồng thời bằng 0<sup>7</sup>. Bạn có thể sử dụng thống kê F từ hồi quy này với  $(k - 1)$  và  $(n - k)$  là bậc tự do trên tử số và mẫu số để kiểm định giả thuyết  $H_0$  này. Nếu thống kê F tính toán từ phương trình (5.2) có ý nghĩa thống kê, chúng ta có thể bác bỏ giả thuyết về phương sai không đổi. Nếu không có ý nghĩa thống kê, thì chúng ta có thể không bác bỏ giả thuyết  $H_0$  này.

Như kết quả trong Bảng 5.3 cho thấy, thống kê F (với 7 bậc tự do trên tử và 42 bậc tự do dưới mẫu) có ý nghĩa thống kê cao, vì giá trị xác suất p của thống kê này chỉ khoảng 2%. Vì thế chúng ta có thể bác bỏ giả thuyết  $H_0$ .

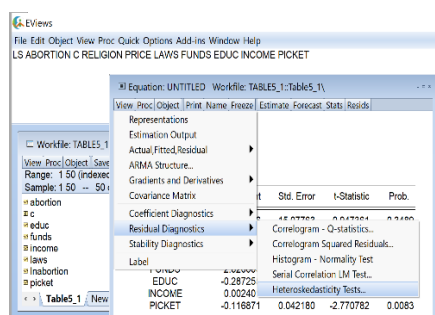
4. Một cách khác, chúng ta có thể sử dụng thống kê Chi bình phương. Chúng ta có thể thấy rằng dưới giả thuyết  $H_0$  về phương sai không đổi, tích của  $R^2_{aux}$  (được tính ở bước 2) và số quan sát ( $n$ ) theo phân phối Chi bình phương, với số bậc tự do bằng với số biến giải thích trong mô hình. Nếu giá trị Chi bình phương tính toán có một giá trị xác suất thấp, chúng ta có thể bác bỏ giả thuyết  $H_0$  về phương sai không đổi<sup>8</sup>. Như kết quả trong Bảng 5.3 cho thấy, giá trị Chi bình phương quan sát ( $=nR^2_{aux}$ ) khoảng bằng 16 có một giá trị xác suất p rất thấp, điều này cho thấy rằng chúng ta có thể bác bỏ giả thuyết  $H_0$  về phương sai không đổi. Nói một cách khác, hồi quy trong Bảng 5.2 bị hiện tượng phương sai thay đổi.

*Một lưu ý cảnh báo:* Kiểm định này là một kiểm định mẫu lớn và có thể không phù hợp trong các mẫu nhỏ<sup>9</sup>.

Tóm lại, chúng ta có thể kết luận rằng dường như hồi quy về tỷ lệ phá thai gặp phải vấn đề phương sai thay đổi.

Trở lại với ví dụ của chúng ta, chúng ta có được các kết quả như được trình bày trong Bảng 5.3.

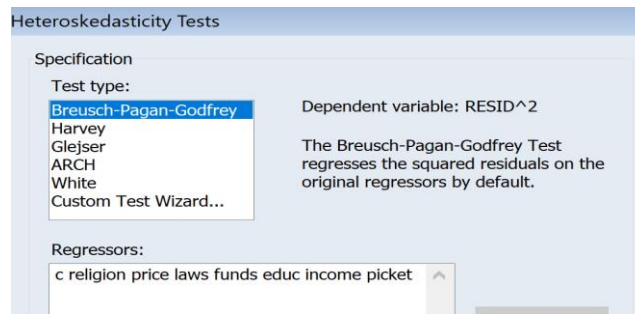
Hướng dẫn kiểm định Breusch-Pagan với Eviews và Stata:



<sup>7</sup> Nếu điều đó xảy ra, hằng số  $A_1$  sẽ gợi cho chúng ta rằng phương sai của hạng nhiễu là một hằng số hoặc phương sai không đổi.

<sup>8</sup> Nhớ lại mối quan hệ giữa các thống kê F và Chi bình phương, đó là:  $mF_m = \chi^2_{m, n}$  khi  $m \rightarrow \infty$ ; nghĩa là, với bậc tự do ở mẫu số lớn, thì số bậc tự do ở tử số nhân với giá trị F xấp xỉ bằng giá trị Chi bình phương với số bậc tự do ở tử số, trong đó  $m$  và  $n$  lần lượt là số bậc tự do trên tử số và dưới mẫu số (xem Phụ lục thống kê).

<sup>9</sup> Bạn có thể lập luận rằng dữ liệu mà chúng ta có thực sự không phải là một mẫu ngẫu nhiên, vì chúng ta có tất cả các bang ở Mỹ. Vì thế, chúng ta thực sự có đầy đủ tổng thể. Nhưng nhớ rằng dữ liệu về tỷ lệ phá thai chỉ cho một năm. Rất có thể là tỷ lệ này sẽ thay đổi từ năm này sang năm khác. Vì vậy, chúng ta có thể xử lý dữ liệu được sử dụng cho một năm như một mẫu từ tất cả các tỷ lệ phá thai có thể có của tất cả các năm mà chúng ta có dữ liệu.



#### Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	2.823820	Prob. F(7,42)	0.0167
Obs*R-squared	16.00112	Prob. Chi-Square(7)	0.0251
Scaled explained SS	10.57563	Prob. Chi-Square(7)	0.1582

**Lưu ý:** Stata thay các biến giải thích bằng  $\hat{Y}_i$ , nên giá trị kiểm định có khác so với kết quả từ Eviews.

```
. reg abortion religion price laws funds educ income picket
```

Source	SS	df	MS	Number of obs	=	50
Model	2862.66338	7	408.951912	F(7, 42)	=	8.20
Residual	2094.96246	42	49.8800585	Prob > F	=	0.0000
Total	4957.62584	49	101.176038	R-squared	=	0.5774
				Adj R-squared	=	0.5070
				Root MSE	=	7.0626

abortion	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
religion	.0200709	.0863805	0.23	0.817	-.1542521 .1943939
price	-.0423631	.0222232	-1.91	0.063	-.0872113 .0024851
laws	-.8731018	2.376566	-0.37	0.715	-5.669206 3.923003
funds	2.820003	2.783475	1.01	0.317	-2.797276 8.437282
educ	-.2872551	.1995545	-1.44	0.157	-.6899725 .1154622
income	.0024007	.0004552	5.27	0.000	.0014821 .0033193
picket	-.1168712	.0421799	-2.77	0.008	-.2019936 -.0317488
_cons	14.28396	15.07763	0.95	0.349	-16.14393 44.71185

```
. hettest
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of abortion

chi2(1) = 8.25

Prob > chi2 = 0.0041

```
. hettest, fstat
```

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of abortion

F(1, 48) = 10.26

Prob > F = 0.0024



## Kiểm định White

Chúng ta tiếp tục tinh thần của kiểm định BP và hồi quy phần dư bình phương theo 7 biến giải thích, bình phương của các biến giải thích, và tích từng cặp giữa các biến giải thích này. Như vậy, chúng ta có tổng cộng 33 hệ số.

Như trong kiểm định BP, chúng ta thu giá trị  $R^2$  từ hồi quy (phụ) này và nhân  $R^2$  với số quan sát. Dưới giả thuyết  $H_0$  là phương sai không đổi, thì tích ( $nR^2$ ) theo phân phối Chi bình phương với số bậc tự do bằng với số hệ số ước lượng. Kiểm định White tổng quát hơn và linh hoạt hơn so với kiểm định BP.

Trong ví dụ hiện tại của chúng ta, nếu chúng ta không đưa các số hạng bình phương và tích chéo theo cặp vào hồi quy phụ, thì chúng ta có  $nR^2 = 15.7812$ , giá trị này theo phân phối Chi bình phương với 7 bậc tự do. Xác suất để có một giá trị Chi bình phương bằng hoặc lớn hơn giá trị như thế khoảng 0.03, giá trị xác suất này khá thấp. Điều này cho thấy rằng chúng ta có thể bác bỏ giả thuyết  $H_0$  về phương sai không đổi.

Nếu chúng ta đưa thêm các số hạng bình phương và tích chéo theo cặp vào hồi quy phụ, thì chúng ta có  $nR^2 = 32.1022$ , có giá trị Chi bình phương với 33 bậc tự do<sup>10</sup>. Xác suất để có một giá trị Chi bình phương như thế khoảng 0.51. Trong trường hợp này, chúng ta sẽ không bác bỏ giả thuyết  $H_0$ .

Như bài tập này cho thấy, kiểm định Chi bình phương theo White rất nhạy cảm đối với việc chúng ta đưa hay bỏ các số hạng bình phương và tích chéo vào/ra hồi quy phụ<sup>11</sup>. Nhớ rằng kiểm định White là một kiểm định mẫu lớn. Vì thế, khi chúng ta đưa các biến giải thích và các số hạng bình phương và tích chéo vào mô hình, dẫn đến mất đi 33 bậc tự do. Nên kết quả của hồi quy phụ có thể rất nhạy cảm, như trường hợp hiện tại của chúng ta.

Để tránh mất quá nhiều bậc tự do, kiểm định White có thể được rút ngắn bằng cách hồi quy phần dư bình phương theo giá trị ước lượng của biến phụ thuộc và bình phương của nó<sup>12</sup>. Nghĩa là, chúng ta hồi quy phương trình sau đây:

$$e_i^2 = \alpha_1 + \alpha_2 \text{Abortionf} + \alpha_3 \text{Abortionf}^2 + v_i \quad (5.3)$$

Trong đó  $\text{Abortionf}$  = giá trị dự báo của tỷ lệ phá thai từ phương trình (5.1). Vì tỷ lệ phá thai ước lượng là một hàm tuyến tính của các biến giải thích trong mô hình của phương trình (5.1), theo một cách mà chúng ta đang gián tiếp đưa các biến giải thích và các bình phương của chúng để ước lượng phương trình (5.3), điều này vẫn trong tinh thần của kiểm định White ban đầu. Nhưng lưu ý rằng trong phương trình (5.3) không có chỗ cho

<sup>10</sup> Bởi vì chúng ta có 7 biến giải thích, 5 biến giải thích bình phương (tức trừ 2 biến giả) và tích chéo của mỗi biến giải thích với các biến giải thích khác. Nhưng lưu ý rằng chúng ta không đưa giá trị bình phương của các biến giả, vì bình phương của một biến giả có giá trị bằng 1 cũng là 1. Cũng lưu ý rằng tích chéo của biến religion với income cũng giống như giống với tích chéo của income với religion, để tránh tính trùng (double-counting).

<sup>11</sup> Đó là lý do tại sao phải lưu ý rằng kiểm định White sức mạnh thống kê (statistical power) yếu. Sức mạnh thống kê của một kiểm định là xác suất bác bỏ giả thuyết  $H_0$  khi giả thuyết đó sai.

<sup>12</sup> Xem Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 4<sup>th</sup> edn, South-Western Publishing, 2009, p. 275.



các số hạng tích chéo, vì vậy loại bỏ các số hạng chéo trong kiểm định White gốc. Vì vậy, phiên bản kiểm định White rút gọn tiết kiệm được nhiều bậc tự do.

Kết quả của hồi quy này được trình bày trong Bảng 5.4. Thống kê mà chúng ta quan tâm trong bảng này là thống kê F, có ý nghĩa thống kê rất cao, vì giá trị xác suất p của nó rất thấp. Vì vậy kiểm định White rút gọn khẳng định lại kiểm định BP và kết luận rằng hàm tỷ lệ phá thai thực sự có hiện tượng phương sai thay đổi. Và kết luận này có được với ít mất mát bậc tự do hơn.

Chú ý rằng mặc dù thống kê F có ý nghĩa, nhưng từng hệ số hồi quy riêng không có ý nghĩa thống kê. Thật tình cờ, nếu bạn bỏ số hạng ABORTIONF bình phương khỏi phương trình (5.3), bạn sẽ thấy số hạng ABORTIONF có ý nghĩa thống kê<sup>13</sup>. Lý do là các số hạng ABORTIONF và bình phương của nó có mối quan hệ hàm số, tạo nên sự đa cộng tuyến. Nhưng hãy nhớ rằng đa cộng tuyến đề cập đến các mối quan hệ tuyến tính giữa các biến và không phải các mối quan hệ phi tuyến, như ở phương trình (5.3).

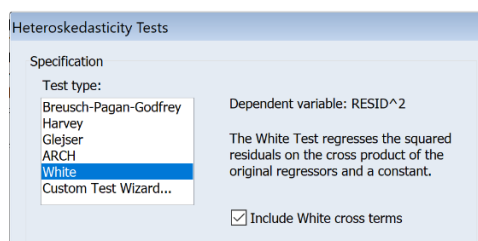
**Bảng 5.4:** Kiểm định White rút gọn.

Dependent Variable: RES^2 Method: Least Squares Sample: 1 50 Included observations: 50 White Heteroskedasticity-Consistent Standard Errors & Covariance				
	Coefficient	Std. Error	t-Statistic	Prob.
C	20.20241	27.09320	0.745663	0.4596
ABORTIONF	-1.455268	3.121734	-0.466173	0.6432
ABORTIONF^2	0.107432	0.081634	1.316014	0.1946
R-squared	0.193083	Mean dependent var	41.89925	
Adjusted R-squared	0.158746	S.D. dependent var	57.93043	
S.E. of regression	53.13374	Akaike info criterion	10.84163	
Sum squared resid	132690.1	Schwarz criterion	10.95635	
Log likelihood	-268.0406	Durbin-Watson stat	1.975605	
F-statistic	5.623182	Prob(F-statistic)	0.006464	

Lưu ý rằng dù chúng ta nên sử dụng kiểm định BP hay kiểm định White hoặc bất kỳ kiểm định nào khác về phương sai thay đổi, thì các kiểm định này sẽ chỉ cho thấy liệu phương sai hạng nhiễu trong một trường hợp cụ thể có thay đổi hay không. Nhưng các kiểm định này không nhất thiết đề xuất nên làm gì nếu chúng ta gặp vấn đề phương sai thay đổi.

[Diễn giải: Kiểm định White trên Eviews và Stata:]

Tương tự như kiểm định BP, như chúng ta chọn:



<sup>13</sup> Hệ số của Abortionf là 3.1801 với giá trị t là 3.20, có ý nghĩa thống kê ở mức 0.002.

#### Heteroskedasticity Test: White

F-statistic	0.869647	Prob. F(33,16)	0.6455
Obs*R-squared	32.10224	Prob. Chi-Square(33)	0.5116
Scaled explained SS	21.21735	Prob. Chi-Square(33)	0.9437

. imtest

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	<b>32.10</b>	<b>33</b>	<b>0.5116</b>
Skewness	<b>11.73</b>	<b>7</b>	<b>0.1099</b>
Kurtosis	<b>0.11</b>	<b>1</b>	<b>0.7415</b>
Total	<b>43.94</b>	<b>41</b>	<b>0.3481</b>

## 5.4 Biện pháp khắc phục

Sau khi biết các hậu quả của phương sai thay đổi, có lẽ chúng ta cần tìm các biện pháp khắc phục. Vấn đề ở đây là chúng ta không biết các phương sai thực khác nhau, tức các  $\sigma_i^2$ , vì hiếm khi chúng được quan sát. Nếu chúng ta có thể quan sát chúng, thì chúng ta có thể có được các ước lượng BLUE bằng cách chia mỗi quan sát cho  $\sigma_i$  (thay đổi) và ước lượng mô hình được chuyển hóa theo OLS. Phương pháp ước lượng này được biết với tên gọi là **bình phương bé nhất có trọng số** (WLS)<sup>14</sup>. Thật không may, các phương sai thực  $\sigma_i^2$  hiếm khi được biết. Như vậy giải pháp là gì?

Trong thực tế, chúng ta thực hiện các phán đoán về các  $\sigma_i^2$  có thể có là gì và chuyển đổi mô hình hồi quy gốc theo một cách sao cho trong mô hình được chuyển hóa phương sai của hạng nhiễu có thể là đồng nhất. Vài cách chuyển hóa được sử dụng trong thực tế như sau<sup>15</sup>:

1. Nếu phương sai thực tỷ lệ với bình phương của một trong số các biến giải thích, chúng ta có thể chia cả hai vế của phương trình (5.1) cho biến đó và chạy hồi quy đã được chuyển hóa. Giả sử trong phương trình (5.1), phương sai của hạng nhiễu tỷ lệ với bình phương của biến income. Vì thế chúng ta chia cả hai vế của phương trình (5.1) cho biến income và ước lượng phương trình chuyển hóa này. Sau đó chúng ta kiểm định phương sai thay đổi cho phương trình này bằng các kiểm định như BP và White. Nếu các kiểm định này chỉ ra rằng không có dấu hiệu của phương sai thay đổi, thì chúng ta giả định rằng hạng nhiễu được chuyển hóa có phương sai không đổi.
2. Nếu phương sai thực tỷ lệ với một trong số các biến giải thích, chúng ta có thể cách **chuyển hóa bình phương** (square transformation), nghĩa là, chia cả hai vế của phương trình (5.1) cho căn bậc hai của biến giải thích được chọn. Sau đó

<sup>14</sup> Vì mỗi quan sát chia cho (tức gán trọng số)  $\sigma_i$ , nên một quan sát với  $\sigma_i$  lớn sẽ bị giảm nhiều hơn so với một quan sát với  $\sigma_i$  thấp.

<sup>15</sup> Để biết chi tiết, xem Gujarati/Porter, pp. 392 – 5.

chúng ta ước lượng hồi quy đã được chuyển hóa và kiểm định phương sai thay đổi cho phương trình này bằng các kiểm định như BP và White. Nếu các kiểm định này thỏa mãn, thì chúng ta có thể sử dụng kết quả hồi quy này.

Có các vấn đề thực tế trong các ứng dụng các thủ tục này. Thứ nhất, làm sao chúng ta biết chọn biến giải thích nào để chuyển hóa nếu có nhiều biến giải thích? Chúng ta có thể tiến hành bằng cách thử - sai, nhưng cách đó có thể sẽ mất thời gian. Thứ hai, nếu vài giá trị của biến giải thích được chọn bằng 0, thì việc chia cho 0 hiển nhiên sẽ là có vấn đề.

Sự lựa chọn nhằm biến giải thích đôi khi có thể tránh được bằng cách sử dụng giá trị  $Y$  ước lượng (tức là  $\hat{Y}$ ), đó là một giá trị trung bình có trọng số của tất cả các biến giải thích trong mô hình, và trọng số chính là các hệ số hồi quy, tức là các  $bs$ .

Cũng cần lưu ý rằng tất cả các phương pháp chuyển hóa này hơi mang tính tình thế. Nhưng chúng ta không có nhiều sự lựa chọn, vì chúng ta đang cố gắng dự đoán phương sai thực của hạng nhiễu là gì. Tất cả những gì mà chúng ta có thể hy vọng là mong sao dự đoán của chúng ta đúng một cách hợp lý.

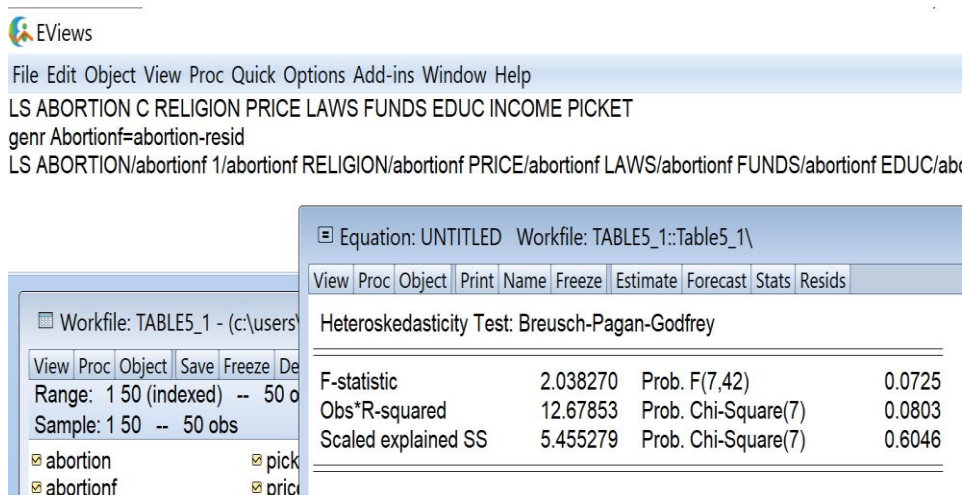
Để minh họa tất cả những cách chuyển hóa này có thể sẽ tốn nhiều thời gian và không gian. Tuy nhiên, chúng ta sẽ chỉ minh họa một trong những cách chuyển hóa này. Nếu chúng ta chia phương trình (5.1) cho tỷ lệ phá thai ước lượng từ phương trình (5.1), chúng ta có kết quả như được trình bày trong Bảng 5.5.

**Bảng 5.5:** Phương trình (5.1) sau khi chuyển hóa.

Dependent Variable: ABORTION/ABORTIONF				
Method: Least Squares				
Sample: 1 50				
Included observations: 50				
	Coefficient	Std. Error	t-Statistic	Prob.
1/ABORTIONF	12.81786	11.22852	1.141545	0.2601
RELIGION/ABORTIONF	0.066088	0.068468	0.965239	0.3400
PRICE/ABORTIONF	-0.051468	0.017507	-2.939842	0.0053
LAWS/ABORTIONF	-1.371437	1.819336	-0.753812	0.4552
FUNDS/ABORTIONF	2.726181	3.185173	0.855897	0.3969
EDUC/ABORTIONF	-0.228903	0.147545	-1.551408	0.1283
INCOME/ABORTIONF	0.002220	0.000481	4.616486	0.0000
PICKET/ABORTIONF	-0.082498	0.031247	-2.640211	0.0116
R-squared	0.074143	Mean dependent var	1.011673	
Adjusted R-squared	-0.080166	S.D. dependent var	0.334257	
S.E. of regression	0.347396	Akaike info criterion	0.868945	
Sum squared resid	5.068735	Schwarz criterion	1.174869	
Log likelihood	-13.72363	Durbin-Watson stat	2.074123	
Note: Abortionf is the abortion rate forecast from Eq. (5.1)				

Chúng ta thực hiện các kiểm định BP và White cho phương trình này, nhưng cả hai kiểm định đều cho thấy vấn đề phương sai thay đổi vẫn còn tồn tại<sup>16</sup>.

<sup>16</sup> Để tiết kiệm không gian, chúng ta không trình bày chi tiết các kết quả kiểm định. Bạn đọc có thể xác nhận kết luận này bằng cách chạy các kiểm định với dữ liệu được cho trong Table 5.1.



#### Heteroskedasticity Test: White

F-statistic	2.121937	Prob. F(8,41)	0.0555
Obs*R-squared	14.64023	Prob. Chi-Square(8)	0.0665
Scaled explained SS	6.299352	Prob. Chi-Square(8)	0.6137

Cũng cần nói thêm rằng chúng ta thực hiện các chuyển hóa nhằm mục đích loại bỏ phương sai thay đổi. Chúng ta có thể quay trở lại hồi quy gốc bằng cách nhân hai vế cho ABORTIONF trong Bảng 5.5.

- Chuyển hóa sang dạng logarit: Thỉnh thoảng, thay vì ước lượng hồi quy phương trình (5.1), chúng ta có thể hồi quy dạng logarit của biến phụ thuộc theo các biến giải thích, các biến giải thích có thể ở dạng tuyến tính hoặc dạng log. Lý do chuyển hóa sang dạng logarit là việc chuyển hóa sang log giúp nén các quy mô trong đó các biến được đo lường, vì thế giúp giảm sự khác biệt gấp 10 lần giữa hai giá trị xuống còn khác biệt gấp đôi. Ví dụ, số 80 gấp 10 lần số 8, nhưng  $\ln(80)$  (=4.3280) thì chỉ khoảng gấp đôi  $\ln(8)$  (=2.0794).

Một hạn chế của việc sử dụng cách chuyển hóa sang dạng log là chúng ta chỉ có thể lấy log của các con số dương mà thôi.

Hồi quy log của tỷ lệ phá thai theo các biến giải thích trong phương trình (5.1), chúng ta có được kết quả trong Bảng 5.6.

Một cách định tính, các kết quả này tương tự các kết quả trong Bảng 5.1, trong đó các biến price, income, và picket có ý nghĩa thống kê. Tuy nhiên, việc giải thích các hệ số hồi quy thì khác với kết quả trong Bảng 5.1. Các hệ số độ dốc là các **hệ số bán co giãn** (semi-elasticities) – nghĩa là, những thay đổi tương đối trong tỷ lệ phá thai theo một thay đổi đơn vị trong giá trị của biến giải thích<sup>17</sup>. Vì vậy, hệ số của price (-0.003) có nghĩa là nếu giá tăng thêm một đôla, thì thay đổi tương đối (trung bình) trong tỷ lệ phá thai là -0.003 hoặc khoảng -0.3% (khi tất cả các biến

<sup>17</sup> Nhớ lại thảo luận của chúng ta về các mô hình bán log.

khác được giữ nguyên không đổi). Tất cả các hệ số khác cũng được giải thích một cách tương tự<sup>18</sup>.

**Bảng 5.6:** Hồi quy dạng logarit tỷ lệ phá thai.

Dependent Variable: LABORTION				
Method: Least Squares				
Date: 10/09/09 Time: 14:45				
Sample: 1 50				
Included observations: 50				
	Coefficient	Std. Error	t-Statistic	Prob.
C	2.833265	0.755263	3.751362	0.0005
RELIGION	0.000458	0.004327	0.105742	0.9163
PRICE	-0.003112	0.001113	-2.795662	0.0078
LAWS	-0.012884	0.119046	-0.108226	0.9143
FUNDS	0.087688	0.139429	0.628907	0.5328
EDUC	-0.014488	0.009996	-1.449417	0.1546
INCOME	0.000126	2.28E-05	5.546995	0.0000
PICKET	-0.006515	0.002113	-3.083638	0.0036
R-squared	0.589180	Mean dependent var	2.904263	
Adjusted R-squared	0.520710	S.D. dependent var	0.511010	
S.E. of regression	0.353776	Akaike info criterion	0.905342	
Sum squared resid	5.256618	Schwarz criterion	1.211266	
Log likelihood	-14.63355	Durbin-Watson stat	1.929785	
F-statistic	8.604924	Prob(F-statistic)	0.000002	
Note : Labortion = log of abortion				

Khi thực hiện các kiểm định BP và White (không có các số hạng bình phương và tích chéo) của hồi quy này, chúng ta nhận thấy rằng hồi quy này không gặp phải phương sai thay đổi. Một lần nữa, kết quả hồi quy này nên được chấp nhận một cách thận trọng, vì mẫu 50 quan sát của chúng ta có thể là không đủ lớn.

Kết luận này nên lên một điểm quan trọng về các kiểm định phương sai thay đổi. Nếu một hoặc hơn một trong số các kiểm định này chỉ ra rằng chúng ta có vấn đề phương sai thay đổi, thì có thể đó không phải là phương sai thay đổi do bản chất của vấn đề mà là do một lỗi sai dạng mô hình, một chủ đề chúng ta sẽ thảo luận ở chương 7.

## Sai sai số chuẩn điều chỉnh phương sai thay đổi hoặc sai số chuẩn mạnh của White<sup>19</sup>

[Diễn giải: Công thức điều chỉnh như sau (xem Wooldridge, 5<sup>th</sup> edn, p. 271):]

- Bước 1: Hồi quy phương trình sau đây:  

$$Y_i = b_1 + b_2X_2 + b_3X_3 + \dots + b_kX_k + e_i \quad (*)$$
 Lưu phần dư  $e_i$ , tạo biến  $e_i^2$ .
- Hồi quy  $X_j$  theo các biến giải thích còn lại trong mô hình (\*),  $X_j$  lần lượt là  $X_2, X_3, \dots, X_k$ .  
 Ví dụ:  $X_j$  là  $X_2$ :  

$$X_j = a_1 + a_3X_3 + a_4X_4 + \dots + a_kX_k + v_{ij} \quad (*.*)$$

<sup>18</sup> Nhưng nhớ lại rằng cảnh báo được đề cập ở chương trước về cách giải thích các biến giả trong các hồi quy dạng mô hình bán log.

<sup>19</sup> Để biết chi tiết, xem Gujarati/Porter, p. 391.

Lưu phần dư  $v_{ij}$ , tạo biến  $v_{ij}^2$ .

Lưu RSS từ phương trình (\*.\*), đặt tên là  $RSS_j$ , và tạo  $RSS_j^2$ .

- Công thức tính phương sai điều chỉnh phương sai thay đổi của White cho hệ số  $b_j$  (tức  $b_2$ ) như sau:

$$\hat{\sigma}_{b_j}^2 = \frac{\sum_{i=1}^n v_{ij}^2 \cdot e_i^2}{RSS_j^2} \quad (*.*.*)$$

Nếu cỡ mẫu lớn, White đã đề xuất một thủ tục để có được các **sai số chuẩn điều chỉnh phương sai thay đổi** (heteroscedasticity-corrected standard errors). [Xem công thức (\*.\*.\*) ở trên]. Trong lý thuyết, các sai số chuẩn điều chỉnh này được biết với tên gọi là các **sai số chuẩn mạnh** (robust standard errors). Quy trình của White bây giờ đã được lập trình sẵn trong nhiều phần mềm kinh tế lượng. Thủ tục này không làm thay đổi các giá trị của các hệ số hồi quy như được cho trong Bảng 5.2, nhưng điều chỉnh các sai số chuẩn để cho phép phương sai thay đổi. Sử dụng Eviews, chúng ta có các kết quả như được trình bày trong Bảng 5.7.

**Bảng 5.7:** Các sai số chuẩn mạnh của hồi quy tỷ lệ phá thai.

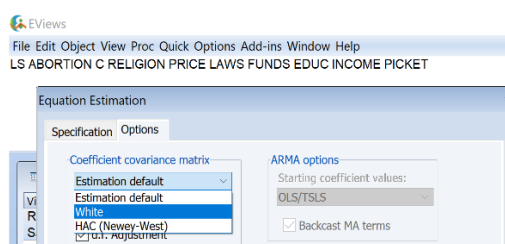
Dependent Variable: ABORTION RATE				
Method: Least Squares				
Sample: 1 50				
Included observations: 50				
White Heteroskedasticity-Consistent Standard Errors & Covariance				
	Coefficient	Std. Error	t-Statistic	Prob.
C	14.28396	14.90146	0.958561	0.3433
RELIGION	0.020071	0.083861	0.239335	0.8120
PRICE	-0.042363	0.025944	-1.632868	0.1100
LAWS	-0.873102	1.795849	-0.486178	0.6294
FUNDS	2.820003	3.088579	0.913042	0.3664
EDUC	-0.287255	0.176628	-1.626329	0.1114
INCOME	0.002401	0.000510	4.705512	0.0000
PICKET	-0.116871	0.040420	-2.891415	0.0060
R-squared	0.577426	Mean dependent var	20.57800	
Adjusted R-squared	0.506997	S.D. dependent var	10.05863	
S.E. of regression	7.062581	Akaike info criterion	6.893145	
Sum squared resid	2094.962	Schwarz criterion	7.199069	
Log likelihood	-164.3286	Durbin-Watson stat	2.159124	
F-statistic	8.198706	Prob(F-statistic)	0.000003	

Nếu bạn so sánh các kết quả này với các kết quả trong Bảng 5.2, bạn sẽ thấy một vài thay đổi. Biến price bây giờ ít ý nghĩa hơn trước, mặc dù các hệ số của income và picket vẫn có cùng mức ý nghĩa. Nhưng lưu ý rằng các hệ số hồi quy ước lượng vẫn giống nhau giữa hai bảng kết quả.

Nhưng đừng quên rằng thủ tục White chỉ có hiệu lực trong các mẫu lớn, nên có thể không đúng trong trường hợp ví dụ hiện tại của chúng ta. Chúng ta hãy xem xét lại trước hết là hàm tiền lương trong Chương 1 và sau đó là hàm số giờ làm việc đã được thảo luận trong Chương 4; cả hai trường hợp mẫu của chúng ta tương đối lớn.



[Diễn giải: Sai số chuẩn điều chỉnh phương sai thay đổi trên Eviews và Stata:]



```
. reg abortion religion price laws funds educ income picket, robust
```

```
Linear regression               Number of obs   =          50
                                F(7, 42)         =          8.88
                                Prob > F           =         0.0000
                                R-squared          =         0.5774
                                Root MSE       =         7.0626
```

abortion	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
religion	.0200709	.0838611	0.24	0.812	-.1491677	.1893095
price	-.0423631	.025944	-1.63	0.110	-.0947202	.009994
laws	-.8731018	1.795849	-0.49	0.629	-4.497272	2.751069
funds	2.820003	3.088579	0.91	0.366	-3.413002	9.053008
educ	-.2872551	.176628	-1.63	0.111	-.6437049	.0691946
income	.0024007	.0005102	4.71	0.000	.0013711	.0034303
picket	-.1168712	.0404201	-2.89	0.006	-.1984422	-.0353002
_cons	14.28396	14.90146	0.96	0.343	-15.78841	44.35632

## Xem xét lại hàm tiền lương

Trong Bảng 1.2, chúng ta đã trình bày một hàm tiền lương với 1.289 công nhân. Vì dữ liệu được sử dụng trong bảng này là dữ liệu chéo, nên rất có thể là kết quả hồi quy bị hiện tượng phương sai thay đổi. Để biết có đúng như thế không, chúng ta sử dụng các kiểm định BP và White, các kiểm định này cho kết quả như sau. [Diễn giải: Dữ liệu là tập tin Table 1.1].

**Bảng 1.2:** Hàm tiền lương.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.183338	1.015788	-7.071691	0.0000
FEMALE	-3.074875	0.364616	-8.433184	0.0000
NONWHITE	-1.565313	0.509188	-3.074139	0.0022
UNION	1.095976	0.506078	2.165626	0.0305
EDUCATION	1.370301	0.065904	20.79231	0.0000
EXPER	0.166607	0.016048	10.38205	0.0000

R-squared	0.323339	Mean dependent var	12.36585
Adjusted R-squared	0.320702	S.D. dependent var	7.896350
S.E. of regression	6.508137	Akaike info criterion	6.588627
Sum squared resid	54342.54	Schwarz criterion	6.612653
Log likelihood	-4240.370	Hannan-Quinn criter.	6.597646
F-statistic	122.6149	Durbin-Watson stat	1.897513
Prob(F-statistic)	0.000000		



#### Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	11.50785	Prob. F(5,1283)	0.0000
Obs*R-squared	55.32705	Prob. Chi-Square(5)	0.0000
Scaled explained SS	254.9588	Prob. Chi-Square(5)	0.0000

Heteroskedasticity Tests

Specification

Test type:

- Breusch-Pagan-Godfrey
- Harvey
- Glejser
- ARCH
- White**
- Custom Test Wizard...

Dependent variable: RESID^2

The White Test regresses the squared residuals on the cross product of the original regressors and a constant.

☐ Include White cross terms

#### Heteroskedasticity Test: White

F-statistic	13.17407	Prob. F(5,1283)	0.0000
Obs*R-squared	62.94665	Prob. Chi-Square(5)	0.0000
Scaled explained SS	290.0715	Prob. Chi-Square(5)	0.0000

Heteroskedasticity Tests

Specification

Test type:

- Breusch-Pagan-Godfrey
- Harvey
- Glejser
- ARCH
- White**
- Custom Test Wizard...

Dependent variable: RESID^2

The White Test regresses the squared residuals on the cross product of the original regressors and a constant.

☒ Include White cross terms

#### Heteroskedasticity Test: White

F-statistic	4.909722	Prob. F(17,1271)	0.0000
Obs*R-squared	79.43116	Prob. Chi-Square(17)	0.0000
Scaled explained SS	366.0357	Prob. Chi-Square(17)	0.0000

**Kiểm định BP:** Khi các phần dư bình phương thu được từ mô hình trong Bảng 1.2 được hồi quy theo các biến giải thích trong hồi quy hàm tiền lương, chúng ta có giá trị  $R^2$  là 0.0429. Nhân giá trị này với số quan sát, 1.289, chúng ta có một giá trị Chi bình phương khoảng 55. Với 5 bậc tự do, tức số biến giải thích trong hàm tiền lương, thì xác suất để có một giá trị bằng hoặc lớn hơn giá trị Chi bình phương đó thực tế là bằng 0, điều này cho thấy rằng hàm tiền lương trong Bảng 1.2 thực sự bị vấn đề phương sai thay đổi.

**Kiểm định White:** Để biết liệu rằng các kết quả kiểm định BP có tin cậy không, chúng ta sử dụng kiểm định White, cả loại trừ và đưa vào các số hạng bình phương và tích chéo. Các kết quả như sau. Khi loại bỏ các số hạng bình phương và tích chéo,  $nR^2 = 62.9466$ , theo phân phối Chi bình phương với 5 bậc tự do. Xác suất để có được một giá trị bằng hoặc lớn hơn giá trị Chi bình phương như thế thực tế bằng 0. Điều này khẳng định rằng hồi quy hàm tiền lương thực sự bị phương sai thay đổi.

**Bảng 5.8:** Hàm tiền lương điều phương sai thay đổi.

Dependent Variable: W Method: Least Squares Sample: 1 1289 Included observations: 1289 White Heteroskedasticity-Consistent Standard Errors & Covariance				
	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.183338	1.090064	-6.589834	0.0000
FEMALE	-3.074875	0.364256	-8.441521	0.0000
NONWHITE	-1.565313	0.397626	-3.936647	0.0001
UNION	1.095976	0.425802	2.573908	0.0102
EDUC	1.370301	0.083485	16.41372	0.0000
EXPER	0.166607	0.016049	10.38134	0.0000
R-squared 0.323339 Mean dependent var 12.36585				
Adjusted R-squared 0.320702 S.D. dependent var 7.896350				
S.E. of regression 6.508137 Akaike info criterion 6.588627				
Sum squared resid 54342.54 Schwarz criterion 6.612653				
Log likelihood -4240.370 Durbin-Watson stat 1.897513				
F-statistic 122.6149 Prob(F-statistic) 0.000000				

Khi chúng ta đưa các số hạng bình phương và tích chéo của các biến giải thích vào phương trình kiểm định, chúng ta có  $nR^2 = 79.4311$ , có phân phối Chi bình phương với 17 bậc tự do (5 biến giải thích, 2 biến bình phương, và 10 số hạng tích chéo của các biến giải thích). Xác suất để có một giá trị bằng hoặc lớn hơn giá trị Chi bình phương như thế thực tế bằng 0.

Tóm lại, chúng ta có bằng chứng mạnh để kết luận rằng hồi quy hàm tiền lương trong Bảng 1.2 gặp phải vấn đề phương sai thay đổi.

Thay vì chuyển hóa hàm tiền lương trong Bảng 1.2 bằng cách chia hai vế cho một hoặc hơn một biến giải thích, chúng ta có thể đơn giản điều chỉnh vấn đề phương sai thay đổi bằng cách tính các sai số chuẩn mạnh theo thủ tục của White. Các kết quả này được trình bày trong Bảng 5.8.

Nếu bạn so sánh các kết quả này với các kết quả trong Bảng 1.2, thì bạn sẽ thấy các hệ số hồi quy giống nhau, nhưng một số các sai số chuẩn đã thay đổi, và điều này làm thay đổi các giá trị t.

### Xem xét lại hàm số giờ làm việc

Xem xét các kết quả được cho trong Bảng 4.3 về số giờ làm việc bởi 753 phụ nữ có gia đình. Các kết quả này không được điều chỉnh phương sai thay đổi. Trên cơ sở các kiểm định BP và White (có hoặc không có các số hạng bình phương và tích chéo), chúng ta

thấy rằng hàm số giờ làm việc trong Bảng 4.3 gặp phải vấn đề phương sai thay đổi<sup>20</sup>. [Diễn giải: Tập tin **Table4.2**].

Vì cỡ mẫu tương đối lớn, nên chúng ta có thể sử dụng thủ tục White để có các sai số chuẩn điều chỉnh phương sai thay đổi. Các kết quả này được trình bày trong Bảng 5.9.

**Bảng 5.9:** Hàm số giờ làm việc điều chỉnh phương sai thay đổi.

reg hours age educ exper faminc hhours hwage kidsl6 wage mtr unemployment if  
hours>0, robust

Linear regression	Number of obs	=	<b>428</b>
	F(10, 417)	=	<b>23.20</b>
	Prob > F	=	<b>0.0000</b>
	R-squared	=	<b>0.3358</b>
	Root MSE	=	<b>640.2</b>

hours	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
age	-17.7274	5.263072	-3.37	0.001	-28.07286	-7.381946
educ	-27.03403	15.70405	-1.72	0.086	-57.903	3.834947
exper	24.20345	4.95372	4.89	0.000	14.46607	33.94082
faminc	.0137806	.0078976	1.74	0.082	-.0017434	.0293046
hhours	-.4864745	.0732871	-6.64	0.000	-.6305326	-.3424163
hwage	-144.9734	17.58257	-8.25	0.000	-179.5349	-110.4119
kidsl6	-180.4415	105.0628	-1.72	0.087	-386.9603	26.07729
wage	-47.43286	9.832833	-4.82	0.000	-66.76096	-28.10476
mtr	-6351.293	1206.585	-5.26	0.000	-8723.04	-3979.547
unemployment	-16.50367	9.632981	-1.71	0.087	-35.43892	2.431585
_cons	8484.524	1154.479	7.35	0.000	6215.2	10753.85

Nếu bạn so sánh các kết quả này với các kết quả trong Bảng 4.3, bạn sẽ thấy một số thay đổi trong các sai số chuẩn ước lượng và các giá trị t. Các biến thu nhập gia đình và số con dưới 6 tuổi bây giờ ít ý nghĩa hơn trước đây, trong khi đó biến tỷ lệ thất nghiệp lại có ý nghĩa hơn.

Điểm cần lưu ý ở đây là nếu cỡ mẫu tương đối lớn, chúng ta nên trình bày các sai số chuẩn được điều chỉnh phương sai thay đổi theo thủ tục của White cùng với các sai số chuẩn của hồi quy OLS thông thường để có ý tưởng gì đó về sự hiện diện của phương sai thay đổi.

<sup>20</sup> Đối với kiểm định BP,  $nR^2 = 38.76$ , có phân phối Chi bình phương với 10 bậc tự do. Xác suất để có giá trị bằng hoặc lớn hơn giá trị Chi bình phương như thế thực tế bằng 0. Đối với kiểm định White,  $nR^2 = 40.19$  khi không có các số hạng bình phương và tích chéo, và  $= 120.23$  khi các số hạng đó được đưa vào phương trình kiểm định. Trong cả hai trường hợp, xác suất để có một giá trị bằng hoặc lớn hơn giá trị Chi bình phương như thế thực tế bằng 0.

Bảng 4.3: Hồi quy số giờ làm việc của phụ nữ

Source	SS	df	MS	Number of obs	=	428
				F(10, 417)	=	21.08
Model	86401477.5	10	8640147.75	Prob > F	=	0.0000
Residual	170909542	417	409855.018	R-squared	=	0.3358
				Adj R-squared	=	0.3199
Total	257311020	427	602601.92	Root MSE	=	640.2

hours	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-17.7274	4.903114	-3.62	0.000	-27.3653	-8.089503
educ	-27.03403	15.79456	-1.71	0.088	-58.0809	4.01285
exper	24.20345	4.653332	5.20	0.000	15.05653	33.35036
faminc	.0137806	.0058661	2.35	0.019	.0022499	.0253113
hhours	-.4864745	.0704622	-6.90	0.000	-.6249799	-.347969
hwage	-144.9734	15.88407	-9.13	0.000	-176.1962	-113.7506
kids16	-180.4415	86.3696	-2.09	0.037	-350.2156	-10.66743
wage	-47.43286	10.30926	-4.60	0.000	-67.69745	-27.16826
mtr	-6351.293	1029.837	-6.17	0.000	-8375.613	-4326.974
unemployment	-16.50367	10.55941	-1.56	0.119	-37.25997	4.252637
_cons	8484.524	987.5952	8.59	0.000	6543.238	10425.81

## 5.5 Tóm tắt và kết luận

Trong chương này chúng ta đã xem xét một trong số các vi phạm về giả định của mô hình hồi quy tuyến tính cổ điển, đó là phương sai thay đổi, thường gặp trong dữ liệu chéo. Mặc dù phương sai thay đổi không phá hủy các tính chất không chệch và nhất quán của các ước lượng OLS, nhưng các ước lượng OLS kém hiệu quả hơn, làm cho việc suy diễn thống kê trở nên kém tin cậy nếu chúng ta không điều chỉnh các sai số chuẩn của OLS thông thường.

Trước khi chúng ta giải quyết vấn đề phương sai thay đổi, chúng ta cần tìm hiểu xem liệu rằng chúng ta có khó khăn gì trong bất kỳ ứng dụng cụ thể nào hay không. Đối với mục đích này, chúng ta có thể phân tích phần dư bình phương từ mô hình gốc hoặc sử dụng vài kiểm định chính thức về phương sai thay đổi, chẳng hạn kiểm định Breusch-Pagan và kiểm định White. Nếu một hoặc hơn một trong số các kiểm định này cho thấy có tồn tại phương sai thay đổi, thì chúng ta có thể tiến hành sửa chữa vấn đề này.

Vấn đề phương sai thay đổi có thể được giải quyết nếu chúng ta biết các phương sai thay đổi, tức các  $\sigma_i^2$ , vì trong trường hợp đó chúng ta có thể chuyển hóa mô hình gốc (5.1) bằng cách chia cả hai vế cho  $\sigma_i$  và ước lượng mô hình đã được chuyển hóa bằng OLS, khi đó chúng ta sẽ có các ước lượng BLUE. Phương pháp ước lượng này được biết với tên gọi là bình phương bé nhất có trọng số (WLS). Thật không may, chúng ta hiếm khi biết các phương sai thực của hạng nhiễu. Vì thế chúng ta cần giải pháp tốt thứ nhì.

Sử dụng các dự toán về bản chất có thể có của  $\sigma_i^2$  chúng ta chuyển hóa mô hình gốc, ước lượng mô hình đó, rồi thực hiện các kiểm định phương sai thay đổi cho mô hình được chuyển hóa. Nếu các kiểm định này cho thấy rằng không có vấn đề phương sai thay đổi trong mô hình được chuyển hóa, thì chúng ta không thể bác bỏ mô hình được chuyển hóa. Tuy nhiên, nếu mô hình được chuyển hóa cho thấy rằng vấn đề phương sai thay đổi vẫn còn tồn tại, thì chúng ta có thể tìm một cách chuyển hóa khác và lặp lại chu trình một lần nữa.

Tuy nhiên, có thể đỡ mất công sức nếu chúng ta có một mẫu đủ lớn, bởi vì trong trường hợp đó chúng ta có thể sử dụng các sai số chuẩn điều chỉnh phương sai thay đổi theo thủ tục của White. Các sai số chuẩn điều chỉnh này được biết với tên gọi là các sai số chuẩn mạnh. Ngày nay, nhiều bộ dữ liệu vi mô do nhiều cơ quan thực hiện có số lượng quan sát lớn, điều này giúp cho việc sử dụng các sai số chuẩn mạnh trong các mô hình hồi quy nếu có hoài nghi vấn đề phương sai thay đổi./.