
CHƯƠNG 11

PHƯƠNG SAI THAY ĐỔI

Phương sai thay đổi (Heteroscedasticity, còn gọi là phương sai của sai số thay đổi) chưa bao giờ là lý do để loại bỏ một mô hình mà nếu không vì lý do này mô hình sẽ rất tốt.*

Nhưng cũng không thể bỏ qua phương sai thay đổi!

Tác giả

Một giả thiết quan trọng trong mô hình hồi quy tuyến tính cổ điển (Giả thiết 4) là các yếu tố nhiễu u_i xuất hiện trong hàm hồi quy tổng thể có phương sai không thay đổi (homoscedasticity, còn gọi là phương sai có điều kiện không đổi); tức là chúng có cùng phương sai. Trong chương này, ta xem xét giá trị của giả thiết này và tìm xem điều gì sẽ xảy ra nếu giả thiết này không được thỏa mãn. Giống như trong Chương 10, ta tìm câu trả lời cho các câu hỏi sau:

1. Đây là bản chất của phương sai thay đổi?
2. Đây là những hậu quả của nó?
3. Làm sao phát hiện ra nó?
4. Đây là các biện pháp sửa chữa vấn đề này?

11.1 BẢN CHẤT CỦA PHƯƠNG SAI THAY ĐỔI

Như đã nêu trong Chương 3, một trong số các giả thiết quan trọng của mô hình hồi quy tuyến tính cổ điển là phương sai của từng yếu tố nhiễu u_i , tùy theo giá trị lựa chọn của các biến giải thích, là một số không đổi, bằng σ^2 . Đây là giả thiết về phương sai không thay đổi (homoscedasticity), hay là khoảng *chênh lệch* (scedasticity) *bằng nhau* (homo), tức là, *phương sai bằng nhau*. Về ký hiệu, ta có:

$$E(u_i^2) = \sigma^2 \quad i = 1, 2, \dots, n \quad (11.1.1)$$

* N. Gregory Mankiw, "A Quick Refresher Course in Macroeconomics", *Journal of Economic Literature* (Khóa học giới thiệu nhanh về kinh tế vĩ mô, Tạp chí Kinh tế), tập XXVIII, 12/1990, trang 1648.

Bằng đồ thị, trong mô hình hồi quy hai biến, phương sai không thay đổi có thể được biểu diễn trong Hình 3.4. Để thuận lợi, hình này được vẽ lại trong Hình 11.1. Như Hình 11.1 biểu diễn, phương sai có điều kiện của Y_i (bằng với phương sai của u_i), tùy thuộc vào giá trị cho trước của X_i , không đổi khi biến X nhận các giá trị khác nhau.

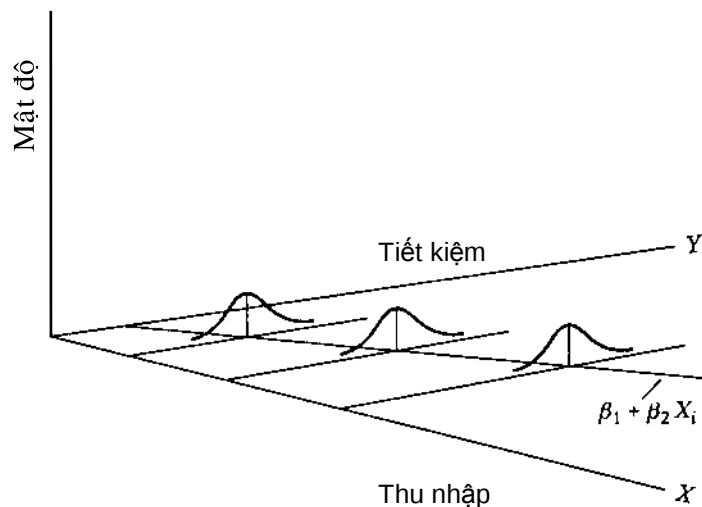
Trái lại, xem Hình 11.2 trong đó cho thấy phương sai có điều kiện của Y_i tăng lên khi X tăng. Ở đây, các phương sai của Y_i không giống nhau. Do vậy có phương sai thay đổi. Về ký hiệu ta có:

$$E(u_i^2) = \sigma_i^2 \quad (11.1.2)$$

Chú ý tới chỉ số ở dưới của σ^2 . Nó làm ta nhớ lại rằng các phương sai có điều kiện u_i (= các phương sai có điều kiện của Y_i) không còn là hằng số nữa.

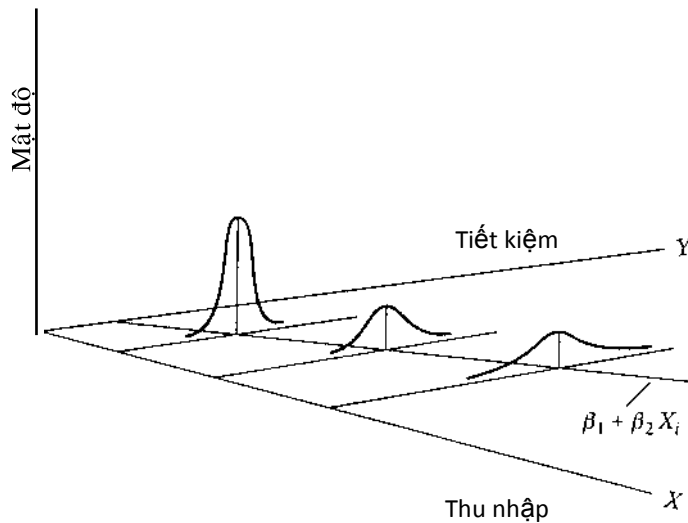
Để làm rõ sự khác nhau giữa phương sai không thay đổi và phương sai thay đổi, giả sử rằng trong mô hình hai biến $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y đại diện cho tiết kiệm và X đại diện cho thu nhập. Hình 11.1 và 11.2 cho thấy rằng khi thu nhập tăng lên, tiết kiệm tính trung bình cũng tăng lên. Nhưng trong Hình 11.1 phương sai của tiết kiệm không đổi ở tất cả các mức thu nhập, trái lại trong Hình 11.2 phương sai tăng lên theo thu nhập. Dường như trong Hình 11.2 các gia đình có thu nhập cao hơn, tính một cách trung bình, tiết kiệm nhiều hơn các gia đình có thu nhập thấp, nhưng độ biến thiên trong tiết kiệm của họ cũng cao hơn.

Có một vài lý do tại sao các phương sai của u_i có thể thay đổi. Sau đây là một số lý do.¹



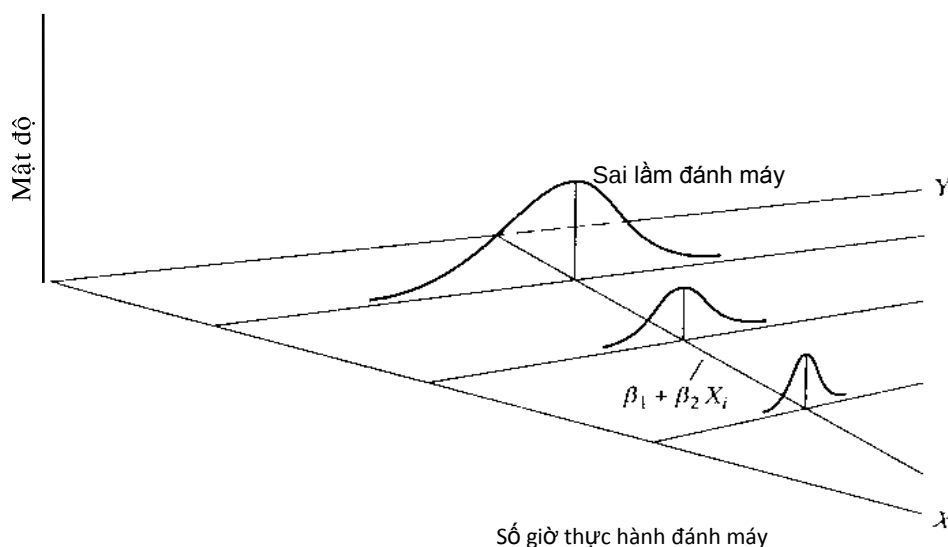
HÌNH 11.1 Các nhiễu có phương sai không thay đổi

¹ Xem Stefan Valavanis, *Econometrics* (Kinh tế lượng), McGraw-Hill, New York, 1959, trang 48.

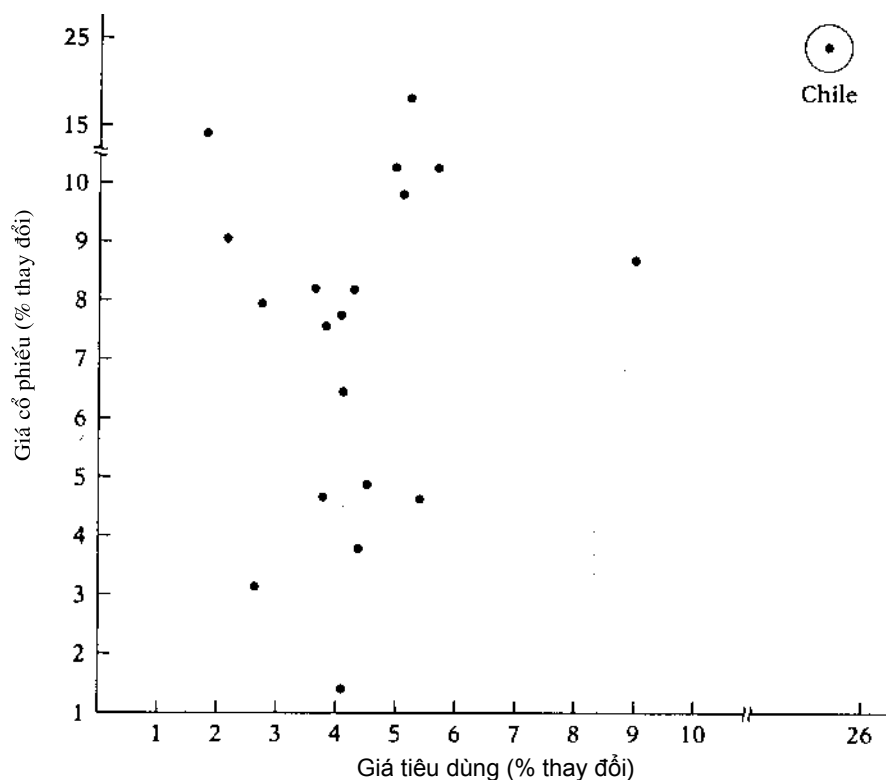
**HÌNH 11.2** Các nhiễu có phương sai thay đổi

1. Theo các *mô hình học tập -sai lầm*, khi mọi người học hỏi, các sai lầm về hành vi của họ ngày càng nhỏ đi theo thời gian. Trong trường hợp này, σ_i^2 được dự kiến là sẽ giảm dần. Ví dụ, xem Hình 11.3 trong đó biểu diễn quan hệ giữa sai sót đánh máy xảy ra trong một khoảng thời gian cho trước với số giờ thực tập đánh máy. Như hình 11.3 mô tả, khi số giờ thực tập đánh máy tăng lên, số các sai sót đánh máy trung bình cũng như phương sai của chúng giảm xuống.
2. Khi thu nhập tăng lên, người dân có nhiều *thu nhập tự định* hơn² và phạm vi lựa chọn về việc sử dụng thu nhập cũng tăng lên. Vì vậy, σ_i^2 có nhiều khả năng tăng lên với thu nhập. Do vậy, trong hồi quy tiết kiệm - thu nhập, ta sẽ thấy σ_i^2 tăng lên theo thu nhập (như trong Hình 11.2) do người dân có nhiều lựa chọn hơn về hành vi tiết kiệm của mình. Tương tự, các công ty có nhiều lợi nhuận hơn thường cho thấy có nhiều biến thiên hơn trong chính sách trả cổ tức so với các công ty có lợi nhuận thấp. Cũng như vậy, công ty có hướng phát triển thì có khả năng cho thấy có nhiều biến thiên lớn trong tỷ lệ trả cổ tức so với công ty có mức độ phát triển không đổi.
3. Khi các kỹ thuật thu nhập số liệu được cải thiện, σ_i^2 có nhiều khả năng giảm. Như vậy, các ngân hàng có thiết bị xử lý số liệu phức tạp thường phạm ít sai lầm trong các báo cáo hàng tháng hay hàng quý về khách hàng của họ hơn là các ngân hàng không có các phương tiện này.

² Như Valavanis trình bày, “Thu nhập tăng lên, và người dân bây giờ không còn tính từng đồng nữa, trái lại trước đây, họ tính từng hào”, *ibid.*, trang 48.

**HÌNH 11.3** Minh họa phương sai thay đổi

4. Phương sai thay đổi cũng có thể nảy sinh do sự hiện diện của **yếu tố tách biệt** (outlier, còn gọi là yếu tố nằm ngoài). Một quan sát nằm ngoài là một quan sát rất khác (có thể rất nhỏ hay rất lớn) với các quan sát khác trong mẫu. Việc bao gồm hay loại trừ một quan sát như thế, đặc biệt là nếu như cỡ mẫu nhỏ, có thể làm thay đổi đáng kể các kết quả phân tích hồi quy. Ví dụ, hãy xem xét đồ thị phân tán trong Hình 11.4. Dựa vào số liệu trong bài tập 11.20, hình 11.4 vẽ tốc độ thay đổi phần trăm của giá cổ phiếu (Y) và giá tiêu dùng (X) trong giai đoạn sau Chiến tranh Thế giới thứ II tới 1969 cho 20 nước. Trong hình vẽ này, quan sát về Y và X của Chile có thể được coi như là một quan sát tách biệt bởi vì các giá trị Y và X của Chile lớn hơn nhiều so với các nước còn lại. Trong các trường hợp này, khó có thể duy trì giả thiết về phương sai không thay đổi. Trong bài tập 11.20 bạn được yêu cầu tìm xem điều gì xảy ra đối với các kết quả hồi quy nếu các quan sát của Chile được loại bỏ khỏi phân tích.
5. Một nguồn tạo ra phương sai thay đổi nữa nảy sinh từ việc vi phạm Giả thiết 9 của mô hình hồi quy tuyến tính cổ điển (CLRM) rằng mô hình hồi quy được xác định một cách đúng đắn. Mặc dù ta sẽ thảo luận nội dung các sai số đặc trưng đầy đủ hơn trong Chương 13, thường thì cái mà có vẻ như phương sai thay đổi có thể là do một số biến quan trọng bị loại bỏ khỏi mô hình. Như vậy, trong hàm cầu một hàng hóa, nếu ta không đưa giá cả của các hàng hóa bổ sung hay cạnh tranh với mặt hàng xem xét vào mô hình (thiên lệch của biến bị loại bỏ), các phần dư thu được từ hồi quy có thể cho thấy một ấn tượng rõ nét rằng phương sai của sai số có thể không cố định. Nhưng nếu các biến bỏ sót được đưa vào mô hình, ấn tượng đó có thể biến mất.



HÌNH 11.4 Quan hệ giữa giá cổ phiếu và giá tiêu dùng.

Lưu ý rằng vấn đề phương sai thay đổi thường phổ biến hơn trong số liệu chéo so với số liệu chuỗi thời gian. Trong số liệu chéo, người ta thường làm việc với các thành viên của một tổng thể tại một thời điểm, như người tiêu dùng riêng biệt hay gia đình họ, công ty, ngành kinh tế, hay khu vực địa lý như bang, quốc gia, thành phố, v.v... Hơn nữa, các thành viên này có thể có quy mô khác nhau như công ty quy mô nhỏ, vừa hay lớn, hay thu nhập thấp, vừa hay cao. Mặt khác, trong số liệu chuỗi thời gian, các biến có xu hướng có thứ tự về độ lớn giống nhau do người ta thường thu thập số liệu của cùng một đối tượng trong một khoảng thời gian. Ví dụ như GNP, chỉ tiêu tiêu dùng, tiết kiệm, hay việc làm tại Hoa Kỳ trong giai đoạn 1950-1994.

Để minh họa cho vấn đề phương sai thay đổi thường xảy ra trong phân tích số liệu chéo, hãy xem Bảng 11.1. Bảng này cho ta số liệu về lương bình quân một lao động trong 10 ngành công nghiệp chế tạo sản phẩm không lâu bền, phân loại theo quy mô lao động của doanh nghiệp hay cơ cấu tổ chức trong năm 1958. Bảng 11.1 cũng cho ta số liệu về năng suất bình quân của 9 nhóm quy mô lao động.

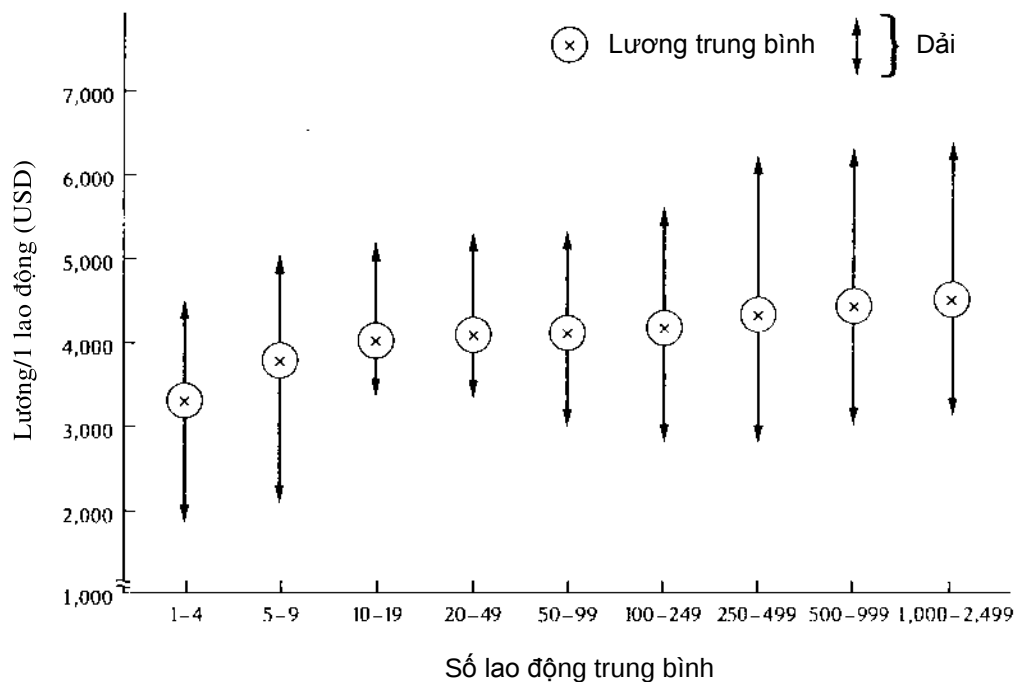
Mặc dù các ngành khác nhau về cơ cấu sản lượng, Bảng 11.1 cho thấy rõ ràng, tính một cách trung bình, các công ty lớn trả lương cao hơn các công ty nhỏ. Ví dụ, các công ty sử dụng từ 1 đến 4 lao động trả trung bình khoảng 4843 USD. Nhưng lưu ý rằng có biến thiên đáng kể trong thu nhập giữa các phân loại quy mô lao động khác nhau. Điều này được biểu thị bằng độ lệch chuẩn ước lượng của thu nhập. Nó cũng được nhận thấy từ số liệu đi cùng, trong đó biểu thị dải thu nhập trong từng nhóm quy mô lao động. Như Hình 11.4 mô tả, dải (giá trị cao nhất - giá

trị thấp nhất), một đại lượng thô về mức độ biến thiên, thay đổi theo nhóm quy mô lao động, từ đó cho thấy phương sai thay đổi trong thu nhập của các nhóm quy mô lao động khác nhau.

11.2 ƯỚC LƯỢNG OLS KHI CÓ SỰ HIỆN DIỆN CỦA PHƯƠNG SAI THAY ĐỔI

Điều gì xảy ra đối với các ước lượng OLS và phương sai của chúng nếu ta đưa vào mô hình phương sai thay đổi bằng cách cho $E(u_i^2) = \sigma_i^2$ nhưng giữ nguyên tất cả các giả thiết khác của mô hình cổ điển? Để trả lời câu hỏi này, ta hãy quay lại với mô hình hai biến:

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



HÌNH 11.5 Lương bình quân một lao động trong quan hệ với quy mô lao động.

Áp dụng công thức thông thường, ước lượng OLS của β_2 là:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum x_i y_i}{\sum x_i^2} \\ &= \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}\end{aligned}\quad (11.2.1)$$

BẢNG 11.1**Mức lương lao động (USD) trong các ngành công nghiệp chế tạo sản phẩm không lâu bền theo quy mô lao động của doanh nghiệp, 1985**

Ngành công nghiệp	Quy mô lao động (số công nhân trung bình)								
	1-4	5-9	10-19	20-49	50-99	100-249	250-499	500-999	1000-2499
Thực phẩm và các sản phẩm tương tự	2.994	3.295	3.565	3.907	4.189	4.486	4.676	4.968	5.342
Sản phẩm thuốc lá	1.721	2.057	3.336	3.320	2.980	2.848	3.072	2.969	3.822
Sản phẩm dệt	3.600	3.657	3.674	3.437	3.340	3.334	3.225	3.163	3.168
Máy mặc và các sản phẩm liên quan	3.494	3.787	3.533	3.215	3.030	2.834	2.750	2.967	3.453
Giấy và các sản phẩm từ gỗ	3.498	3.847	3.913	4.135	4.445	4.885	5.132	5.342	5.326
In và xuất bản	3.611	4.206	4.695	5.083	5.301	5.269	3.182	5.395	5.552
Hóa chất và các sản phẩm tương tự	3.875	4.660	4.930	5.005	5.114	5.248	5.630	5.870	5.876
Sản phẩm dầu lửa và than	4.616	5.181	5.317	5.337	5.421	5.710	6.316	6.455	6.347
Sản phẩm cao su và nhựa	3.538	3.984	4.014	4.287	4.221	4.539	4.721	4.905	5.481
Da và sản phẩm từ da	3.016	3.196	3.149	3.317	3.414	3.254	3.177	3.346	4.067
Lương lao động trung bình	3.396	3.787	4.013	4.014	4.146	4.241	4.387	4.538	4.843
Độ lệch chuẩn	743,7	851,4	727,8	805,1	929,9	1080,6	1243,2	1307,7	1112,5
Năng suất trung bình	9.355	8.584	7.962	8.275	8.389	9.418	9.795	10.281	11.750

Nguồn: The Census of Manufacturers (Tổng điều tra ngành công nghiệp chế tạo, Bộ Thương mại Hoa Kỳ, 1958 (đã tính toán)).

nhưng phương sai của nó bây giờ được biểu diễn bởi công thức sau (xem Phụ lục 11A, Mục 11A.1):

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (11.2.2)$$

Rõ ràng công thức này khác với công thức phương sai thông thường trong trường hợp có giả thiết về phương sai không thay đổi; cụ thể là:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad (11.2.3)$$

Tất nhiên, nếu $\sigma_i^2 = \sigma^2$ đối với mỗi i , hai công thức sẽ đồng nhất. (Tại sao?)

Nhớ lại rằng $\hat{\beta}_2$ là ước lượng tuyến tính không thiên lệch tốt nhất (BLUE) nếu giả thiết của mô hình cổ điển, bao gồm cả giả thiết về phương sai không thay đổi, được thỏa mãn. Ước lượng này có còn là ước lượng tuyến tính không thiên lệch tốt nhất nữa không khi ta bỏ giả thiết về phương sai không thay đổi và thay nó bằng giả thiết về phương sai thay đổi? Ta có thể dễ dàng chứng minh rằng $\hat{\beta}_2$ vẫn là tuyến tính và không thiên lệch. Trên thực tế, như trình bày trong Phụ lục 3A, Mục 3A.2, để tạo sự không thiên lệch của $\hat{\beta}_2$, không nhất thiết là các yếu tố nhiễu (u_i) phải có phương sai không thay đổi. Thực tế, phương sai của u_i , không thay đổi hay thay đổi, không có vai trò trong việc xác định tính chất không thiên lệch.

Với điều kiện $\hat{\beta}_2$ vẫn tuyến tính không thiên lệch, nó có “hiệu quả” và “tốt nhất”, tức là nó có phương sai nhỏ nhất trong nhóm các ước lượng tuyến tính không thiên lệch hay không? Và phương sai trong Phương trình (11.2.2) có phải là phương sai nhỏ nhất không? Câu trả lời là không cho cả hai câu hỏi: $\hat{\beta}_2$ không còn là phương sai tốt nhất và phương sai nhỏ nhất không phải là (11.2.2). Vậy, đâu là ước lượng tuyến tính không thiên lệch tốt nhất trong trường hợp có sự hiện diện của phương sai thay đổi? Câu trả lời được giải quyết trong mục sau.

11.3 PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU TỔNG QUÁT (GLS)

Tại sao ước lượng β_2 theo phương pháp bình phương tối thiểu thông thường (OLS) không phải là tốt nhất, mặc dù nó vẫn không bị thiên lệch? Về trực giác, ta có thể nhận thấy lý do từ Hình 11.5. Như Hình này mô tả, mức thu nhập giữa các nhóm quy mô lao động có độ biến thiên đáng kể. Nếu ta thực hiện hồi quy mức lương bình quân một lao động theo quy mô lao động, ta sẽ thấy cần sử dụng kiến thức cho rằng thu nhập có tính biến thiên lớn giữa các nhóm. Một cách lý tưởng, ta muốn thiết kế một chương trình ước lượng qua đó các quan sát từ tổng thể với độ biến thiên cao sẽ có trọng số thấp hơn những quan sát từ tổng thể có độ biến thiên nhỏ hơn. Xem xét Hình 11.5, ta sẽ cho trọng số lớn hơn đối với các quan sát từ các nhóm quy mô lao động như 10-19 và 20-49 so với các quan sát từ nhóm quy mô lao động như 5-9 và 250-499 bởi vì các quan sát từ nhóm quy mô lao động như 10-19 và 20-49 phân bố gần các giá trị trung bình của chúng hơn, và từ đó cho phép ta ước lượng hàm hồi quy tổng thể (PRF) chính xác hơn.

Tuy nhiên, phương pháp OLS thông thường không tuân theo cách làm này và do vậy không sử dụng “thông tin” về tính biến thiên không bằng nhau của biến phụ thuộc Y , như lương

lao động trong Hình 11.5: Phương pháp OLS cho mỗi quan sát các trọng số hay tầm quan trọng như nhau. Nhưng một phương pháp ước lượng, gọi là **phương pháp bình phương tối thiểu tổng quát (GLS)**, đưa các thông tin này vào mô hình và do vậy có khả năng đưa ra các ước lượng tuyến tính không thiên lệch tốt nhất (BLUE). Để tìm hiểu xem điều này được thực hiện như thế nào, hãy tiếp tục với mô hình hai biến đã quen thuộc:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (11.3.1)$$

Để dễ dàng hơn cho việc biến đổi đại số, ta viết (11.3.1) dưới dạng

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \quad (11.3.2)$$

với $X_{0i} = 1$ với mỗi i . Người đọc có thể nhận thấy rằng hai công thức này đồng nhất với nhau.

Bây giờ giả thiết rằng các phương sai thay đổi σ_i^2 đã biết. Chia 2 vế (11.3.2) cho σ_i , ta có:

$$\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i} \right) + \beta_2 \left(\frac{X_i}{\sigma_i} \right) + \left(\frac{u_i}{\sigma_i} \right) \quad (11.3.3)$$

Để dễ trình bày, ta viết (11.3.3) dưới dạng

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \quad (11.3.4)$$

với các biến sao hay đã biến đổi là các biến ban đầu chia cho σ_i đã biết. Ta sử dụng ký hiệu β_1^* và β_2^* , các tham số của mô hình đã biến đổi, để phân biệt chúng với các tham số OLS thông thường β_1 và β_2 .

Mục đích của việc biến đổi mô hình gốc là gì? Để tìm hiểu, chú ý tới đặc điểm sau của sai số đã biến đổi u_i^* :

$$\begin{aligned} \text{var}(u_i^*) &= E(u_i^*)^2 = E\left(\frac{u_i}{\sigma_i}\right)^2 \\ &= \frac{1}{\sigma_i^2} E(u_i^2) && \text{do } \sigma_i^2 \text{ đã biết} \\ &= \frac{1}{\sigma_i^2} (\sigma_i^2) && \text{do } E(u_i^2) = \sigma_i^2 \\ &= 1 \end{aligned} \quad (11.3.5)$$

và không đổi. Tức là, phương sai của yếu tố nhiễu đã biến đổi u_i^* bây giờ là phương sai không thay đổi. Do ta vẫn giữ nguyên các giả thiết khác của mô hình cổ điển, kết quả u^* có phương sai không thay đổi cho thấy nếu ta áp dụng OLS đối với mô hình đã biến đổi (11.3.3), nó sẽ cho ta các ước lượng BLUE. Nói ngắn gọn, β_1^* và β_2^* bây giờ là ước lượng BLUE chứ không phải các ước lượng OLS $\hat{\beta}_1$ và $\hat{\beta}_2$.

Phép biến đổi các biến gốc để các biến đã biến đổi thỏa mãn các giả thiết của mô hình cổ điển và sau đó áp dụng phương pháp OLS đối với chúng được gọi là phương pháp bình phương tối thiểu tổng quát. *Nói ngắn gọn, GLS là OLS đối với các biến đã biến đổi để thỏa mãn các giả thiết bình phương tối thiểu tiêu chuẩn.* Các ước lượng tính được như vậy được gọi là các **ước lượng GLS**, và chính các ước lượng này mới có tính chất BLUE.

Phương pháp thực tế để ước lượng β_1^* và β_2^* được trình bày như sau. Thứ nhất, ta viết hàm hồi quy mẫu (SRF) của (11.3.3)

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i} \right) + \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i} \right) + \left(\frac{\hat{u}_i}{\sigma_i} \right)$$

hay

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \quad (11.3.6)$$

Bây giờ, để tính các ước lượng GLS, ta tối thiểu hóa

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

tức là,

$$\sum \left(\frac{\hat{u}_i}{\sigma_i} \right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i} \right) - \hat{\beta}_1^* \left(\frac{X_{0i}}{\sigma_i} \right) - \hat{\beta}_2^* \left(\frac{X_i}{\sigma_i} \right) \right]^2 \quad (11.3.7)$$

Cách thức thực tế để tối thiểu hóa (11.3.7) tuân theo kỹ thuật giải tích tiêu chuẩn và được trình bày trong Phụ lục 11A, Mục 11A.2. Như trình bày trong phụ lục, ước lượng β_2^* theo GLS là:

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \quad (11.3.8)$$

và phương sai của nó được tính bởi

$$\text{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \quad (11.3.9)$$

với $w_i = 1/\sigma_i^2$.

Sự khác nhau giữa OLS và GLS

Nhớ lại từ Chương 3 rằng trong OLS ta tối thiểu hóa

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \quad (11.3.10)$$

nhưng trong GLS ta tối thiểu hóa (11.3.7). (11.3.7) cũng có thể được viết như sau:

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \quad (11.3.11)$$

với $w_i = 1/\sigma_i^2$ [kiểm chứng rằng (11.3.11) và (11.3.7) đồng nhất].

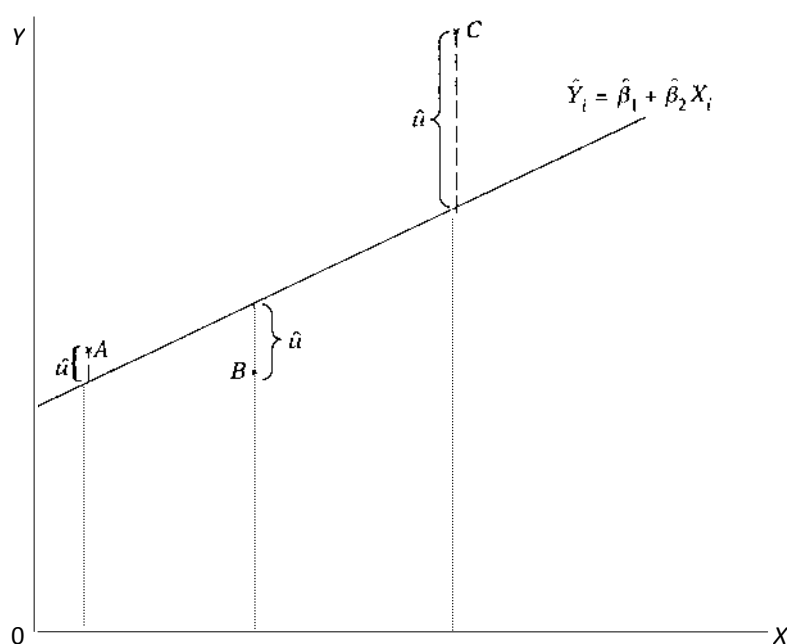
Như vậy, trong GLS ta tối thiểu hóa *tổng bình phương phần dư có trọng số* với $w_i = 1/\sigma_i^2$ đóng vai trò các trọng số, nhưng trong OLS ta tối thiểu hóa RSS không có trọng số hay (tương đương với) các trọng số bằng nhau. Như (11.3.7) biểu thị, trong GLS trọng số của từng quan sát tỷ lệ nghịch với giá trị σ_i của nó, tức là trong quá trình tối thiểu hóa RSS(11.3.11), các quan sát

từ một tổng thể với giá trị σ_i lớn hơn sẽ có trọng số tương đối nhỏ hơn và những quan sát từ tổng thể với giá trị σ_i nhỏ hơn sẽ có trọng số lớn hơn theo tỷ lệ. Để nhìn nhận sự khác nhau giữa OLS và GLS một cách rõ ràng hơn, hãy xem xét đồ thị phân tán giả thiết trong Hình 11.6.

Trong phương pháp OLS (không có trọng số), từng giá trị $\hat{u}_i^2$ gắn với các điểm A, B, C sẽ nhận cùng một trọng số trong quá trình tối thiểu hóa RSS. Rõ ràng là trong trường hợp này giá trị $\hat{u}_i^2$ gắn với điểm C sẽ chiếm ưu thế trong RSS. Nhưng trong GLS quan sát thái cực C sẽ có trọng số tương đối nhỏ hơn so với hai quan sát kia. Như đã nêu ở trên, đây là phương pháp đúng đắn bởi vì để ước lượng hàm hồi quy tổng thể (PRF) một cách tin cậy hơn, ta cho trọng số lớn hơn đối với các quan sát nằm gần xung quanh giá trị trung bình (tổng thể) của chúng so với các quan sát nằm rải rác ở xa.

Do (11.3.11) tối thiểu hóa RSS có tính trọng số, nó được gọi là **bình phương tối thiểu có trọng số (WLS)**, và các ước lượng tính được trong phép toán này và trình bày trong (11.3.8) và (11.3.9) được gọi là các **ước lượng WLS**. Nhưng WLS chỉ là một trường hợp đặc biệt của kỹ thuật ước lượng tổng quát hơn, GLS. Trong bối cảnh có phương sai thay đổi, người ta coi hai khái niệm WLS và GLS tương đương nhau. Trong các chương sau, ta sẽ tiếp cận với các trường hợp đặc biệt khác của GLS.

Trước khi chuyển sang phần tiếp theo, chú ý rằng nếu $w_i = w =$ hằng số với mọi i , $\hat{\beta}_2^*$ đồng nhất với $\hat{\beta}_2$ và $\text{var}(\hat{\beta}_2^*)$ đồng nhất với phương sai thông thường (nghĩa là phương sai không thay đổi) $\text{var}(\hat{\beta}_2)$ tính theo công thức (11.2.3). Điều này hoàn toàn không có gì ngạc nhiên. (Tại sao?) (Xem bài tập 11.8).



HÌNH 11.6 Đồ thị phân tán giả thiết

11.4 CÁC HẬU QUẢ CỦA VIỆC SỬ DỤNG OLS KHI CÓ SỰ HIỆN DIỆN CỦA PHƯƠNG SAI THAY ĐỔI

Như ta đã thấy, cả $\hat{\beta}_2^*$ và $\hat{\beta}_2$ đều là các ước lượng (tuyến tính) không thiên lệch: trong việc lấy mẫu lặp lại, tính một cách trung bình, $\hat{\beta}_2^*$ và $\hat{\beta}_2$ sẽ bằng với giá trị đúng của β_2 ,³ tức là, cả hai là các ước lượng không thiên lệch. Nhưng ta biết rằng chỉ có $\hat{\beta}_2^*$ là hiệu quả, tức là, có phương sai nhỏ nhất. Điều gì xảy ra với khoảng tin cậy, kiểm định giả thiết và các thủ tục khác của chúng ta nếu ta tiếp tục sử dụng ước lượng OLS của $\hat{\beta}_2$? Ta phân biệt hai trường hợp.

Ước lượng OLS có tính tới phương sai thay đổi

Giả sử ta sử dụng $\hat{\beta}_2$ và sử dụng công thức phương sai trong (11.2.2), trong đó có tính đến vấn đề phương sai thay đổi. Sử dụng phương sai này, và giả sử đã biết σ_i^2 , ta có thể thiết lập các khoảng tin cậy và các giả thiết kiểm định với các kiểm định t và F thông thường được không? Câu trả lời chung là không bởi vì ta có thể chỉ ra rằng $\text{var}(\hat{\beta}_2^*) \leq \text{var}(\hat{\beta}_2)$,⁴ có nghĩa là các khoảng tin cậy, dựa vào $\text{var}(\hat{\beta}_2)$ sẽ lớn một cách không cần thiết. Do vậy, các kiểm định t và F có nhiều khả năng cho ta các kết quả không chính xác bởi vì $\text{var}(\hat{\beta}_2)$ quá lớn và cái mà có thể là một hệ số không có ý nghĩa về thống kê (do giá trị t sẽ nhỏ hơn mức thích hợp) có thể trên thực tế lại có ý nghĩa nếu các khoảng tin cậy đúng được thiết lập trên cơ sở của phương pháp GLS.

Ước lượng OLS không tính đến phương sai thay đổi

Tình hình trở nên rất nghiêm trọng nếu ta không chỉ sử dụng $\hat{\beta}_2$ mà còn tiếp tục sử dụng công thức phương sai thông thường (phương sai không thay đổi) trong (11.2.3) thậm chí nếu phương sai thay đổi tồn tại hay được nghi là tồn tại: Chú ý rằng thường thì hay xảy ra trường hợp thứ hai, bởi vì chạy chương trình hồi quy OLS chuẩn và bỏ qua (hay không chú ý) tới phương sai thay đổi sẽ cho ta phương sai $\hat{\beta}_2$ như trong (11.2.3). Trước hết, $\text{var}(\hat{\beta}_2)$ trong (11.2.3) là ước lượng thiên lệch của $\text{var}(\hat{\beta}_2)$ trong (11.2.2), tức là, tính một cách trung bình, nó ước lượng quá cao hay quá thấp $\text{var}(\hat{\beta}_2)$ trong (11.2.2), và nói chung ta không thể nói thiên lệch là dương (ước lượng quá cao) hay âm (ước lượng quá thấp) bởi vì nó phụ thuộc vào bản chất của quan hệ giữa σ_i^2 và các giá trị của biến X , như có thể thấy rõ từ (11.2.2) (xem bài tập 11.9). Thiên lệch phát sinh từ thực tế là $\hat{\sigma}^2$, ước lượng quy ước của σ^2 , cụ thể là, $\sum \hat{u}_i^2 / (n-2)$ không còn là ước không thiên lệch của σ^2 nữa khi có phương sai thay đổi. Do vậy, ta không thể dựa vào các khoảng tin cậy và các kiểm định t và F tính theo quy ước nữa.⁵ **Nói ngắn gọn, nếu ta tiếp tục sử dụng các thủ tục**

³ Cũng có thể chỉ ra rằng cả $\hat{\beta}_2^*$ và $\hat{\beta}_2$ đều là các ước lượng nhất quán, tức là, chúng hội tụ ở giá trị đúng của β_2 khi kích thước mẫu n tăng lên vô hạn.

⁴ Có thể xem chứng minh chính thức trong Phoebus J. Dhrymes, *Introductory Econometrics* (Kinh tế lượng giới thiệu), Springer-Verlag, New York, 1978, trang 110-111. Nhân đây, lưu ý rằng sự mất hiệu quả của $\hat{\beta}_2$ [nghĩa là $\text{var}(\hat{\beta}_2) > \text{var}(\hat{\beta}_2^*)$ bao nhiêu] phụ thuộc vào các giá trị mẫu của các biến X và giá trị của σ_i^2 .

⁵ Từ (5.3.6) ta biết rằng khoảng tin cậy 100(1- α)% đối với β_2 là $[\hat{\beta}_2 \pm t_{\alpha/2} \text{se}(\hat{\beta}_2)]$. Nhưng nếu $\text{se}(\hat{\beta}_2)$ không thể được ước lượng một cách không thiên lệch, ta có thể tin tưởng gì vào khoảng tin cậy tính theo quy ước?

kiểm định thông thường mặc dù có phương sai thay đổi, những kết luận hay sự suy diễn có thể dẫn ta tới sai lầm.

Để làm sáng tỏ thêm nội dung này, ta tham khảo nghiên cứu **Monte Carlo** do Davidson và MacKinnon thực hiện.⁶ Họ xem xét mô hình đơn giản sau, theo cách ký hiệu của ta là

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (11.4.1)$$

Họ giả sử rằng $\beta_1 = 1$, $\beta_2 = 1$, và $u_i \sim N(0, X_i^\alpha)$. Như biểu thức sau cùng cho thấy, phương sai của sai số thay đổi và quan hệ với giá trị của biến hồi quy độc lập X lũy thừa α . Ví dụ, nếu $\alpha = 1$, phương sai của sai số tỷ lệ với giá trị của X ; nếu $\alpha = 2$, phương sai của sai số tỷ lệ với bình phương giá trị của X , và v.v... Trong Mục 11.6 ta sẽ xem xét logic đằng sau phương pháp này. Dựa vào việc lặp lại 20.000 lần và cho phép α nhận các giá trị khác nhau, họ tính được các sai số chuẩn của hai hệ số hồi quy sử dụng OLS [xem Phương trình (11.2.3)], OLS có tính tới phương sai thay đổi [xem Phương trình (11.2.2)] và GLS [xem Phương trình (11.3.9)]. Ta tóm tắt các kết quả với các giá trị lựa chọn của α trong bảng sau:

Giá trị α	Sai số chuẩn của $\hat{\beta}_1$			Sai số chuẩn của $\hat{\beta}_2$		
	OLS	OLS _{het.}	GLS	OLS	OLS _{het.}	GLS
0,5	0,164	0,134	0,110	0,285	0,277	0,243
1,0	0,142	0,101	0,048	0,246	0,247	0,173
2,0	0,116	0,074	0,0073	0,200	0,220	0,109
3,0	0,100	0,064	0,0013	0,173	0,206	0,056
4,0	0,089	0,059	0,0003	0,154	0,195	0,017

Chú ý: OLS_{het.} có nghĩa là OLS có tính tới phương sai thay đổi.

Đặc điểm nổi bật nhất của các kết quả này là OLS, có hiệu chỉnh hay không hiệu chỉnh khi nảy sinh phương sai thay đổi, đều luôn ước lượng quá cao giá trị đúng của sai số chuẩn tính bởi phương pháp GLS (đúng đắn), đặc biệt là đối với các giá trị cao hơn của α , và do vậy cho thấy rõ tính ưu việt của GLS. Các kết quả này cũng cho thấy rằng nếu ta không sử dụng GLS và dựa vào OLS – có tính đến hay không tính đến vấn đề thay đổi – bức tranh là phức hợp. Các sai số chuẩn OLS thông thường hoặc quá lớn (đối với tung độ gốc) hoặc thường quá nhỏ (đối với hệ số góc) trong quan hệ với các giá trị tính được bằng OLS có tính tới phương sai thay đổi. Thông điệp rất rõ ràng: Trong trường có phương sai thay đổi, ta sử dụng GLS. Tuy nhiên, vì các lý do giải thích ở phần sau của chương, trên thực tế, không phải lúc nào cũng có thể áp dụng GLS một cách dễ dàng.

Từ thảo luận ở trên, rõ ràng là phương sai thay đổi là một vấn đề khó khăn nghiêm trọng về tiềm năng và các nhà nghiên cứu cần phải biết trong một tình huống cụ thể vấn đề này có xuất hiện hay không. Nếu sự hiện diện của nó được phát hiện, ta có thể đưa ra biện pháp chỉnh sửa, như sử dụng hồi quy bình phương tối thiểu có trọng số hay kỹ thuật khác. Tuy nhiên, trước khi quay lại xem xét các phương pháp hiệu chỉnh khác nhau, trước hết ta phải tìm xem phương sai thay đổi có xuất hiện hay có nhiều khả năng xuất hiện trong một trường hợp cụ thể hay không. Nội dung này được thảo luận trong phần sau.

⁶ Russell Davidson & James G. MacKinnon, *Estimation and Inference in Econometrics* (Ước lượng và Suy luận trong Kinh tế Lượng), Oxford University Press, New York, 1993, trang 549-550.

11.5 PHÁT HIỆN PHƯƠNG SAI THAY ĐỔI

Cũng như trường hợp đa cộng tuyến, câu hỏi quan trọng trên thực tế là: Làm sao ta biết được có phương sai thay đổi trong một tình huống cụ thể? Cũng như trong đa cộng tuyến, không có các quy tắc bất di bất dịch để phát hiện ra phương sai thay đổi mà chỉ có vài quy tắc kinh nghiệm. Nhưng đây là tình huống không thể tránh được do ta chỉ có thể biết được σ_i^2 nếu có toàn bộ tổng thể Y tương ứng với các giá trị X , như là tổng thể trình bày trong Bảng 2.1 hay Bảng 11.1. Nhưng những số liệu đó là một ngoại lệ chứ không phải là một quy tắc sử dụng trong hầu hết các điều tra kinh tế. Về mặt này, các nhà kinh tế lượng khác với các nhà khoa học trong các lĩnh vực như nông nghiệp và sinh học. Trong các lĩnh vực đó, nhà nghiên cứu có nhiều khả năng kiểm soát các đối tượng của mình. Thường thì trong nghiên cứu kinh tế, chỉ có một giá trị mẫu của Y tương ứng với một giá trị cụ thể của X . Và không có cách nào mà ta có thể biết được σ_i^2 chỉ từ một quan sát của Y . Do vậy, trong phần lớn các trường hợp liên quan tới điều tra kinh tế lượng, phương sai thay đổi có lẽ là vấn đề trực giác, khả năng dự đoán qua rèn luyện, kinh nghiệm thực nghiệm có trước, hay suy đoán tuyệt đối.

Với vấn đề trình bày trước ở trên, hãy xem xét một số phương pháp không chính thức và chính thức để phát hiện phương sai thay đổi. Như thảo luận sau đây sẽ làm sáng tỏ, phần lớn các phương pháp được dựa vào việc xem xét các phần dư OLS $\hat{u}_i$ do chúng là những yếu tố mà ta quan sát, chứ không phải yếu tố nhiễu u_i . Người ta hy vọng rằng chúng là những ước lượng tốt của u_i . Hy vọng này sẽ là hiện thực nếu cỡ mẫu khá lớn.

Các phương pháp không chính thức

Bản chất của vấn đề. Thường thì bản chất của vấn đề đang xem xét gợi ý cho ta về phương sai thay đổi có khả năng xảy ra hay không. Ví dụ, tiếp tục nghiên cứu công trình tiên phong của Prais và Houthakker về ngân sách gia đình, trong đó họ tìm ra rằng phương sai phần dư xung quanh hồi quy của tiêu dùng đối với thu nhập tăng theo thu nhập, bây giờ người ta giả thiết tổng quát rằng trong các điều tra tương tự ta có thể kỳ vọng rằng các phương sai không bằng nhau giữa các yếu tố nhiễu.⁷ Trên thực tế, trong số liệu chéo liên quan tới các đơn vị không đồng nhất, phương sai thay đổi có thể là quy luật chứ không phải là trường hợp ngoại lệ. Như vậy, trong phân tích số liệu chéo liên quan tới chi đầu tư trong quan hệ với doanh thu, mức lãi suất, v.v..., phương sai thay đổi thường xảy ra nếu các công ty quy mô nhỏ, vừa và lớn được đưa vào cùng trong một mẫu.

Phương pháp đồ thị. Nếu không có một sự tiên nghiệm hay thông tin thực nghiệm về bản chất của phương sai thay đổi, trên thực tế người ta có thể thực hiện phân tích hồi quy dựa trên giả thiết là không có phương sai thay đổi và sau đó thực hiện kiểm tra sau khi chạy hồi quy về phần dư bình phương $\hat{u}_i^2$ để xem chúng có biểu thị một mẫu hình hệ thống không. Mặc dù $\hat{u}_i^2$ không đồng nhất với u_i^2 , chúng có thể được sử dụng làm số gần đúng đặc biệt nếu cỡ mẫu đủ lớn.⁸ Một sự thẩm tra $\hat{u}_i^2$ có thể cho ta thấy được các mẫu hình như trong Hình 11.7.

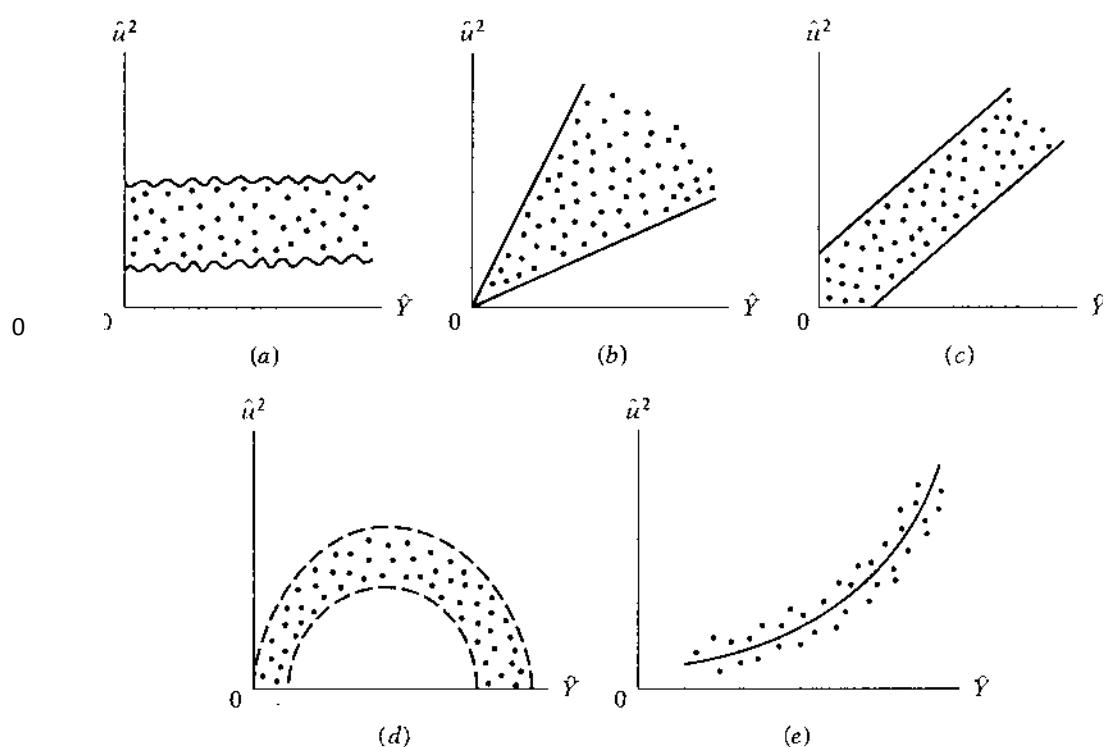
⁷ S. J. Prais & H. X. Houthakker, *The Analysis of Family Budgets* (Phân tích Ngân sách Gia đình), Cambridge University Press, New York, 1955.

⁸ Về quan hệ giữa $\hat{u}_i^2$ và u_i^2 , xem E. Malinvaud, *Statistical Methods of Econometrics* (Các Phương pháp Thống kê của Kinh tế Lượng), North Holland Publishing Company, Amsterdam, 1970, trang 88-89.

Trong Hình 11.7, $\hat{u}_i^2$ được vẽ trong quan hệ với $\hat{Y}_i$, giá trị ước lượng của Y_i từ đường hồi quy. Ý tưởng là để tìm xem giá trị trung bình ước lượng của Y có quan hệ một cách hệ thống với phần dư bình phương hay không. Trong Hình 11.7a ta thấy rằng không có một mẫu hình hệ thống giữa hai biến, từ đó gợi ý rằng có lẽ không có sự hiện diện của phương sai thay đổi trong số liệu. Tuy nhiên, các hình 11.7b đến e cho thấy một mẫu hình cụ thể. Ví dụ, Hình 11.7c đưa ra một quan hệ tuyến tính, trái lại Hình 11.7d và e chỉ ra quan hệ bậc hai giữa $\hat{u}_i^2$ và $\hat{Y}_i$. Sử dụng kiến thức này, mặc dù không chính thức, người ta có thể chuyển số liệu theo cách thích hợp để các số liệu sau khi biến đổi không còn phương sai thay đổi. Trong Mục 11.6, ta sẽ xem xét một số phép biến đổi này.

Thay cho việc vẽ $\hat{u}_i^2$ theo $\hat{Y}_i$, người ta có thể vẽ chúng theo một biến giải thích, đặc biệt nếu như việc vẽ $\hat{u}_i^2$ theo $\hat{Y}_i$ tạo ra mẫu hình như Hình 11.7a. Một đồ thị vẽ như trình bày trong Hình 11.8, có thể cho ta các mẫu hình giống như trong Hình 11.7. (Trong trường hợp mô hình hai biến, vẽ $\hat{u}_i^2$ theo $\hat{Y}_i$ tương đương với việc vẽ nó theo X_i , và do vậy mà Hình 11.8 giống với Hình 11.7. Nhưng điều này sẽ khác đi khi ta xem xét mô hình có hai hay nhiều biến giải thích X ; trong ví dụ này, $\hat{u}_i^2$ có thể được vẽ theo mọi biến X có trong mô hình).

Ví dụ, một mẫu hình như trong Hình 11.8c, cho thấy rằng phương sai của yếu tố nhiễu có quan hệ tuyến tính với biến X . Như vậy, nếu trong hồi quy tiết kiệm - thu nhập, ta tìm thấy một mẫu hình như trong Hình 11.8c thì điều đó cho thấy rằng phương sai thay đổi có thể *tỷ lệ* với giá trị của biến thu nhập. Thông tin này có thể trợ giúp ta trong phép biến đổi số liệu theo cách để trong hồi quy đối với số liệu đã biến đổi, phương sai của yếu tố nhiễu sẽ không thay đổi. Ta sẽ quay trở lại nội dung này trong phần sau.



HÌNH 11.7 Các mẫu hình giả thiết của phần dư bình phương ước lượng

Các phương pháp chính thức

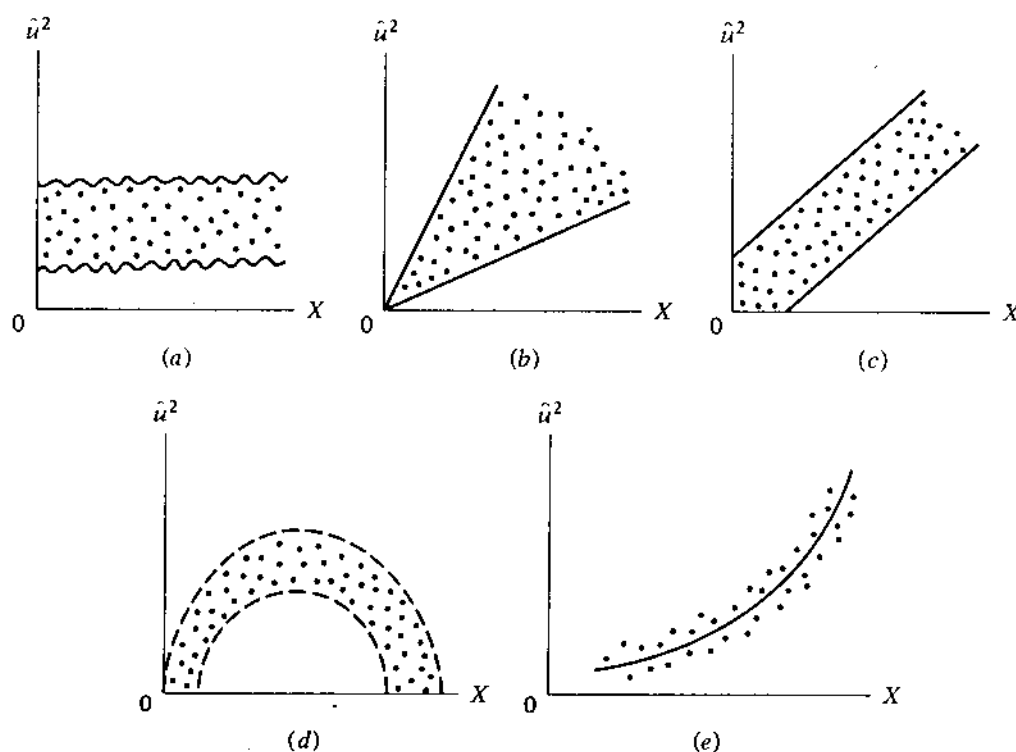
Kiểm định Park.⁹ Park hợp thức hóa phương pháp đồ thị bằng cách cho rằng σ_i^2 là một hàm số của biến giải thích X_i . Dạng hàm số mà Park gợi ý là

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

hay

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \quad (11.5.1)$$

với v_i là số hạng nhiễu ngẫu nhiên.



HÌNH 11.8 Đồ thị phân tán của các phần dư bình phương ước lượng theo X .

Do ta thường không biết được σ_i^2 , Park đề nghị sử dụng $\hat{u}_i^2$ như một biến thay thế và thực hiện hồi quy sau:

$$\begin{aligned} \ln \hat{u}_i^2 &= \ln \sigma^2 + \beta \ln X_i + v_i \\ &= \alpha + \beta \ln X_i + v_i \end{aligned} \quad (11.5.2)$$

Nếu β có ý nghĩa về mặt thống kê, nó sẽ cho thấy là có phương sai thay đổi trong số liệu. Nếu β không có ý nghĩa, ta có thể chấp nhận giả thiết về phương sai không thay đổi. Như vậy, kiểm định Park là một quy trình hai bước. Trong bước một, ta chạy hồi quy OLS mà không xem xét

⁹ R. E. Park, "Estimation with Heteroscedastic Error Terms" (Ước lượng với các số hạng sai số có phương sai thay đổi), *Econometrica*, tập 34, số 4, 10/1966, trang 888. Kiểm định Park là một trường hợp đặc biệt của kiểm định tổng quát do A. C. Harvey đề xuất trong "Estimating Regression Models with Multiplicative Heteroscedasticity" (Ước lượng các mô hình hồi quy với phương sai thay đổi tích), *Econometrica*, tập 44, số 3, 1976, trang 461-465.

tới phương sai thay đổi. Ta tính được $\hat{u}_i$ từ hồi quy này, và sau đó trong bước hai ta chạy hồi quy (11.5.2).

Mặc dù có ý nghĩa về mặt thực nghiệm, kiểm định Park có một số vấn đề khó khăn. Goldfeld và Quandt đã lập luận rằng sai số v_i trong (11.5.2) có thể không thỏa mãn các giả thiết OLS và bản thân nó có thể là phương sai thay đổi.¹⁰ Tuy vậy, người ta có thể sử dụng kiểm định Park như một phương pháp giải thích chặt chẽ.

Ví dụ 11.1. Quan hệ giữa lương và năng suất. Để minh họa phương pháp Park, ta sử dụng số liệu trong Bảng 11.1 để chạy hồi quy sau:

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

Với Y = lương trung bình tính theo nghìn USD, X = năng suất trung bình tính theo nghìn USD, và i = quy mô lao động thứ i của tổ chức kinh tế. Sau đây là các kết hồi quy:

$$\begin{aligned} Y_i &= 1992,3452 + 0,2329X_i \\ se &= (936,4791) \quad (0,0998) \\ t &= (2,1275) \quad (2,333) \quad R^2 = 0,4375 \end{aligned} \quad (11.5.3)$$

Các kết quả cho thấy hệ số độ dốc ước lượng có ý nghĩa ở mức 5% trên cơ sở của kiểm định t một phía. Phương trình cho thấy khi năng suất lao động tăng lên, ví dụ, 1 USD, lương (thù lao lao động) tính trung bình tăng lên 23 xen.

Các phần dư nhận được từ hồi quy (11.5.3) được tính hồi quy theo X_i như trong Phương trình (11.5.2). Từ đó ta có các kết quả sau:

$$\begin{aligned} \ln \hat{u}_i^2 &= 35,817 - 2,8099 \ln X_i \\ se &= (38,319) \quad (4,216) \\ t &= (0,934) \quad (-0,667) \quad R^2 = 0,0595 \end{aligned} \quad (11.5.4)$$

Rõ ràng, không có quan hệ có ý nghĩa thống kê giữa hai biến. Theo kiểm định Park, ta có thể kết luận rằng không có phương sai thay đổi trong phương sai của sai số.¹¹

Kiểm định Glejser.¹² Kiểm định Glejser tương tự như tinh thần của kiểm định Park. Sau khi tính được các phần dư $\hat{u}_i$ từ hồi quy OLS, Glejser thực hiện hồi quy các giá trị tuyệt đối của $\hat{u}_i$ theo biến X , các giá trị này được cho rằng có quan hệ mật thiết với σ_i^2 . Trong các thử nghiệm của mình, Glejser sử dụng các dạng hàm số sau:

$$\begin{aligned} |\hat{u}_i| &= \beta_1 + \beta_2 X_i + v_i \\ |\hat{u}_i| &= \beta_1 + \beta_2 \sqrt{X_i} + v_i \\ |\hat{u}_i| &= \beta_1 + \beta_2 \frac{1}{X_i} + v_i \end{aligned}$$

¹⁰ Stephen M. Goldfeld & Richard E. Quandt, *Nonlinear Methods in Econometrics* (Các Phương pháp Phi tuyến trong Kinh tế Lượng), North Holland Publishing Company, Amsterdam, 1972, trang 93-94.

¹¹ Dạng hàm cụ thể do Park lựa chọn chỉ là đề gợi ý. Một dạng hàm khác có thể tạo ra quan hệ có ý nghĩa. Ví dụ, ta có thể sử dụng $\hat{u}_i^2$ thay cho $\ln \hat{u}_i^2$ làm biến phụ thuộc.

¹² H. Glejser, "A New Test for Heteroscedasticity", *Journal of the American Statistical Association* (Một kiểm định mới về phương sai thay đổi, Tạp chí của Hiệp hội Thống kê Hoa Kỳ), tập 64, 1969, trang 316-323.

$$\begin{aligned} |\hat{u}_i| &= \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i \\ |\hat{u}_i| &= \sqrt{\beta_1 + \beta_2 X_i} + v_i \\ |\hat{u}_i| &= \sqrt{\beta_1 + \beta_2 X_i^2} + v_i \end{aligned}$$

với v_i là số hạng sai số.

Một lần nữa, trên khía cạnh thực nghiệm hay thực tiễn, ta có thể sử dụng phương Glejser. Nhưng Goldfeld và Quandt chỉ ra rằng số hạng sai số v_i có một số vấn đề theo đó giá trị kỳ vọng của v_i khác không, v_i có tương quan chuỗi (xem Chương 12), và khá châm biếm là nó lại có phương sai thay đổi.¹³ Một khó khăn nữa của phương pháp Glejser là các mô hình như

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i \text{ và } |\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

có các tham số phi tuyến và do vậy không thể ước lượng bằng quy trình OLS.

Glejser tìm thấy rằng đối với các mẫu lớn, 4 mô hình đầu tiên trong số các mô hình ở trên cho ta các kết quả nói chung là thỏa đáng trong việc phát hiện ra phương sai thay đổi. Do vậy, về mặt thực tiễn, phương pháp Glejser có thể được sử dụng cho các mẫu lớn. Đối với các mẫu nhỏ thì nó chỉ được xem như là một phương tiện định tính để tìm hiểu về phương sai thay đổi. Về ứng dụng phương pháp Glejser, xem Mục 11.7.

Kiểm định tương quan theo hạng của Spearman (Spearman rank correlation test). Trong bài tập 3.8, ta định nghĩa hệ số tương quan hạng của Spearman như sau:

$$r_s = 1 - 6 \left[\frac{\sum d_i^2}{n(n^2 - 1)} \right] \quad (11.5.5)$$

với d_i = sự khác nhau giữa các hạng quy cho hai đặc điểm khác nhau của thành phần hay hiện tượng thứ i và n = số các thành phần hay hiện tượng được xếp thứ hạng. Hệ số tương quan hạng ở trên có thể được sử dụng để phát hiện ra phương sai thay đổi như sau: Giả sử $Y_i = \beta_0 + \beta_1 X_i + u_i$.

Bước 1. Thực hiện hồi quy Y theo X và tính phần dư $\hat{u}_i$.

Bước 2. Bỏ qua dấu của $\hat{u}_i$, tức là, lấy giá trị tuyệt đối $|\hat{u}_i|$, xếp thứ hạng cả $|\hat{u}_i|$ lẫn X_i (hay Y_i) theo thứ tự tăng lên hay giảm xuống và tính hệ số tương quan hạng Spearman theo công thức trên.

Bước 3. Giả thiết rằng hệ số tương quan hạng của tổng thể ρ_s bằng không và $n > 8$, ý nghĩa của r_s mẫu có thể được kiểm định bằng kiểm định t như sau:¹⁴

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \quad (11.5.6)$$

với số bậc tự do = $n - 2$.

¹³ Về chi tiết, xem Goldfeld & Quandt, op. Cit., Chương 3.

¹⁴ Xem G. Udny Yule & M. G Kendall, *An Introduction to the Theory of Statistics* (Giới thiệu Lý thuyết Thống kê), Charles Griffin & Company, London, 1953, trang 455.

Nếu giá trị t tính được lớn hơn giá trị t tới hạn, ta có thể chấp nhận giả thiết về phương sai thay đổi; ngược lại ta có thể bác bỏ nó. Nếu mô hình hồi quy có nhiều hơn một biến X , r_s có thể được tính giữa $|\hat{u}_i|$ và từng biến X một cách riêng rẽ và có thể được kiểm định về ý nghĩa thống kê bằng kiểm định t theo như Phương trình (11.5.6).

Ví dụ 11.2. Minh họa kiểm định tương quan hạng. Để minh họa kiểm định tương quan hạng, xem xét số liệu trong Bảng 11.2 là một mẫu con lấy từ số liệu trong bảng thuộc bài tập 5.16, trong đó yêu cầu bạn ước lượng đường thị trường vốn của lý thuyết cơ cấu đầu tư chứng khoán, cụ thể, $E_i = \beta_1 + \beta_2 \sigma_i$, với E là suất sinh lợi kỳ vọng của cơ cấu đầu tư và σ là độ lệch chuẩn của suất sinh lợi. Do số liệu liên quan tới 10 quỹ hỗ tương có quy mô và mục tiêu đầu tư khác nhau, trong sự tiên nghiệm ta có thể dự kiến có phương sai thay đổi. Để kiểm định giả thiết này, ta áp dụng phương pháp tương quan hạng. Các phép tính cần thiết cũng được chỉ trong Bảng 11.2.

Áp dụng công thức (11.5.5), ta có:

$$\begin{aligned} r_s &= 1 - 6 \frac{110}{10(100-1)} \\ &= 0,3333 \end{aligned} \quad (11.5.7)$$

Áp dụng kiểm định t trong (11.5.6), ta có:

$$\begin{aligned} t &= \frac{(0,3333)(\sqrt{8})}{\sqrt{1-0,1110}} \\ &= 0,9998 \end{aligned} \quad (11.5.8)$$

Với 8 bậc tự do, giá trị t không có ý nghĩa ở mức ý nghĩa 10%; **giá trị p** bằng 0,17. Như vậy, không có bằng chứng về mối quan hệ hệ thống giữa biến giải thích và các giá trị tuyệt đối của phần dư. Điều này có thể cho thấy rằng không có phương sai thay đổi.

BẢNG 11.2

Kiểm định tương quan hạng về phương sai thay đổi

Tên quỹ	E_i , suất sinh lợi bq năm %	σ_i , độ lệch chuẩn của suất sinh lợi/năm	$\hat{E}_i^*$	$ \hat{u}_i \uparrow$ phần dư, $\left (E_i - \hat{E}_i) \right $	Thứ hạng của $ \hat{u}_i $	Thứ hạng của σ_i	d , hiệu số giữa hai thứ hạng	d^2
Boston Fund	12,4	12,1	11,37	1,03	9	4	5	25
Delaware Fund	14,4	21,4	15,64	1,24	10	9	1	1
Equity Fund	14,6	18,7	14,40	0,20	4	7	-3	9
Fundamental Investors	16,0	21,7	15,78	0,22	5	10	-5	25
Investors Mutual	11,3	12,5	11,56	0,26	6	5	1	1
Loomis-Sales Mutual Fund	10,0	10,4	10,59	0,59	7	2	5	25
Massachusetts Investors Trust	16,2	20,8	15,37	0,83	8	8	0	0
New England Fund	10,4	10,2	10,50	0,10	3	1	2	4
Putnam Fund of Boston	13,1	16,0	13,16	0,06	2	6	-4	16
Wellington Fund	11,3	12,0	11,33	0,03	1	3	-2	4
Total							0	110

* Tính từ hồi quy: $\hat{E}_i = 5,8194 + 0,4590\sigma_i$.

† Giá trị tuyệt đối của các phần dư.

Chú ý: Thứ tự được xếp theo giá trị tăng dần.

Kiểm định Goldfeld-Quandt.¹⁵ Phương pháp phổ biến này được áp dụng nếu ta giả sử rằng phương sai thay đổi, σ_i^2 , quan hệ đồng biến với một trong số các biến giải thích trong mô hình hình hồi quy. Để đơn giản hóa, ta hãy xem xét mô hình hai biến thông thường sau:

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

Giả sử σ_i^2 quan hệ đồng biến với X_i như sau

$$\sigma_i^2 = \sigma^2 X_i^2 \quad (11.5.9)$$

với σ^2 là hằng số.¹⁶

Giả thiết (11.5.9) mặc định rằng σ_i^2 tỷ lệ thuận với bình phương của biến X . Prais và Houthakker trong nghiên cứu ngân sách gia đình của họ đã nhận thấy rằng giả thiết như vậy là thích hợp. (Xem Mục 11.6).

Nếu (11.5.9) thích hợp, điều đó có nghĩa là σ_i^2 có thể lớn hơn khi giá trị của X_i lớn hơn. Nếu trường hợp này xảy ra, phương sai thay đổi có rất nhiều khả năng hiện diện trong mô hình. Để kiểm định một cách rõ ràng, Goldfeld và Quandt đưa ra các bước sau:

- Bước 1.** Xếp thứ tự hay sắp hạng các quan sát theo các giá trị của X_i , bắt đầu từ giá trị thấp nhất của X .
- Bước 2.** Loại bỏ quan sát nằm ở giữa, với c được xác định là một tiên nghiệm, và chia $(n - c)$ quan sát còn lại làm hai nhóm, mỗi nhóm chứa $(n - c)/2$ quan sát.
- Bước 3.** Thiết lập các hồi quy OLS riêng rẽ cho $(n - c)/2$ quan sát đầu và cuối, và tính tổng các bình phương phần dư RSS_1 và RSS_2 tương ứng, RSS_1 đại diện cho RSS từ hồi quy tương ứng với các giá trị nhỏ hơn của X_i (nhóm phương sai nhỏ) và RSS_2 đại diện cho RSS từ hồi quy tương ứng với các giá trị lớn hơn của X_i (nhóm phương sai lớn). Mỗi RSS này có:

$$\frac{(n - c)}{2} - k \quad \text{hay} \quad \left(\frac{n - c - 2k}{2} \right) \text{df (bậc tự do)}$$

với k là số các tham số cần ước lượng, bao gồm cả tung độ gốc. (Tại sao?) Trong trường hợp hai biến, tất nhiên là k bằng 2.

- Bước 4.** Tính tỷ lệ

$$\lambda = \frac{RSS_2/\text{df}}{RSS_1/\text{df}} \quad (11.5.10)$$

¹⁵ Goldfeld & Quandt, op. cit., Chương 3.

¹⁶ Đây chỉ là một giả thiết hợp lý. Trên thực tế, điều kiện cần thiết là σ_i^2 có quan hệ đơn điệu với X_i .

Nếu u_i được giả thiết là có phân phối chuẩn (một giả thiết mà ta thường đưa ra), và nếu giả thiết về phương sai thay đổi có hiệu lực, thì có thể chỉ ra rằng λ trong (11.5.10) tuân theo phân phối F với số bậc tự do (df) tử số và mẫu số là $(n - c - 2k)/2$.

Nếu trong một ứng dụng, giá trị $\lambda (= F)$ tính được lớn hơn F tới hạn tại mức ý nghĩa đã chọn, ta có thể bác bỏ giả thiết về phương sai không thay đổi, tức là, ta có thể nói rằng phương sai thay đổi có rất nhiều khả năng tồn tại.

Trước khi minh họa kiểm định này, ta cần giải thích về việc loại bỏ c quan sát ở giữa. Các quan sát này được loại bỏ để làm rõ hay tăng thêm sự khác biệt giữa nhóm phương sai nhỏ (nghĩa là RSS_1) và nhóm phương sai lớn (nghĩa là RSS_2). Nhưng khả năng của kiểm định Goldfeld-Quandt trong việc thực hiện điều này một cách thành công phụ thuộc vào việc c được lựa chọn như thế nào.¹⁷ Đối với mô hình hai biến, các thử nghiệm Monte Carlo do Goldfeld và Quandt thực hiện cho thấy c vào khoảng 8 nếu cỡ mẫu khoảng 30, và khoảng 16 nếu cỡ mẫu vào khoảng 60. Nhưng Judge và các đồng tác giả lưu ý rằng $c = 4$ nếu $n = 30$ và $c = 10$ nếu n vào khoảng 60 là thỏa đáng trên thực tiễn.¹⁸

Trước khi tiếp tục thảo luận, có thể cần lưu ý rằng trong trường hợp có nhiều hơn một biến X trong mô hình, xếp thứ hạng của các quan sát, tức là bước một trong kiểm định, có thể được thực hiện theo bất cứ biến nào trong số các biến X . Như vậy trong mô hình: $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$, ta có thể xếp thứ tự số liệu theo mọi biến trong số các biến X . Nếu trong tiên nghiệm, ta không biết chắc biến X nào là thích hợp, ta có thể thực hiện kiểm định đối với từng biến X , hay bằng kiểm định Park đối với từng biến X .

Ví dụ 11.3. Kiểm định Goldfeld-Quandt. Để minh họa kiểm định Goldfeld-Quandt, ta trình bày trong Bảng 11.3 số liệu chéo của 30 gia đình về chi tiêu tiêu dùng trong quan hệ với thu nhập. Giả sử ta mặc định rằng chi tiêu tiêu dùng quan hệ tuyến tính với thu nhập nhưng số liệu có phương sai thay đổi. Ta mặc định thêm rằng bản chất của phương sai thay đổi giống như trong (11.5.9). Việc xếp lại thứ tự cần thiết các số liệu để áp dụng kiểm định cũng được trình bày trong Bảng 11.3.

Bỏ 4 quan sát ở giữa, các hồi quy OLS dựa vào 13 quan sát đầu và 13 quan sát cuối và các tổng bình phương phần dư tương ứng của chúng được trình bày tiếp sau (sai số chuẩn được đặt trong ngoặc).

Hồi quy dựa vào 13 quan sát đầu:

$$\begin{array}{rcll} Y_i = 3,4094 & + & 0,6968X_i & \\ (8,7049) & & (0,0744) & r^2 = 0,8887 \\ & & & RSS_1 = 377,17 \\ & & & \text{Bậc tự do} = 11 \end{array}$$

Hồi quy dựa vào 13 quan sát sau:

¹⁷ Về mặt kỹ thuật, **năng lực** của kiểm định phụ thuộc vào việc c được lựa chọn như thế nào. Trong thống kê, *năng lực của một kiểm định* được tính bởi xác suất bác bỏ giả thiết không khi nó sai [nghĩa là bằng $1 - \text{Xác suất (sai lầm loại II)}$]. Ở đây, *giả thiết không* là các phương sai của hai nhóm như nhau, nghĩa là phương sai không thay đổi. Về thảo luận sâu hơn, xem M. M. và C. Giaccotto, “Một Nghiên cứu các Kiểm định Mới và Hiện hữu về Phương sai thay đổi trong Mô hình Tuyến tính Tổng quát”, *Journal of Econometrics* (Tập chí Kinh tế Lượng), tập 26, 1984, trang 355-373.

¹⁸ George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, và Tsoung-Chao Lee, *Introduction to the Theory of Econometrics* (Giới thiệu Lý thuyết Kinh tế Lượng), John Wiley & Sons, New York, 1982, trang 422.

$$Y_i = -28,0272 + 0,7941X_i$$

$$(30,6421) \quad (0,1319) \quad r^2 = 0,7681$$

$$RSS_1 = 1536,8$$

$$\text{Bậc tự do} = 11$$

BẢNG 11.3

Số liệu giả thiết về chi tiêu tiêu dùng $Y(\text{USD})$ và thu nhập $X(\text{USD})$ để minh họa kiểm định Goldfeld-Quandt

Y	X	Số liệu xếp bậc theo các giá trị X	
		Y	X
55	80	55	80
65	100	70	85
70	85	75	90
80	110	75	100
79	120	65	105
84	115	74	110
98	130	80	115
95	140	84	120
90	125	79	125
75	90	90	130
74	105	98	140
110	160	95	145
113	150	108	150
125	165	113	160
108	145	110	165
115	180	125	180
140	225	115	185
120	200	130	190
145	240	135	200
130	185	120	205
152	220	140	210
144	210	144	220
175	245	152	225
180	260	140	230
135	190	137	240
140	205	145	245
178	265	175	250
191	270	189	260
137	230	178	265
189	250	191	270

} 4 quan sát
ở giữa

Từ các kết quả hồi quy này, ta có

$$\lambda = \frac{RSS_2/df}{RSS_1/df} = \frac{1536,8/11}{377,17/11}$$

$$\lambda = 4,07$$

Giá trị tới hạn của F đối với 11 bậc tự do ở tử số và 11 bậc tự do ở mẫu số tại mức 5% là 2,82. Do giá trị $F(= \lambda)$ lớn hơn giá trị tới hạn, ta có thể kết luận rằng có phương sai thay đổi trong phương sai của sai số. Tuy nhiên, nếu mức ý nghĩa được cố định tại 1%, ta có thể không bác bỏ giả thiết về phương sai không thay đổi. (Tại sao?) Lưu ý rằng giá trị p của λ quan sát được là 0,014.

Kiểm định Preusch-Pagan-Godfrey.¹⁹ Sự thành công của kiểm định Goldfeld-Quandt không chỉ phụ thuộc vào giá trị của c (số các quan sát ở giữa bị loại bỏ) mà còn vào việc xác định đúng biến X để dùng nó xếp thứ tự các quan sát. Hạn chế này của kiểm định có thể vượt qua được nếu ta xem xét **kiểm định Preusch-Pagan-Godfrey (BPG)**.

Để minh họa kiểm định này, hãy xem xét mô hình hồi quy tuyến tính k biến sau:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \quad (11.5.11)$$

Giả thiết rằng phương sai của sai số σ_i^2 được biểu diễn như sau:

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi}) \quad (11.5.12)$$

tức là, σ_i^2 là một hàm số của các biến phi ngẫu nhiên Z ; một số hay tất cả X có thể đóng vai trò của Z . Một cách cụ thể, giả thiết rằng

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi} \quad (11.5.13)$$

tức là, σ_i^2 là hàm tuyến tính của các biến Z . Nếu $\alpha_2 = \alpha_3 = \dots = \alpha_m$, thì $\sigma_i^2 = \alpha_1 =$ hằng số. Do vậy, để kiểm định xem σ_i^2 có không thay đổi hay không, ta có thể kiểm định giả thiết là $\alpha_2 = \alpha_3 = \dots = \alpha_m = 0$. Đây là ý tưởng cơ bản đằng sau kiểm định Breusch-Pagan. Sau đây là quy tắc thực tế:

Bước 1. Ước lượng (11.5.11) bằng OLS và tính các phần dư $u_1, u_2, \dots, u_n$.

Bước 2. Tính $\tilde{\sigma}^2 = \sum \hat{u}_i^2 / n$. Nhớ lại từ Chương 4 rằng đây là ước lượng hợp lý tối đa (ML) của σ^2 . [Lưu ý: Ước lượng OLS là $\sum \hat{u}_i^2 / (n - k)$].

Bước 3. Lập các biến p_i với định nghĩa sau:

$$p_i = \hat{u}_i^2 / \tilde{\sigma}^2 \quad (11.5.14)$$

Biến p_i đơn giản là từng phần dư bình phương chia cho $\tilde{\sigma}^2$.

Bước 4. Thực hiện hồi quy đối với p_i vừa thiết lập theo các biến Z như sau:

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi} + v_i \quad (11.5.15)$$

với v_i là phần dư trong hồi quy này.

Bước 5. Tính ESS (tổng các bình phương giải thích được) từ (11.5.14) và định nghĩa

¹⁹ T. Breusch & A. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation" (Một kiểm định đơn giản về phương sai thay đổi và biến thiên hệ số ngẫu nhiên), *Econometrica*, tập 47, 1979, trang 1287-1294. Xem đồng thời L. Godfrey, "Testing for Multiplicative Heteroscedasticity", *Journal of Econometrics* (Kiểm định phương sai thay đổi tích, Tạp chí Kinh tế Lượng), tập 8, 1978, trang 227-236. Do giống nhau nên các kiểm định này được gọi chung là kiểm định Breusch-Pagan-Godfrey về phương sai thay đổi.

$$T = \frac{1}{2} (ESS) \quad (11.5.15)$$

Giả sử u_i có phân phối chuẩn, ta có thể chỉ ra rằng nếu có phương sai không thay đổi và nếu cỡ mẫu n tăng lên vô hạn, thì

$$T \underset{asy}{\sim} \chi_{m-1}^2 \quad (11.5.16)$$

tức là, T tuân theo phân phối Chi-bình phương với $(m - 1)$ bậc tự do. (*Lưu ý: asy có nghĩa là tiệm cận*).

Do vậy, nếu trong một ứng dụng giá trị T tính được ($= \chi^2$) vượt giá trị tới hạn của χ^2 tại mức ý nghĩa đã chọn, ta có thể bác bỏ giả thiết về phương sai không thay đổi; nếu không ta không bác bỏ nó.

Ví dụ 11.4. Kiểm định Breusch-Pagan-Godfrey. Ta hãy quay lại với số liệu (Bảng 11.3) được dùng để minh họa kiểm định Goldfeld-Quandt về phương sai thay đổi. Thực hiện hồi quy Y theo X , ta có các kết quả sau:

Bước 1.

$$\begin{aligned} Y_i &= 9,2903 + 0,6378X_i \\ se &= (5,2314) \quad (0,0286) \quad RSS = 2361,153 \\ R^2 &= 0,9466 \end{aligned} \quad (11.5.17)$$

Bước 2.

$$\tilde{\sigma}^2 = \sum \hat{u}_i^2 / 30 = 2361,153 / 30 = 78,7051$$

Bước 3. Chia các phần dư u_i tính được từ hồi quy (11.5.7) cho 78,7051 để thiết lập biến p_i .

Bước 4. Giả thiết rằng p_i có quan hệ tuyến tính với $X_i = (Z_i)$ như theo (11.5.3), ta có hồi quy:

$$\begin{aligned} p_i &= -0,7426 + 0,0101X_i \\ se &= (0,7529) \quad (0,0041) \quad ESS = 10,4280 \\ R^2 &= 0,18 \end{aligned} \quad (11.5.18)$$

Bước 5.

$$T = \frac{1}{2} (ESS) = 5,2140 \quad (11.5.19)$$

Với giả thiết của kiểm định BPG, T trong (11.5.19) một cách tiệm cận, tuân theo phân phối Chi-bình phương với số bậc tự do bằng 1. [*Lưu ý: Chỉ có một biến hồi quy độc lập trong (11.5.8)*]. Bây giờ, từ bảng Chi-bình phương ta thấy rằng với 1 bậc tự do, giá trị Chi-bình phương tới hạn 5% là 3,8414 và giá trị F tới hạn 1% là 6,6349. Như vậy, giá trị Chi-bình phương quan sát được là 5,2140 có ý nghĩa ở mức 5% nhưng không có ý nghĩa ở mức 1%. Do vậy, ta có cùng kết luận như kiểm định Goldfeld-Quandt. Nhưng lưu ý rằng, nói một cách chặt chẽ, kiểm định BPG là một kiểm định tiệm cận, hay mẫu lớn và, trong ví dụ của ta, 30 quan sát có thể không tạo thành một mẫu lớn. Cũng cần phải chỉ ra rằng trong các mẫu nhỏ, kiểm định nhạy cảm với giả thiết cho rằng các yếu tố nhiễu u_i có

phân phối chuẩn. Tất nhiên, ta có thể kiểm định giả thiết về quy luật chuẩn bằng các kiểm định Chi-bình phương hay **Bera-Jarque** thảo luận trước đây.²⁰

Kiểm định tổng quát về phương sai thay đổi của White. Không như kiểm định Goldfeld-Quandt trong đó yêu cầu việc xếp lại thứ tự các quan sát theo biến X , vốn được cho rằng đã gây ra phương sai thay đổi, hay kiểm định BGP nhạy cảm với giả thiết về quy luật chuẩn, kiểm định tổng quát về phương sai không thay đổi do White đề xuất không lệ thuộc vào giả thiết về quy luật chuẩn và được dễ dàng thực hiện.²¹ Để minh họa ý tưởng cơ bản, hãy xem xét mô hình hồi quy ba biến sau (tổng quát hóa thành mô hình k biến có thể được dễ dàng thực hiện):

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (11.5.20)$$

Kiểm định White được tiến hành như sau:

Bước 1. Với số liệu cho trước, ta ước lượng (11.5.20) và tính các phần dư u_i .

Bước 2. Sau đó, ta thực hiện hồi quy (bổ trợ) sau:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i \quad (11.5.21)^{22}$$

Tức là, các phần dư bình phương từ hồi quy gốc được hồi quy theo các biến X ban đầu hay biến hồi quy độc lập, giá trị bình phương của chúng, và (các) tích chéo giữa các biến hồi quy độc lập. Ta cũng có thể đưa ra các lũy thừa cao hơn đối với các biến hồi quy độc lập. Lưu ý rằng có một số hạng không đổi trong phương trình này mặc dù hồi quy ban đầu có thể chứa hay không chứa nó. Tính R^2 từ hồi quy (bổ trợ).

Bước 3. Với *giả thiết không* cho rằng không có phương sai thay đổi, ta có thể chỉ ra rằng cỡ mẫu (n) nhân với R^2 tính được từ hồi quy bổ trợ, một cách tiệm cận, tuân theo phân phối Chi-bình phương với số bậc tự do (df) bằng với số lượng các biến hồi quy độc lập (loại trừ số hạng không đổi) trong hồi quy bổ trợ. Tức là,

$$n \cdot R^2 \xrightarrow{\text{asy}} \chi_{df}^2 \quad (11.5.22)$$

với df (số bậc tự do) được xác định ở trên. Trong ví dụ của ta, số bậc tự do là 5 do có 5 biến hồi quy độc lập trong hồi quy bổ trợ.

Bước 4. Nếu giá trị Chi-bình phương tính được trong (11.5.22) vượt giá trị Chi-bình phương tới hạn tại mức ý nghĩa lựa chọn, ta kết luận rằng có phương sai thay đổi. Nếu nó không vượt giá trị Chi-bình phương tới hạn thì không có phương sai thay đổi, tức là ta nói rằng trong hồi quy bổ trợ (11.5.21), $\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$ (xem chú thích 22).

²⁰ Về khía cạnh này, xem R. Koenker, "A Note on Studentizing a Test for Heteroscedasticity", *Journal of Econometrics* (Một lưu ý về thực hiện kiểm định t về phương sai thay đổi, *Tạp chí Kinh tế Lượng*), tập 17, 1981, trang 1180-1200.

²¹ H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity" (Ước lượng ma trận hiệp phương sai nhất quán phương sai thay đổi và kiểm định trực tiếp phương sai thay đổi), *Econometrica*, tập 48, 1980, trang 817-818.

²² Đằng sau quy trình này là giả thiết cho rằng phương sai của sai số u_i , σ_i^2 , có quan hệ hàm số với các biến hồi quy độc lập, bình phương của chúng, và tích chéo giữa chúng. Nếu tất cả các hệ số độ dốc riêng phần trong hồi quy đều đồng thời bằng 0, thì phương sai của sai số là hằng số không thay đổi bằng với α_1 .

Ví dụ 11.5. Kiểm định về phương sai thay đổi của White. Dựa vào số liệu chéo của 41 quốc gia, Stephen Lewis ước lượng mô hình hồi quy sau:²³

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \quad (11.5.23)$$

với Y = tỷ lệ giữa thuế ngoại thương (thuế xuất nhập khẩu) và thu nhập của chính phủ, X_2 = tỷ lệ giữa tổng giá trị xuất khẩu nhập khẩu và GNP, và X_3 = GNP bình quân đầu người; $\ln$ là logarit tự nhiên. Giả thiết của Lewis là Y và X_2 sẽ có quan hệ đồng biến (kim ngạch ngoại thương càng cao thì thu nhập từ thuế ngoại thương càng cao) và Y và X_3 sẽ có quan hệ nghịch biến (khi thu nhập tăng lên, chính phủ sẽ dễ thu thuế trực thu – ví dụ như thuế lợi tức – hơn so với thuế ngoại thương).

Các kết quả thực nghiệm hỗ trợ giả thiết này. Đối với mục đích của chúng ta, điểm quan trọng là trong số liệu có phương sai thay đổi hay không. Do đây là số liệu chéo gồm nhiều quốc gia khác nhau, ta dự đoán là có phương sai thay đổi trong phương sai của sai số. Bằng cách áp dụng **kiểm định về phương sai thay đổi của White** đối với các phần dư tính được từ hồi quy (11.5.23), ta tính được các kết quả sau:²⁴

$$\begin{aligned} \hat{u}_i^2 = & -5,8417 + 2,5629 \ln \text{Trade}_i + 0,6918 \ln \text{GNP}_i \\ & - 0,408 (\ln \text{Trade}_i)^2 - 0,049 (\ln \text{GNP}_i)^2 \\ & + 0,0015 (\ln \text{Trade}_i) (\ln \text{GNP}_i) \end{aligned} \quad R^2 = 0,1148 \quad (11.5.24)$$

Lưu ý: Các sai số chuẩn không được trình bày do chúng không thích hợp cho mục đích của chúng ta ở đây.

Bây giờ, $n \cdot R^2 = 41(0,1148) = 4,7068$ một cách tiệm cận, có phân phối Chi-bình phương với 5 bậc tự do (tại sao?). Giá trị Chi-bình phương tới hạn ở mức 5% với 5 bậc tự do là 11,0705, giá trị tới hạn ở mức 10% là 9,2363, và giá trị tới hạn ở mức 25% là 6,62568. Với tất cả các mục đích thực tiễn, ta có thể kết luận rằng, trên cơ sở của kiểm định White, không có phương sai thay đổi.

Sau đây là một bình luận cần thiết về kiểm định White. Nếu một mô hình có vài biến hồi quy độc lập, thì việc đưa ra tất cả các biến hồi quy độc lập, bình phương của chúng (hay lũy thừa bậc cao hơn), và tích chéo có thể nhanh chóng sử dụng hết số bậc tự do. Do vậy, ta phải lưu ý khi sử dụng kiểm định này. Đôi khi ta có thể bỏ các số hạng tích chéo. Trong các trường hợp mà thống kê kiểm định có ý nghĩa, phương sai thay đổi có thể không nhất thiết là nguyên nhân, mà lại là các sai số đặc trưng. Vấn đề này sẽ được đề cập trong Chương 13 (nhớ lại điểm #5 trong Mục 11.1). Nói một cách khác, **kiểm định White có thể là một kiểm định về phương sai thay đổi (thuần túy) hay sai số đặc trưng hay cả hai.**

Các kiểm định khác về phương sai thay đổi. Có một vài kiểm định khác về phương sai thay đổi, mỗi loại dựa trên các giả thiết nhất định. Người đọc có quan tâm có thể xem tài liệu tham khảo.²⁵

²³ Stephen R. Lewis, “Government Revenue from Foreign Trade”, *Manchester School of Economics and Social Studies* (Thu nhập của chính phủ từ ngoại thương, *Trường Kinh tế và Xã hội học Manchester*), tập 31, 1963, trang 39-47.

²⁴ Các kết quả này, với thay đổi trong ký hiệu, được lấy từ William F. Lott & Subhash C. Ray, *Applied Econometrics: Problems with Data Sets*, Instructor’s Manual (Kinh tế lượng ứng dụng: các vấn đề với những tập hợp số liệu, Sổ tay Người hướng dẫn), Chương 22, trang 137-140.

²⁵ Xem M. J. Harrison & B. P. MacCabe, “A Test for Heteroscedasticity Based on Ordinary Least Square Residuals”, *Journal of the American Statistical Association* (Một kiểm định về phương sai thay đổi dựa trên các phần dư tính theo phương pháp bình phương tối thiểu thông thường, Tạp chí của Hiệp hội Thống kê Hoa Kỳ, tập 74, 1979, trang 494-499; J. Szroeter, “A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models” (Một lớp các kiểm định thông số về phương sai thay đổi trong các mô hình kinh tế lượng tuyến tính),

11.6. CÁC BIỆN PHÁP KHẮC PHỤC

Như ta đã thấy, phương sai thay đổi không làm mất đi tính các tính chất không thiên lệch và nhất quán của các ước lượng OLS, nhưng chúng không còn hiệu quả nữa, thậm chí khi tiệm cận (nghĩa là quy mô mẫu lớn). Việc mất tính hiệu quả làm cho quy tắc kiểm định giả thiết có giá trị mơ hồ. Do vậy, cần phải có các biện pháp khắc phục. Có hai cách thức xử lý: khi biết σ_i^2 và khi chưa biết σ_i^2 .

Khi biết σ_i^2 : Phương pháp bình phương tối thiểu có trọng số

Như ta đã thấy trong Mục 11.3, nếu biết được σ_i^2 , phương pháp đơn giản nhất để hiệu chỉnh phương sai thay đổi là bằng bình phương tối thiểu có trọng số vì vậy các ước lượng theo phương pháp này là các ước lượng tuyến tính không thiên lệch tốt nhất (BLUE).

Ví dụ 11.6. Minh họa phương pháp bình phương tối thiểu có trọng số. Để minh họa phương pháp này, giả sử ta muốn nghiên cứu quan hệ giữa lương và quy mô lao động qua số liệu trong Bảng 11.1. Để đơn giản hóa, ta tính quy mô lao động bằng 1 (1-4 lao động), 2 (5-9 lao động),..., 9 (1000-2499 lao động), mặc dù ta cũng có thể tính nó bằng trung điểm của các nhóm lao động khác nhau trong bảng (xem bài tập 11.21).

BẢNG 11.4

Minh họa hồi quy bình phương tối thiểu có trọng số

Lương Y	Quy mô LĐ X	σ_i	Y_i/σ_i	X_i/σ_i
3396	1	743,4	4,5664	0,0013
3787	2	851,4	4,4480	0.0023
4013	3	727,8	5,5139	0.0041
4104	4	805,06	5,0978	0.0050
4146	5	929,9	4,4585	0.0054
4241	6	1080,6	3,9247	0.0055
4387	7	1243,2	3,5288	0.0056
4538	8	1307,7	3,4702	0.0061
4843	9	1112,5	4,3532	0.0081

Lưu ý: Trong hồi quy (11.6.2), biến phụ thuộc là (Y_i/σ_i) và các biến độc lập là $(1/\sigma_i)$ và (X_i/σ_i) .

Nguồn: Số liệu về Y và σ_i (độ lệch chuẩn của lương) được lấy từ Bảng 11.1. Quy mô lao động: 1 = 1-4 lao động, 2 = 5-9 lao động, v.v... Số liệu này cũng được lấy từ Bảng 11.1.

Bây giờ, gọi Y là lương trung bình một lao động (USD) và X là quy mô lao động, ta chạy hồi quy sau [Xem Phương trình (11.3.6)]:

$$Y_i / \sigma_i = \hat{\beta}_1^* (1 / \sigma_i) + \hat{\beta}_2^* (X_i / \sigma_i) + (\hat{u}_i / \sigma_i) \quad (11.6.1)$$

Econometrica, tập 46, 1978, trang 1311-1327; M. A. Evans & M. L. King, “A Further Class of Tests for Heteroscedasticity”, *Journal of Econometrics* (Một lớp các kiểm định nữa về phương sai thay đổi, Tạp chí Kinh tế Lượng), tập 37, 1988, trang 265-276.

với σ_i là các độ lệch chuẩn của mức lương như báo cáo trong Bảng 11.1. Số liệu thô cần thiết để chạy hồi quy này được trình bày trong Bảng 11.4.

Trước khi chuyển sang phân tích các kết quả hồi quy, lưu ý rằng (11.6.1) không có tung độ gốc. (Tại sao?) Do vậy, ta sẽ phải sử dụng mô hình hồi quy qua gốc tọa độ để ước lượng β_1^* và β_2^* , một chủ đề đã thảo luận trong Chương 6. Nhưng phần lớn các phần mềm thống kê ngày nay có tính năng khử tung độ gốc (ví dụ, xem SAS). Cũng lưu ý một đặc điểm thú vị nữa của (11.6.1): Nó có hai biến giải thích, $(1/\sigma_i)$ và (X_i/σ_i) , trái lại nếu ta sử dụng OLS thực hiện hồi quy mức lương theo quy mô lao động, hồi quy sẽ có một biến giải thích, X_i . (Tại sao?)

Sau đây là các kết quả hồi quy của bình phương tối thiểu có trọng số (WLS):

$$\begin{aligned} \widehat{(Y_i/\sigma_i)} &= 3406,639(1/\sigma_i) + 154,153(X_i/\sigma_i) \\ &\quad (80,983) \quad (16,959) \\ t &= (42,066) \quad (9,090) \quad R^2 = 0,9993^{26} \end{aligned} \quad (11.6.2)$$

Để so sánh, ta đưa ra các kết quả hồi quy OLS thông thường hay không tính trọng số:

$$\begin{aligned} Y_i &= 3417,833 + 148,767X_i \\ &\quad (81,136) \quad (14,418) \\ t &= (42,125) \quad (10,318) \quad R^2 = 0,9383 \end{aligned} \quad (11.6.3)$$

Trong bài tập 11.7 bạn được yêu cầu so sánh hai hồi quy này.

Khi chưa biết σ_i^2

Như đã lưu ý ở trên, nếu biết giá trị đúng của σ_i^2 , ta có thể sử dụng phương pháp WLS để tính các ước lượng BLUE. Vì giá trị đúng của σ_i^2 ít khi biết được nên câu hỏi đặt ra là có cách nào tính được các ước lượng *nhất quán* (về mặt thống kê) của các phương sai và tích sai của các ước lượng OLS thậm chí nếu có phương sai thay đổi hay không? Câu trả lời là có.

Các sai số chuẩn và phương sai nhất quán - thay đổi của White. White đã chỉ ra rằng ước lượng này có thể được thực hiện sao cho có thể đưa ra suy luận thống kê có giá trị một cách tiệm cận (nghĩa là cỡ mẫu lớn) về các giá trị đúng của tham số.²⁷ Ta sẽ không trình bày các chi tiết toán học do chúng nằm ngoài phạm vi của cuốn sách này. Nhưng một số phần mềm máy tính (ví dụ, TSP, ET, SHAZAM) giờ đây trình bày các phương sai và sai số chuẩn đã hiệu chỉnh phương sai thay đổi cùng với các phương sai và sai số chuẩn OLS thông thường.²⁸

Ví dụ 11.7. Minh họa quy tắc của White. Ví dụ, ta tóm tắt các kết quả sau do Greene thực hiện:²⁹

$$\begin{aligned} Y_i &= 832,91 - 1834,2(\text{Thu nhập}) + 1587,04(\text{Thu nhập})^2 \\ \text{OLS se} &= (327,3) \quad (829,0) \quad (519,1) \\ t &= (2,54) \quad (2,21) \quad (3,06) \end{aligned} \quad (11.6.4)$$

²⁶ Như lưu ý trong chú thích 3 của Chương 6, R^2 của hồi quy qua gốc tọa độ không trực tiếp so sánh với R^2 của mô hình có tung độ gốc. R^2 tính được là 0,9993 có tính đến sự khác biệt này. (Xem phần mềm SAS về chi tiết R^2 được hiệu chỉnh như thế nào cho việc thiếu tung độ gốc. Xem cả Phụ lục 6A, Mục 6A1).

²⁷ Xem H. White, op. Cit.

²⁸ Nói một cách kỹ thuật hơn, chúng được gọi là **các ước lượng ma trận tích sai nhất quán-phương sai thay đổi** (heteroscedasticity-consistent covariance matrix estimators), viết tắt là **HCCME**.

²⁹ William H. Greene, *Econometric Analysis* (Phân tích Kinh tế Lượng), Xuất bản lần thứ 2, Macmillan, New York, 1993, trang 385.

$$\begin{array}{lll} \text{White se} = & (460,9) & (1243,0) & (830,0) \\ t = & (1,81) & (-1,48) & (1,91) \end{array}$$

với Y = chi tiêu bình quân đầu người cho trường công lập theo tiểu bang năm 1979 và Thu nhập = thu nhập bình quân đầu người theo tiểu bang năm 1979. Mẫu gồm 50 bang cộng với Washington, D.C.

Như các kết quả trên biểu thị, các sai số chuẩn đã hiệu chỉnh phương sai thay đổi của White lớn hơn đáng kể so với các sai số chuẩn OLS và do vậy các giá trị t ước lượng nhỏ hơn nhiều so với các giá trị tính được theo phương pháp OLS. Trên cơ sở của các ước lượng OLS, cả hai biến hồi quy độc lập đều có ý nghĩa thống kê ở mức 5%, trái lại trên cơ sở của các ước lượng White thì chúng không có ý nghĩa. Tuy nhiên, cần phải chỉ ra rằng các sai số chuẩn đã hiệu chỉnh phương sai thay đổi của White có thể lớn hơn hay nhỏ hơn các sai số chuẩn không hiệu chỉnh.

Do giờ đây các ước lượng phương sai nhất quán - thay đổi của White có trong các phần mềm hồi quy chính thức, người đọc nên báo cáo chúng trong kết quả hồi quy của mình. Như Wallace và Silver lưu ý:

Nói chung, nên sử dụng phương pháp White (có trong các chương trình hồi quy) thường xuyên, có lẽ so sánh kết quả với kết quả OLS thông thường để kiểm tra xem phương sai thay đổi có gây ra vấn đề nghiêm trọng nào không trong một tập hợp số liệu cụ thể.³⁰

Các giả thiết hợp lý về dạng phương sai thay đổi. Ngoài yêu cầu về mẫu lớn, một thiếu sót của quy tắc White là các ước lượng tính được có thể không hiệu quả như là các ước lượng tính được bằng các phương pháp trong đó biến đổi số liệu để phản ánh các thể loại cụ thể của phương sai thay đổi. Để minh họa điều này, hãy quay lại với mô hình hồi quy hai biến:

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

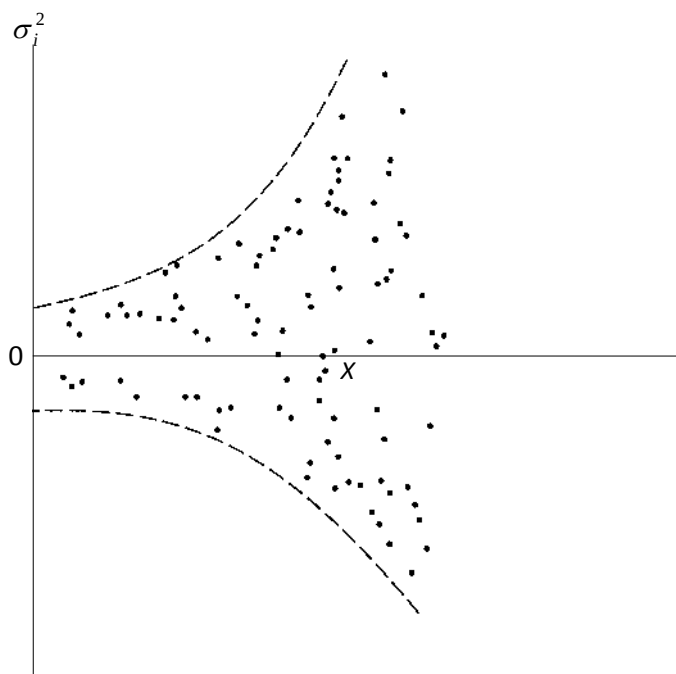
Bây giờ, ta xem xét một số giả thiết về phương sai thay đổi.

Giả thiết 1: Phương sai của sai số tỷ lệ thuận với X_i^2 :

$$E(u_i^2) = \sigma^2 X_i^2 \quad (11.6.5)^{31}$$

³⁰ T. Dudley Wallace & J. Lew Silver, *Econometrics: An Introduction*, (Kinh tế Lượng: Giới thiệu), Reading, Mass. 1988, trang 265.

³¹ Nhớ lại rằng ta đã gặp phải giả thiết này trong phần thảo luận về kiểm định Goldfeld-Quandt.

**HÌNH 11.9** Phương sai của sai số tỷ lệ thuận với X^2 .

Nếu, do tính chất “suy đoán”, các phương pháp đồ thị, hay phương pháp Park & Glejser, người ta tin rằng phương sai của u_i tỷ lệ thuận với bình phương của biến giải thích X (xem Hình 11.9), ta có thể biến đổi mô hình gốc như sau. Chia mô hình gốc cho X_i :

$$\begin{aligned}\frac{Y_i}{X_i} &= \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} \\ &= \beta_1 \frac{1}{X_i} + \beta_2 + v_i\end{aligned}\quad (11.6.6)$$

với v_i là số hạng nhiễu đã biến đổi, bằng u_i/X_i . Bây giờ, ta có thể dễ dàng chứng minh rằng:

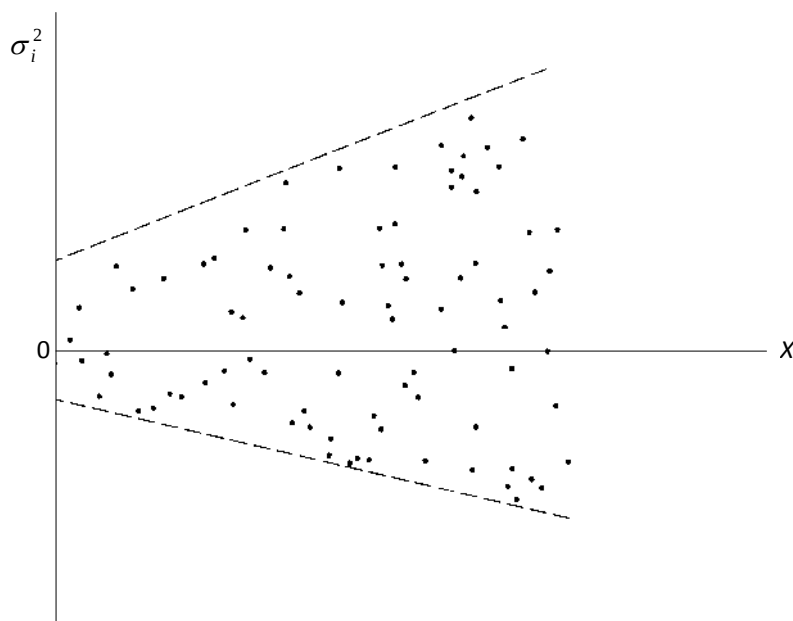
$$\begin{aligned}E(v_i^2) &= E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E(u_i^2) \\ &= \sigma^2 \quad \text{sử dụng (11.6.5)}\end{aligned}$$

Vậy, phương sai của v_i giờ đây không thay đổi, và ta có thể tiếp tục áp dụng OLS cho phương trình đã biến đổi (11.6.6), thực hiện hồi quy Y_i/X_i theo $1/X_i$.

Lưu ý rằng trong hồi quy đã biến đổi, tung độ gốc β_2 là hệ số góc của mô hình ban đầu và hệ số độ dốc. Do vậy, để trở lại mô hình ban đầu ta sẽ phải nhân (11.6.6) ước lượng được với X_i . Một áp dụng của phép biến đổi này được trình bày trong bài tập 11.17.

Giả thiết 2: Phương sai của sai số tỷ lệ thuận với X_i . **Biến đổi căn bậc hai:**

$$E(u_i^2) = \sigma^2 X_i \quad (11.6.7)$$

**HÌNH 11.10** Phương sai của sai số tỷ lệ thuận với X .

Nếu ta tin rằng phương sai của u_i , thay vì tỷ lệ với bình phương của X_i , tỷ lệ với X_i , thì mô hình ban đầu có thể được biến đổi như sau (xem Hình 11.10):

$$\begin{aligned}\frac{Y_i}{\sqrt{X_i}} &= \frac{\beta_1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ &= \beta_1 \frac{1}{\sqrt{X_i}} + \beta_2 \sqrt{X_i} + v_i\end{aligned}\quad (11.6.8)$$

với $v_i = u_i / \sqrt{X_i}$ và với $X_i > 0$.

Với giả thiết 2, ta có thể dễ dàng chứng minh rằng $E(v_i^2) = \sigma^2$, một trường hợp phương sai không thay đổi. Do vậy, ta sẽ phải sử dụng mô hình hồi quy qua gốc tọa độ để ước lượng β_1 và β_2 . Sau khi đã thực hiện (11.6.8), ta có thể quay trở lại với mô hình ban đầu đơn giản bằng cách nhân (11.6.8) với $\sqrt{X_i}$.

Giả thiết 3: Phương sai của sai số tỷ lệ thuận với bình phương giá trị trung bình của Y .

$$E(u_i^2) = \sigma^2 [E(Y_i)]^2 \quad (11.6.9)$$

Phương trình (11.6.9) mặc định rằng phương sai của u_i tỷ lệ với bình phương giá trị kỳ vọng của Y (xem Hình 11.7.e). Bây giờ

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

Do vậy, nếu biến đổi hàm số ban đầu như sau:

$$\begin{aligned}\frac{Y_i}{E(Y_i)} &= \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)} \\ &= \beta_1 \left(\frac{1}{E(Y_i)} \right) + \beta_2 \frac{X_i}{E(Y_i)} + v_i\end{aligned}\quad (11.6.10)$$

với $v_i = u_i/E(Y_i)$, ta có thể thấy rằng $E(v_i^2) = \sigma^2$; tức là, các yếu tố nhiễu v_i có phương sai không thay đổi. Vậy, hồi quy (11.6.10) sẽ thỏa mãn giả thiết về phương sai không thay đổi của mô hình hồi quy tuyến tính cổ điển.

Tuy nhiên, phép biến đổi (11.6.10) không thực hiện được do $E(Y_i)$ phụ thuộc vào β_1 và β_2 đều chưa biết. Tất nhiên, ta biết $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ là một ước lượng của $E(Y_i)$. Do vậy, ta có thể tiến hành theo hai bước: Thứ nhất, ta chạy hồi quy OLS thông thường, không để ý tới phương sai thay đổi, và tính Y_i . Sau đó, sử dụng Y_i ước lượng được, ta biến đổi mô hình như sau:

$$\frac{Y_i}{\hat{Y}_i} = \beta_1 \left(\frac{1}{\hat{Y}_i} \right) + \left(\frac{X_i}{\hat{Y}_i} \right) + v_i \quad (11.6.11)$$

với $v_i = (u_i/Y_i)$. Trong Bước 2, ta chạy hồi quy (11.6.11). Mặc dù $\hat{Y}_i$ không chính xác là $E(Y_i)$, chúng là những ước lượng nhất quán; tức là, khi kích thước mẫu tăng lên vô hạn, chúng hội tụ tại giá trị đúng của $E(Y_i)$. Vậy, phép biến đổi (11.6.11) sẽ được thực hiện một cách thỏa đáng trên thực tiễn nếu cỡ mẫu đủ lớn.

Giả thiết 4: Biến đổi lôgarít như

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i \quad (11.6.12)$$

thường làm giảm phương sai thay đổi khi so sánh với hồi quy $Y_i = \beta_1 + \beta_2 X_i + u_i$.

Kết quả này nảy sinh là do biến đổi lôgarít làm thu lại tỷ lệ mà các biến được đo lường, do vậy làm giảm sự khác biệt 10 lần giữa hai giá trị xuống còn sự khác biệt 2 lần. Vậy, số 80 gấp 10 lần số 8, nhưng $\ln 80 (= 4,3280)$ gấp khoảng hai lần $\ln 8 (= 2,0794)$.

Một lợi thế nữa của phép biến đổi lôgarít là hệ số độ dốc β_2 đo độ co giãn của Y theo X , tức là, thay đổi theo tỷ lệ phần trăm của Y đối với thay đổi tỷ lệ phần trăm của X . Ví dụ, nếu Y là tiêu dùng và X là thu nhập, β_2 trong (11.6.12) sẽ cho ta độ co giãn thu nhập, trái lại trong mô hình gốc β_2 chỉ đo tốc độ thay đổi của tiêu dùng trung bình khi thu nhập thay đổi đi một đơn vị. Đó là một lý do tại sao các mô hình log khá phổ biến trong kinh tế lượng thực nghiệm. (Về một số vấn đề liên quan tới biến đổi log, xem bài tập 11.4).

Để kết thúc phần thảo luận về các biện pháp khắc phục, ta tái nhấn mạnh rằng tất cả các phép biến đổi thảo luận ở trên là để phục vụ cho mục tiêu cụ thể; ta đặc biệt quan tâm tới bản chất của σ_i^2 . Phép biến đổi nào thảo luận ở trên sẽ hữu hiệu phụ thuộc vào bản chất của vấn đề và tính chất nghiêm trọng của phương sai thay đổi. Ta cũng cần phải lưu ý thêm một số vấn đề liên quan tới các phép biến đổi:

1. Khi mở rộng mô hình hơn hai biến, ta có thể không biết trước các biến X nào cần phải lựa chọn để biến đổi số liệu.³²
2. Biến đổi log thảo luận trong Giả thiết 4 không áp dụng được nếu một số giá trị của Y và X bằng không hay âm.³³
3. Ta còn gặp phải vấn đề **tương quan giả** (spurious correlation). Thuật ngữ này, do Karl Pearson đặt ra, nói về trường hợp có tương quan giữa các tỷ số của các biến mặc dù các biến gốc không có tương quan hay ngẫu nhiên.³⁴ Vậy, trong mô hình $Y_i = \beta_1 + \beta_2 X_i + u_i$, Y và X có thể không tương quan nhưng trong mô hình biến đổi $Y_i/X_i = \beta_1(1/X_i) + \beta_2$, Y_i/X_i và $1/X_i$ thường được tìm thấy là có tương quan.
4. Khi σ_i^2 không được biết trực tiếp và được ước lượng từ một hay nhiều phép biến đổi mà ta đã thảo luận ở trên, tất cả các quy tắc kiểm định của ta sử dụng kiểm định t , F , v.v..., nói một cách chặt chẽ, chỉ có giá trị trong các mẫu lớn. Do vậy, ta phải cẩn thận trong việc giải thích các kết quả dựa vào các phép biến đổi khác nhau trong các mẫu nhỏ hay mẫu có giới hạn.³⁵

11.7 VÍ DỤ KẾT LUẬN

Để kết thúc thảo luận về phương sai thay đổi, chúng ta đưa ra một ví dụ minh họa về các phương pháp khác nhau trong việc phát hiện ra nó và một số biện pháp khắc phục.

Ví dụ 11.8: Chi tiêu cho nghiên cứu và phát triển (R&D) ở Hoa Kỳ, 1988. Số liệu về chi tiêu cho nghiên cứu và phát triển của 18 nhóm ngành kinh tế trong quan hệ với doanh thu và lợi nhuận được trình bày lại trong Bảng 11.5. Do số liệu chéo trình bày trong Bảng 11.5 khá không đồng nhất, hồi quy R&D theo doanh thu (hay lợi nhuận) có nhiều khả năng xuất hiện phương sai thay đổi. Sau đây là các kết quả hồi quy R&D theo doanh thu:

$$\begin{aligned}\widehat{\text{R\&D}} &= 192,99 + 0,0319 \text{ Doanh thu}_i \\ \text{se} &= (990,00) \quad (0,0083) \\ t &= (0,1948) \quad (3,8434) \quad r^2 = 0,4783\end{aligned}\tag{11.7.1}$$

Như kỳ vọng, chi tiêu R&D và doanh thu có tương quan đồng biến. Giá trị t tính được “dường như” có ý nghĩa thống kê ở mức 0,002 (kiểm định hai phía). Dĩ nhiên, nếu phương sai thay đổi, ta không thể tin tưởng vào các sai số chuẩn ước lượng hay các giá trị t ước lượng. Áp dụng kiểm định Park cho các phần dư ước lượng từ (11.7.1), ta có các kết quả sau:

$$\begin{aligned}\widehat{\ln \hat{u}_i^2} &= 5,6877 + 0,7014 \ln \text{Doanh thu}_i \\ \text{se} &= (6,6977) \quad (0,6033) \\ t &= (0,8572) \quad (1,1626) \quad r^2 = 0,0779\end{aligned}\tag{11.7.2}$$

³² Tuy nhiên, trên thực tiễn, ta có thể vẽ $\hat{u}_i^2$ theo từng biến và quyết định biến X nào có thể được sử dụng để biến đổi số liệu. (Xem Hình 11.8).

³³ Đôi khi ta có thể sử dụng $\ln(Y_i + k)$ hay $\ln(X_i + k)$, với k là một số dương lựa chọn sao cho tất cả các giá trị của Y và X trở nên dương. Xem bài tập 11.22.

³⁴ Ví dụ, nếu X_1 , X_2 , và X_3 không có tương quan lẫn nhau $r_{12} = r_{13} = r_{23} = 0$ và ta nhận thấy rằng (các giá trị của) các tỷ số X_1/X_2 và X_2/X_3 có tương quan, thì đây là trường hợp tương quan giả. “Tổng quát hơn, tương quan có thể được xem là giả nếu đó được tạo ra bởi phương pháp xử lý số liệu chứ không phải có trong số liệu ban đầu.” M. G. Kendall & W. R. Backland, *A Dictionary of Statistical Terms* (Tự điển Thuật ngữ Thống kê), Hafner Publishing, New York, 1972, trang 143.

³⁵ Về chi tiết, xem George G. Judge và các tác giả, op. Cit., Mục 14.4, trang 415-420.

Trên cơ sở của kiểm định Park, ta không có lý do nào để bác bỏ giả thiết về phương sai không thay đổi.

BẢNG 11.5

Đổi mới công nghệ tại Hoa Kỳ: Chi tiêu cho nghiên cứu và phát triển (R&D) tại Hoa Kỳ, 1988 (tính theo triệu USD)

Nhóm ngành kinh tế	Doanh thu	Chi R&D	Lợi nhuận
1. Bao bì & đóng gói	6.375,3	62,5	185,1
2. Tài chính phi ngân hàng	11.626,4	92,9	1569,5
3. Dịch vụ	14.655,1	178,3	276,8
4. Kim loại và khai thác mỏ	21.869,2	258,4	2.828,1
5. Nhà ở và xây dựng	26.408,3	494,7	225,9
6. Công nghiệp chế tạo chung	32.405,6	1.083,0	3.751,9
7. Các ngành dịch vụ giải trí	35.107,7	1.620,6	2.884,1
8. Giấy và sản phẩm từ rừng	40.295,4	421,7	4.645,7
9. Thực phẩm	70.761,6	509,2	5.036,4
10. Y tế	80.552,8	6620,1	13.869,9
11. Hàng không	95.294,0	3.918,6	4.487,8
12. Hàng tiêu dùng	101.314,1	1595,3	10.278,9
13. Điện và điện tử	116.141,3	6.107,5	8.787,3
14. Hóa chất	122.315,7	4.454,1	16.438,8
15. Tập đoàn	141.649,9	3.163,8	9.761,4
16. Thiết bị văn phòng và máy tính	175.025,8	13.210,7	19.774,5
17. Nhiên liệu	230.614,5	1.703,8	22.626,6
18. Thiết bị tự động	293.543,0	9.528,2	18.415,4

Lưu ý: Các ngành được xếp theo thứ tự doanh thu tăng dần.

Nguồn: *Business Week*, số đặc biệt 1989, Bảng tính điểm R&D, trang 180-224.

Trên cơ sở của kiểm định Glejser, ta có các kết quả sau (để đỡ tốn diện tích, ở đây chỉ trình bày các giá trị của t):

$$\begin{aligned} |\hat{u}_i| &= 578,57 + 0,0119 \ln \text{Doanhthu}_i \\ t &= (0,8525) \quad (2,0931) \quad r^2 = 0,2150 \end{aligned} \quad (11.7.3)$$

$$\begin{aligned} |\hat{u}_i| &= -507,02 + 7,2970 \sqrt{\text{Doanhthu}_i} \\ t &= (-0,5032) \quad (2,3704) \quad r^2 = 0,2599 \end{aligned} \quad (11.7.4)$$

$$\begin{aligned} |\hat{u}_i| &= 2.273,7 + 19.925.000(1/\text{Doanhthu}_i) \\ t &= (3,7601) \quad (-1,6175) \quad r^2 = 0,1405 \end{aligned} \quad (11.7.5)$$

Như phương trình (11.7.3) và (11.7.4) cho thấy, giả thiết về các phương sai không thay đổi có thể bị bác bỏ. Do vậy, các sai số chuẩn ước lượng và giá trị t không thể được chấp nhận tại giá trị bên ngoài của chúng. Trong bài tập 11.23 người đọc được yêu cầu áp dụng các kiểm định Breusch-Pagan và White về phương sai thay đổi cho số liệu trong Bảng 11.5.

Do dường như có nghi ngờ về giả thiết phương sai không thay đổi, hãy xem liệu ta có thể biến đổi số liệu để giảm tính nghiêm trọng của phương sai thay đổi hay không, nếu không phải là hoàn toàn loại bỏ nó. Về các phần dư tính được từ hồi quy (11.7.1), ta có thể thấy rằng phương sai của sai số tỷ lệ với biến doanh thu (kiểm tra điều này), và do vậy, theo Giả thiết 2 thảo luận ở trên, ta có thể sử dụng biến đổi căn bình phương. Phép biến đổi này cho ta các kết quả sau:

$$\frac{R \& D}{\sqrt{\text{Doanhthu}_i}} = -246,68 \frac{1}{\sqrt{\text{Doanhthu}_i}} + 0,00368 \sqrt{\text{Doanhthu}_i}$$

$$\begin{array}{lll} se = (341,13) & (0,0071) & \\ t = (-0,6472) & (5,1723) & R^2 = 0,6258 \end{array} \quad (11.7.6)$$

Nếu bạn nhân (11.7.6) với $\sqrt{\text{Doanhthu}_i}$ ở cả hai vế, bạn sẽ có các kết quả có thể so sánh với hồi quy gốc (11.7.1). Có rất ít khác biệt giữa hai hệ số độ dốc. Nhưng lưu ý rằng so với (11.7.1) sai số chuẩn của hệ số độ dốc trong (11.7.6) nhỏ hơn, từ đó cho thấy hồi quy OLS gốc đã ước lượng sai số chuẩn quá cao trên thực tế. Như lưu ý ở trên, khi xuất hiện phương sai thay đổi, các ước lượng OLS của sai số chuẩn bị thiên lệch và ta không thể nói trước được là thiên lệch theo hướng nào. Trong ví dụ hiện tại, thiên lệch có hướng đi lên, tức là, nó *ước lượng quá cao* sai số chuẩn. Đồng thời lưu ý rằng (11.7.6) đại diện cho *bình phương tối thiểu có trọng số* (tại sao?).

Trong bài tập 11.25, người đọc được yêu cầu tính các sai số chuẩn đã hiệu chỉnh phương sai thay đổi của White cho ví dụ ở trên và so sánh các kết quả với kết quả trong (11.7.6).

11.8 TÓM TẮT VÀ KẾT LUẬN

1. Một giả thiết quan trọng của mô hình hồi quy tuyến tính cổ điển là các yếu tố nhiễu u_i có cùng phương sai, σ^2 . Nếu giả thiết này không được thỏa mãn thì có sự hiện diện của phương sai thay đổi.
2. Phương sai thay đổi không làm mất đi tính chất không thiên lệch và nhất quán của các ước lượng OLS.
3. Nhưng các ước lượng này không còn có phương sai nhỏ nhất hay là các ước lượng hiệu quả. Tức là chúng không còn là các ước lượng tuyến tính không thiên lệch tốt nhất (BLUE).
4. Các ước lượng BLUE được tính từ phương pháp bình phương tối thiểu có trọng số, với điều kiện là các phương sai thay đổi, σ_i^2 , đã biết.
5. Khi có phương sai thay đổi, các phương sai của các ước lượng OLS không được tính từ các công thức OLS thông thường. Nhưng nếu ta vẫn sử dụng các công thức OLS thông thường, các kiểm định t và F dựa vào chúng có thể gây ra những kết luận sai lầm.
6. Trình bày các hậu quả của phương sai thay đổi dễ hơn là phát hiện ra nó. Có một số các kiểm định chuẩn đoán, nhưng ta không thể đảm bảo chắc rằng kiểm định nào hữu hiệu trong một trường hợp cụ thể.
7. Thậm chí nếu như ta nghi ngờ và phát hiện ra phương sai thay đổi, việc hiệu chỉnh vấn đề không phải là việc dễ dàng. Nếu mẫu lớn, ta có thể tính các sai số chuẩn của các ước lượng OLS đã hiệu chỉnh phương sai thay đổi của White và thực hiện suy luận thống kê dựa trên các sai số chuẩn này.
8. Trái lại, dựa vào các phần dư OLS, ta có thể thực hiện dự đoán qua rèn luyện về dạng có thể có của phương sai thay đổi và biến đổi số liệu gốc theo cách mà trong số liệu đã biến đổi không còn phương sai thay đổi.
9. Cuối cùng, các nhiễu phần dư OLS có thể không chỉ thay đổi mà còn tự tương quan. Một kỹ thuật gọi là **phương sai thay đổi có điều kiện tự tương quan**, viết tắt là **ARCH**, có thể được sử dụng để giải quyết vấn đề này. Ta sẽ thảo luận điều này trong Chương 12 khi xem xét chi tiết nội dung tự tương quan.

BÀI TẬP**Câu hỏi**

11.1. Cho biết với lý do ngắn gọn các phát biểu sau là đúng, sai hay không chắc chắn:

- (a) Trong trường hợp có phương sai thay đổi các ước lượng OLS bị thiên lệch và không hiệu quả.
- (b) Nếu có phương sai thay đổi, các kiểm định t và F thông thường không có giá trị.
- (c) Trong trường hợp có phương sai thay đổi phương pháp OLS thông thường luôn luôn ước lượng quá cao sai số chuẩn của các ước lượng.
- (d) Nếu các phần dư ước lượng từ một hồi quy cho thấy có một mẫu hình hệ thống, điều đó có nghĩa là có phương sai thay đổi trong số liệu.
- (e) Không có một kiểm định tổng quát nào về phương sai thay đổi mà không cần có bất kỳ giả thiết về biến số nào mà số hạng sai số có tương quan với.
- (f) Nếu một mô hình hồi quy bị xác định sai (ví dụ một biến quan trọng bị bỏ sót), các phần dư OLS sẽ có một mẫu hình rõ nét.
- (g) Nếu một biến làm hồi quy có phương sai thay đổi bị bỏ sót (một cách không đúng đắn), các phần dư (OLS) sẽ thay đổi.

11.2 Trong một hồi quy mức lương trung bình (W) theo số lao động (N) trong một mẫu ngẫu nhiên gồm 30 công ty, ta có các kết quả hồi quy sau:*

$$\begin{array}{llll} W = & 7,5 & + & 0,009N \\ t = & \text{không có} & & (16,10) \end{array} \quad R^2 = 0,90 \quad (1)$$

$$\begin{array}{llll} W/N = & 0,008 & + & 7,8 \, 1/N \\ t = & (14,43) & & (76,58) \end{array} \quad R^2 = 0,99 \quad (2)$$

- (a) Bạn giải thích hai hồi quy này như thế nào?
- (b) Đây là giả thiết của tác giả khi đi từ Phương trình (1) sang (2)? Ông ta có lo lắng về phương sai thay đổi hay không? Làm sao mà bạn biết được?
- (c) Bạn có thể tìm thấy mối liên quan giữa các hệ số độ dốc và tung độ gốc của hai mô hình không?
- (d) Bạn có thể so sánh các giá trị R^2 của hai mô hình không? Tại sao có hay tại sao không?

11.3. (a) Bạn có thể ước lượng các thông số của các mô hình

$$\begin{aligned} \hat{u}_i^2 &= \sqrt{\beta_1 + \beta_2 X_i} + v_i \\ \hat{u}_i^2 &= \sqrt{\beta_1 + \beta_2 X_i^2} + v_i \end{aligned}$$

bằng phương pháp bình phương tối thiểu thông thường được không? Tại sao được hay tại sao không?

* Xem Dominick Salvatore, *Managerial Economics* (Kinh tế học Quản lý), McGraw-Hill, New York, 1989, trang 157.

- (b) Nếu không, bạn có thể đưa ra một phương pháp, không chính thức hoặc chính thức, để ước lượng các thông số của các mô hình như trên được không?
- 11.4.** Mặc dù các mô hình log như trong (11.6.12) thường làm giảm phương sai thay đổi, ta phải cẩn thận đến tính chất của yếu tố nhiễu của những mô hình như thế. Ví dụ, mô hình

$$Y_i = \beta_1 X_i^{\beta_2} u_i \quad (1)$$

có thể được viết thành

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + \ln u_i \quad (2)$$

- (a) Nếu $\ln u_i$ có kỳ vọng bằng 0, phân phối của u_i phải như thế nào?
- (b) Nếu $E(u_i) = 1$, $E(\ln u_i)$ có bằng 0 không? Tại sao có hay tại sao không?
- (c) Nếu $E(\ln u_i)$ khác 0, có thể thực hiện biện pháp gì để cho nó bằng 0?
- 11.5.** Chứng tỏ rằng β_2^* trong (11.3.8) cũng có thể được biểu diễn bằng

$$\beta_2^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{2*}}$$

và $\text{var}(\beta_2^*)$ trong (11.3.9) cũng có thể được biểu diễn bằng

$$\text{var}(\beta_2^*) = \frac{1}{\sum w_i x_i^{2*}}$$

với $y_i^* = Y_i - \bar{Y}^*$ và $x_i^* = X_i - \bar{X}^*$ đại diện cho các độ lệch từ giá trị trung bình trọng số của $\bar{Y}^*$ và $\bar{X}^*$ định nghĩa như sau

$$\bar{Y}^* = \sum w_i Y_i / \sum w_i$$

$$\bar{X}^* = \sum w_i X_i / \sum w_i$$

- 11.6.** Phục vụ cho các mục đích sư phạm, Hanushek & Jackson ước lượng mô hình sau:

$$C_t = \beta_1 + \beta_2 \text{GNP}_t + \beta_3 D_t + u_t \quad (1)$$

với C_t = tổng chi tiêu tiêu dùng tư nhân trong năm t , GNP_t = tổng sản phẩm quốc dân trong năm t , và D = chi tiêu quốc phòng trong năm t , mục đích của phân tích là nghiên cứu tác động của chi tiêu quốc phòng tới các loại chi tiêu khác trong nền kinh tế.

Mặc định rằng $\sigma_t^2 = \sigma^2(\text{GNP}_t)^2$, các tác giả đã biến đổi (1) và ước lượng

$$C_t/\text{GNP}_t = \beta_1(1/\text{GNP}_t) + \beta_2 + \beta_3(D_t/\text{GNP}_t) + u_t/\text{GNP}_t \quad (2)$$

Sau đây là các kết quả thực nghiệm dựa vào số liệu trong giai đoạn 1946-1975 (các sai số chuẩn được trình bày trong ngoặc)*:

$$C_t = 26,19 + 0,6248 \text{GNP}_t - 0,4398 D_t$$

$$(2,73) \quad (0,0060) \quad (0,0736) \quad R^2 = 0.999$$

* Eric A. Hanushek & John E. Jackson, *Statistical Methods for Social Scientists* (Các phương pháp thống kê cho các nhà khoa học xã hội), Academic, New York, 1977, trang 160.

$$\widehat{C_t/ \text{GNP}_t} = 25,92(1/\text{GNP}_t) + 6242 - 0,4315(D_t/ \text{GNP}_t) \quad R^2 = 0,875$$

(2,22) (0,0068) (0,0597)

- (a) Các tác giả đã có giả thiết gì về bản chất của phương sai thay đổi? Bạn có thể xác minh giả thiết được không?
- (b) So sánh các kết quả của hai hồi quy. Phép biến đổi mô hình gốc có cải thiện kết quả không, tức là, có giảm các sai số chuẩn ước lượng không? Tại sao có hay tại sao không?
- (c) Bạn có thể so sánh hai giá trị R^2 không? Tại sao có hay tại sao không? (Gợi ý: Xem xét các biến phụ thuộc).
- 11.7.** Tham chiếu các hồi quy ước lượng (11.6.2) và (11.6.3). Các kết quả hồi quy khá giống nhau. Điều gì có thể tạo ra kết quả này? (Gợi ý: Tham khảo bài tập 11.13).
- 11.8.** Chứng minh rằng nếu $w_i = w$ là hằng số, đối với mỗi i , β_2^* và $\hat{\beta}_2$ cũng như phương sai của nó là đồng nhất.
- 11.9.** Tham chiếu các công thức (11.2.2) và (11.2.3). Giả sử

$$\sigma_i^2 = \sigma^2 k_i$$

với σ^2 là hằng số và k_i là các trọng số đã biết, không nhất thiết phải bằng nhau.

Sử dụng giả thiết này, hãy chứng tỏ rằng phương sai trong (11.2.2) có thể được biểu diễn là:

$$\text{var}(\beta_2) = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 k_i}{\sum x_i^2}$$

Số hạng thứ nhất ở vế phải là công thức phương sai trong (11.2.3), tức là $\text{var}(\beta_2)$ trong trường hợp phương sai không thay đổi. Bạn có thể nói gì về bản chất của mối quan hệ giữa $\text{var}(\beta_2)$ trong trường hợp phương sai không thay đổi và thay đổi? (Gợi ý: Xem xét số hạng thứ hai ở vế phải của công thức trên). Bạn có thể đưa ra các kết luận tổng quát nào về các mối quan hệ giữa (11.2.2) và (11.2.3)?

- 11.10.** Trong mô hình

$$Y_i = \beta_2 X_i + u_i \quad (\text{Lưu ý: không có tung độ gốc})$$

bạn được cho biết rằng $\text{var}(u_i) = \sigma^2 X_i^2$. Chỉ ra rằng

$$\text{var}(\beta_2) = (\sigma^2 \sum X_i^4) / (\sum X_i^2)^2$$

Bài tập

- 11.11.** Với số liệu trong Bảng 11.1, thực hiện hồi quy mức thù lao (lương) trung bình Y theo năng suất trung bình X , xem quy mô lao động là đơn vị quan sát. Giải thích các kết quả của bạn, và xem liệu các kết quả của bạn có phù hợp với các kết quả trong (11.5.3) hay không?
- (a) Từ hồi quy trên, tính các phần dư u_i .
- (b) Theo kiểm định Park, thực hiện hồi quy $\ln \hat{u}_i^2$ theo $\ln X_i$ và kiểm chứng lại hồi quy (11.5.4)

- (c) Theo phương pháp Glejser, thực hiện hồi quy $|\hat{u}_i|$ theo X_i và sau đó hồi quy $|\hat{u}_i|$ theo $\sqrt{X_i}$ và bình luận về các kết quả của bạn.
- (d) Tìm tương quan hạng giữa $|\hat{u}_i|$ và X_i và bình luận về bản chất của phương sai thay đổi, nếu có, trong số liệu.

11.12. Bảng sau cho biết số liệu về tỷ lệ doanh thu/tiền mặt trong các ngành công nghiệp chế tạo của Hoa Kỳ phân theo quy mô tài sản của doanh nghiệp trong giai đoạn 1971-I đến 1973-IV. (Số liệu trên cơ sở quý). Tỷ lệ doanh thu/tiền mặt có thể được coi là đại lượng đo tốc độ thu nhập của khu vực công ty, tức là, số lần trung chuyển đô la.

Quy mô tài sản (triệu USD)

Năm và quý	1-10	10-25	25-50	50-100	100-250	250-1.000	1.000 +
1971-I	6,696	6,929	6,858	6,966	7,819	7,557	7,860
-II	6,826	7,311	7,299	7,081	7,907	7,685	7,351
-III	6,338	7,035	7,082	7,145	7,691	7,309	7,088
-IV	6,272	6,265	6,874	6,485	6,778	7,120	6,765
1972-I	6,692	6,236	7,101	7,060	7,104	7,584	6,717
-II	6,818	7,010	7,719	7,009	8,064	7,457	7,280
-III	6,783	6,934	7,182	6,923	7,784	7,142	6,619
-IV	6,779	6,988	6,531	7,416	7,279	3,928	6,919
1973-I	7,291	7,428	7,272	7,571	7,583	7,053	6,630
-II	7,766	9,071	7,818	8,692	8,608	7,571	6,805
-III	7,733	8,357	8,090	8,357	7,680	7,654	6,772
-IV	8,316	7,621	7,766	7,867	7,666	7,380	7,072

Nguồn: Quarterly Financial Report for Manufacturing Corporations, Federal Trade Commission and the Securities and Exchange Commission, (Báo cáo Tài chính Hàng Quý cho các Công ty Công nghiệp Chế tạo, Ủy ban Thương mại Liên Bang và Ủy ban Chứng khoán và Giao dịch), Chính phủ Hoa Kỳ, các số khác nhau (đã tính toán).

- (a) Với mỗi quy mô tài sản, tính giá trị trung bình và độ lệch chuẩn của tỷ lệ doanh thu/tiền mặt.
- (b) Vẽ giá trị trung bình theo độ lệch chuẩn tính trong (a), sử dụng quy mô tài sản làm đơn vị quan sát.
- (c) Bằng phương tiện của một mô hình hồi quy thích hợp, quyết định xem độ lệch chuẩn của tỷ số có tăng lên theo giá trị trung bình không. Nếu không, làm sao bạn có thể hợp lý hóa kết quả?
- (d) Nếu có quan hệ mang ý nghĩa thống kê giữa hai biến, làm sao bạn biến đổi số liệu để không còn phương sai thay đổi?

11.13 Kiểm định phương sai không thay đổi của phương sai của Bartlett.* Giả sử có k phương sai mẫu độc lập, $s_1^2, s_2^2, \dots, s_k^2$, với $f_1, f_2, \dots, f_k$ bậc tự do, mỗi phương sai từ các tổng thể trên theo phân phối chuẩn với giá trị trung bình μ và phương sai σ_i^2 . Giả sử tiếp rằng ta muốn kiểm định giả thiết không $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$; tức là, mỗi phương sai mẫu là một ước lượng của cùng một phương sai tổng thể σ^2 .

* Xem "Properties of Sufficiency and Statistical Tests", *Proceedings of the Royal Society of London* (Các tính chất đủ và kiểm định thống kê, *Tài liệu lưu của Hiệp hội Hoàng gia Luân Đôn*), A, tập 160, 1937, trang 268.

Nếu giả thiết H_0 đúng, thì

$$s^2 = \frac{\sum_{i=1}^k f_i s_i^2}{\sum f_i} = \frac{\sum f_i s_i^2}{f}$$

cho ta một ước lượng của ước lượng phương sai tổng thể chung (tổng hợp) σ^2 , với $f_i = (n_i - 1)$, n_i là số các quan sát trong nhóm thứ i và với $f = \sum_{i=1}^k f_i$.

Barlett đã chỉ ra rằng *giả thiết không* có thể được kiểm định bằng tỷ số A/B , có phân phối gần đúng như phân phối χ^2 với $k - 1$ bậc tự do, và

$$A = f \ln s^2 - \sum (f_i \ln s_i^2)$$

và

$$B = 1 + \frac{1}{3(k-1)} \left[\sum \left(\frac{1}{f_i} \right) - \frac{1}{f} \right]$$

Áp dụng kiểm định của Barlett cho số liệu trong Bảng 11.1 và kiểm chứng rằng giả thiết các phương sai tổng thể của mức lương bằng nhau trong từng nhóm quy mô lao động của doanh nghiệp không thể bị bác bỏ tại mức ý nghĩa 5%.

Lưu ý: f_i , số bậc tự do của từng phương sai mẫu, là 9, do n_i của từng mẫu (nghĩa là nhóm quy mô lao động) bằng 10.

11.14. Số liệu trong bảng sau là mức lương trung vị của các nhà kinh tế nam và nữ theo lĩnh vực chuyên môn trong năm 1964.

- Tìm mức lương trung bình và độ lệch chuẩn của mức lương của hai nhóm các nhà kinh tế.
- Có sự khác biệt đáng kể không giữa hai độ lệch chuẩn? (Bạn có thể sử dụng kiểm định Barlett).
- Giả sử bạn muốn dự báo mức lương trung vị của các nhà kinh tế nam từ mức lương trung vị của các nhà kinh tế nữ. Hãy xây dựng một mô hình hồi quy thích hợp cho mục đích này. Nếu bạn dự đoán có phương sai thay đổi trong mô hình này, làm sao bạn có thể xử lý nó?

Lĩnh vực chuyên môn hóa	Mức lương trung vị (nghìn USD)	
	Nam	Nữ
Kinh doanh, tài chính, v.v...	9,3	13,0
Kinh tế lao động	10,3	12,0
Tiền tệ-ngân sách	8,0	11,6
Lý thuyết kinh tế chung	8,7	10,8
Dân số, các chương trình phúc lợi, v.v...	12,0	11,5
Các hệ thống kinh tế và phát triển kinh tế	9,0	12,2

Nguồn: “The Structure of Economists’ Employment and Salaries” (Cơ cấu việc làm và mức lương của các nhà kinh tế), Ủy ban của Hội Khoa học Quốc gia, Báo cáo về Nghề Kinh tế, Tạp chí Kinh tế Hoa Kỳ, tập 55, số 4, 12/1965, trang 62.

11.15. Số liệu sau đây cho ta mức lương trung vị của các nhà kinh tế phân theo bằng cấp và tuổi:

Tuổi (năm)	Mức lương trung vị (nghìn USD)	
	Bằng cao học	Bằng tiến sĩ
25-29	8,0	8,8
30-34	9,2	9,6
35-39	11,0	11,0
40-44	12,8	12,5
45-49	14,2	13,6
50-54	14,7	14,3
55-59	14,5	15,0
60-64	13,5	15,0
65-69	12,0	15,0

Nguồn: “The Structure of Economists’ Employment and Salaries” (Cơ cấu việc làm và mức lương của các nhà kinh tế), Ủy ban của Hội Khoa học Quốc gia, Báo cáo về Nghề Kinh tế, Tạp chí Kinh tế Hoa Kỳ, tập 55, số 4, 12/1965, trang 37.

- (a) Các phương sai của mức lương trung vị của các nhà kinh tế có bằng cao học và tiến sĩ có bằng nhau không?
- (b) Nếu bằng nhau, làm sao bạn có thể kiểm định giả thiết cho rằng mức lương trung bình của hai nhóm bằng nhau?
- (c) Các nhà kinh tế có bằng cao học hưởng lương cao hơn các nhà kinh tế có bằng tiến sĩ trong độ tuổi từ 35 đến 54. Làm sao bạn có thể giải thích kết quả này nếu bạn tin rằng nhà kinh tế có bằng tiến sĩ phải có mức lương cao hơn nhà kinh tế có bằng cao học?

11.16. Tham chiếu số liệu về năng suất trung bình và độ lệch chuẩn của năng suất trong bảng sau. Giả sử rằng

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

với Y = năng suất trung bình trong nhóm lao động thứ i , X_i = quy mô lao động tính bằng 1 = 1-4 lao động, 2 = 5-9 lao động, ..., 9 = 1000-2499 lao động.

Sử dụng số liệu, ước lượng hồi quy bình phương tối thiểu có trọng số và không có trọng số theo Phương trình (11.6.2) và (11.6.3) và bình luận về kết quả thu được.

Quy mô lao động (số lao động)	Năng suất trung bình (USD)	Độ lệch chuẩn của năng suất (USD)
1-4	9.355	2.487
5-9	8.584	2.642
10-19	7.962	3.055
20-49	8.275	2.706
50-99	8.389	3.119
100-249	9.418	4.493
250-499	9.795	4.910
500-999	10.281	5.893
1.000-2.499	11.750	5.550

Nguồn: *Census of Manufactures* (Tổng điều tra công nghiệp chế tạo), Bộ Thương mại Hoa Kỳ, 1958 (đã tính toán).

11.17. Trong một điều tra khoảng 9966 nhà kinh tế trong năm 1964, các số liệu sau đây được tính:

Tuổi (năm)	Mức lương trung vị (USD)
20-24	7.800
25-29	8.400
30-34	9.700
35-39	11.500
40-44	13.000
45-49	14.800
50-54	15.000
55-59	15.000
60-64	15.000
65-69	14.500
70 +	12.000

Nguồn: “The Structure of Economists’ Employment and Salaries”
(Cơ cấu việc làm và mức lương của các nhà kinh tế), Ủy ban của Hội
Khoa học Quốc gia, Báo cáo về Nghề Kinh tế, Tạp chí Kinh tế Hoa
Kỳ, tập 55, số 4, 12/1965, trang 36.

(a) Xây dựng một mô hình hồi quy hợp lý để giải thích mức lương trung vị trong quan hệ với tuổi.

Lưu ý: Để phục vụ cho mục đích hồi quy, giả sử rằng các mức lương trung vị ứng với điểm giữa của các lớp tuổi. Tức là, 7.800 USD ứng với độ tuổi 22,5 năm, và v.v... Đối với lớp tuổi cuối cùng, giả sử độ tuổi tối đa là 75 năm.

(b) Giả sử rằng phương sai của số hạng nhiễu tỷ lệ với bình phương độ tuổi, biến đổi số liệu để làm cho các số hạng nhiễu có phương sai không thay đổi.

(c) Lặp lại (b), giả sử rằng phương sai tỷ lệ với độ tuổi. Phép biến đổi nào có vẻ hợp lý hơn?

(d) Nếu không có phép biến đổi nào ở trên hợp lý, giả sử rằng phương sai tỷ lệ với kỳ vọng có điều kiện của mức lương trung vị, tùy thuộc theo độ tuổi nhất định. Làm sao bạn có thể biến đổi số liệu để tạo ra phương sai không thay đổi?

11.18 Bảng sau cho biết số liệu về quyền rút tiền đặc biệt (SDR), còn được gọi là *vàng giấy*, và cán cân thanh toán của 10 quốc gia trong năm 1974.

Quốc gia	SDR, triệu USD	Cán cân thanh toán, triệu USD
Bi	715	346
Canada	574	26
Pháp	248	-83*
Đức	1.763	-466
Ý	221	-4.633
Nhật Bản	529	1.241
Hà Lan	595	985
Thụy Điển	131	-802
Anh Quốc	843	-4.355
Hoa Kỳ	2.370	-8.374

* Dấu âm biểu thị thâm thủng cán cân thanh toán

Nguồn: International Financial Statistics (Thống kê Tài chính Quốc tế), Quỹ Tiền tệ Quốc tế, 12/1975.

Bởi vì SDR được sử dụng làm đồng tiền quốc tế, quy mô SDR được dự kiến là có quan hệ với tình hình cán cân thanh toán của một quốc gia.

- (a) Từ số liệu ở trên, có một mối quan hệ rõ nét nào giữa SDR và cán cân thanh toán hay không? Trả lời bằng hồi quy của SDR theo cán cân thanh toán.
- (b) Sử dụng kết quả của bạn trong (a), kiểm định riêng rẽ các giả thiết sau: phương sai của nhiễu tỷ lệ với
- (i) bình phương giá trị cán cân thanh toán
- (ii) giá trị kỳ vọng có điều kiện của SDR tùy thuộc theo giá trị của cán cân thanh toán.
- (c) Bạn có thể sử dụng phép biến đổi log thảo luận ở phần trên để biến đổi số liệu không? Tại sao hay tại sao không?
- (d) Áp dụng kiểm định tương quan hạng đối với u_i tính được từ hồi quy trong (a) và số liệu cán cân thanh toán. Bạn có thể nói gì không về phương sai thay đổi trên cơ sở của kiểm định này?

11.19. Bạn được cho biết các số liệu sau:

RSS_1 dựa vào 30 quan sát đầu = 55, số bậc tự do = 25

RSS_2 dựa vào 30 quan sát đầu = 140, số bậc tự do = 25

Thực hiện kiểm định Goldfeld-Quandt về phương sai thay đổi ở mức ý nghĩa 5%.

11.20. Bảng sau cho biết số liệu về tỷ lệ thay đổi phần trăm giá cổ phiếu (Y) và giá tiêu dùng (X) trong số liệu chéo của 20 quốc gia.

Giá cổ phiếu và giá tiêu dùng, giai đoạn sau Chiến tranh Thế giới thứ II (tới 1969)

Quốc gia	Tốc độ thay đổi (%/năm)	
	Giá cổ phiếu, Y	Giá tiêu dùng, X
1. Úc	5,0	4,3
2. Ao	11,1	4,6
3. Bỉ	3,2	2,4
4. Canada	7,9	2,4
5. Chilê	25,5	26,4
6. Đan Mạch	3,8	4,2
7. Phần Lan	11,1	5,5
8. Pháp	9,9	4,7
9. Đức	13,3	2,2
10. Ấn Độ	1,5	4,0
11. Ai Len	6,4	4,0
12. Israel	8,9	8,4
13. Ý	8,1	3,3
14. Nhật Bản	13,5	4,7
15. Mêhicô	4,7	5,2
16. Hà Lan	7,5	3,6
17. Niu Di Lân	4,7	3,6
18. Thụy Điển	8,0	4,0
19. Anh Quốc	7,5	3,9

20. Hoa Kỳ

9,0

2,1

Nguồn: Phillip Cagan, *Common Stock Values and Inflation: The Historical Record of Many Countries*, National Bureau of Economic Research (Giá trị cổ phiếu thường và lạm phát: thống kê lịch sử của nhiều quốc gia, Cục Nghiên cứu Kinh tế Quốc gia), tài liệu bổ sung, 3/1974, Bảng 1, trang 4.

(a) Vẽ số liệu lên đồ thị phân tán.

(b) Thực hiện hồi quy Y theo X và xem xét các phần dư từ hồi quy này. Bạn quan sát thấy gì?

(c) Do số liệu Chilê có vẻ như không tiêu biểu (hay nằm ngoài), lặp lại hồi quy trong (b), bỏ số liệu Chilê. Bây giờ xem xét các phần dư từ hồi quy này. Bạn quan sát thấy gì?

(d) Nếu trên cơ sở của các kết quả trong (b) bạn kết luận là có phương sai thay đổi trong phương sai của sai số nhưng trên cơ sở các kết quả trong (c) bạn đảo ngược kết luận của mình, bạn rút ra kết luận tổng quát gì?

11.21 Tính các hồi quy tương tự như (11.6.2) và (11.6.3) bằng cách tính quy mô lao động theo trung điểm của nhóm trong các nhóm quy mô lao động (ví dụ 2,5, 7,0) và so sánh kết quả của bạn với các kết quả trong (11.6.2) và (11.6.3).

Bạn thích phương pháp tính quy mô lao động nào và tại sao? Đây là vấn đề khó khăn tiềm năng nếu bạn có một nhóm quy mô lao động mở, ví dụ, 2500 lao động hay nhiều hơn?

11.22. Tham chiếu bảng trong bài tập 11.8. Cộng thêm số 8500 vào tất cả các số liệu cân cân thanh toán trong bảng, và

(a) thực hiện hồi quy các SDR theo các số liệu cân cân thanh toán mới sử dụng mô hình tuyến tính.

(b) thực hiện hồi quy các SDR theo các số liệu cân cân thanh toán mới sử dụng mô hình tuyến tính log hay mô hình log kép.

(c) so sánh hai kết quả hồi quy và bình luận các kết quả của bạn.

(d) Tại sao phải cộng thêm số 8500 vào số liệu cân cân thanh toán? Thêm bất cứ số khác có được không?

11.23. Tham chiếu số liệu trong Bảng 11.5. Thực hiện hồi quy chi tiêu nghiên cứu và phát triển (R&D) theo doanh thu và tìm trên cơ sở của kiểm định Breusch-Pagan và White xem phương sai của sai số trong hồi quy này có không thay đổi hay không.

11.24. Lặp lại bài tập 11.23, lần này thực hiện hồi quy R&D theo lợi nhuận.

11.25. Tại Ví dụ 11.8 thảo luận trong Mục 11.7, tính các sai số chuẩn đã hiệu chỉnh phương sai thay đổi của White trong hồi quy chi tiêu R&D theo doanh thu và so sánh các kết quả của bạn với hồi quy (11.7.1). Bạn rút ra kết luận quan trọng nào từ ví dụ này?

PHỤ LỤC 11A**11.A.1. CHỨNG MINH PHƯƠNG TRÌNH (11.2.2)**

Từ Phụ lục 3A, Mục 3A.3, ta có

$$\begin{aligned}\text{var}(\hat{\beta}_2) &= E(k_1^2 u_1^2 + k_2^2 u_2^2 + \dots + k_n^2 u_n^2 + 2 \text{ tích chéo của hai số hạng}) \\ &= E(k_1^2 u_1^2 + k_2^2 u_2^2 + \dots + k_n^2 u_n^2)\end{aligned}$$

do các kỳ vọng của các số hạng tích chéo bằng không theo giả thiết rằng không có tương quan chuỗi.

$$\text{var}(\hat{\beta}_2) = k_1^2 E(u_1^2) + k_2^2 E(u_2^2) + \dots + k_n^2 E(u_n^2)$$

do k_i đã biết. (Tại sao?)

$$\text{var}(\hat{\beta}_2) = k_1^2 \sigma_1^2 + k_2^2 \sigma_2^2 + \dots + k_n^2 \sigma_n^2$$

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

$$\begin{aligned}\text{var}(\hat{\beta}_2) &= \sum k_i^2 \sigma_i^2 \\ &= \sum \left[\left(\frac{x_i}{\sum x_i^2} \right)^2 \sigma_i^2 \right] \quad \text{do } k_i = \frac{x_i}{\sum x_i^2} = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \quad (11.2.2)\end{aligned}$$

11.A.2 PHƯƠNG PHÁP BÌNH PHƯƠNG TỐI THIỂU CÓ TRỌNG SỐ

Để minh họa phương pháp này, ta sử dụng mô hình hai biến $Y_i = \beta_1 + \beta_2 X_i + u_i$. Phương pháp bình phương tối thiểu không có trọng số sẽ cực tiểu hóa

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \quad (1)$$

để tính các ước lượng, trái lại phương pháp bình phương tối thiểu có trọng số sẽ cực tiểu hóa tổng bình phương phần dư có trọng số:

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)^2 \quad (2)$$

với $\hat{\beta}_1^*$ và $\hat{\beta}_2^*$ là các ước lượng bình phương tối thiểu có trọng số và với các trọng số w_i tính như sau:

$$w_i = \frac{1}{\sigma_i^2} \quad (3)$$

tức là, các trọng số tỷ lệ nghịch với phương sai của u_i hay Y_i , tùy thuộc vào giá trị cụ thể của X_i . Ta hiểu rằng $\text{var}(u_i | X_i) = \text{var}(Y_i | X_i) = \sigma_i^2$.

Lấy vi phân (2) theo $\hat{\beta}_1^*$ và $\hat{\beta}_2^*$, ta có

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_1^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i) (-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_2^*} = 2 \sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i) (-X_i)$$

Cho các biểu thức trên bằng 0, ta có hai phương trình chuẩn sau:

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \quad (4)$$

$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2 \quad (5)$$

Chú ý sự giống nhau giữa các phương trình chuẩn và các phương trình chuẩn của phương pháp bình phương tối thiểu không có trọng số.

Giải các phương trình này đồng thời, ta có

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \quad (6)$$

và

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \quad (7)=(11.3.8)$$

Phương sai của $\hat{\beta}_2^*$ trong (11.3.9) có thể tính được theo cách tính β_2 trong Phụ lục 3A, Mục 3A.3.

Chú ý: $\bar{Y}^* = \sum w_i Y_i / \sum w_i$ và $\bar{X}^* = \sum w_i X_i / \sum w_i$. Như có thể được kiểm chứng dễ dàng, các trung bình trọng số này trùng với các trung bình thông thường hay không có trọng số $\bar{Y}$ và $\bar{X}$ khi $w_i = w =$ hằng số, với mọi i .