Chẩn đoán phần dư

Như đã nêu trong một bài giảng trước đây, các phần dư bình phương tối thiểu phản ánh các tính chất của những thành phần nhiễu ngẫu nhiên. Do điều này đúng, nên chúng ta có thể sử dụng những phần dư này để kiểm tra một số trong các giả định đã đưa ra về những nhiễu ngẫu nhiên. Làm việc này là quan trọng vì những tính chất thống kê tốt đẹp của các hàm ước lượng bình phương tối thiểu phụ thuộc vào các giả định cổ điển.

Chúng ta sẽ kiểm tra hai vi phạm các giả định cổ điển : sự không nhất quán của phương sai và tự tương quan. Trong mỗi trường hợp chúng ta sẽ xác định vấn đề, kiểm tra những hậu quả trên các hàm ước lượng bình phương tối thiểu, xác định các kiểm định đối với sự hiện diện của những vấn đề này, và cuối cùng chúng ta sẽ đề xuất các hàm ước lượng có các tính chất thống kê tốt đẹp.

Tính không nhất quán của phương sai

Định nghĩa

Trong trường hợp thứ nhất mà chúng ta sẽ xem xét là tình thế xuất hiện khi các phương sai của sai số là không bất biến trên các quan sát. Thay cho :

$$E\left[\varepsilon_i^2\right] = \sigma^2$$

chúng ta có điều này :

$$E\left[\varepsilon_i^2\right] = \sigma_i^2$$

Tính không nhất quán của phương sai thường gặp trong các tập hợp dữ liệu chéo và dữ liệu dạng bảng.

Các tính chất thống kê của các hàm ước lượng bình phương tối thiểu khi hiện diện tính không nhất quán của phương sai

Xét mô hình hồi qui đơn biến :

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

Chúng ta sẽ tập trung vào hàm ước lượng bình phương tối thiểu của hệ số độ dốc:

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \beta_2 + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}$$

Giá trị kỳ vọng là :

$$E\left[\hat{\beta}_2\right] = \beta_2 + \frac{\sum x_i E[\varepsilon_i]}{\sum x_i^2} = \beta_2$$

Sự có mặt của tính không nhất quán của phương sai không tác động lên tính chệch của hà m ước lượng bình phương tối thiểu nà y.

Bây giờ hã y xét phương sai của hà m ước lượng nà y:

$$VAR\left[\hat{\beta}_2\right] = \frac{\sum x_i^2 \, VAR[\varepsilon_i]}{\left(\sum x_i^2\right)^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

Sự giản lược mà chúng ta thu được với một phương sai không đổi không xuất hiện ở đây.

Vấn đề là liệu đây có phải là phương sai thiểu có thể đạt tới bằng bất cứ hà m ước lượng không chệch tuyến tính nà o không . Định lý Aitken chỉ ra rằng khi thà nh phần nhiễu ngẫu nhiên có phương sai không đồng nhất, thì một hà m ước lượng thay thế được gọi là hà m ước lượng bình phương tối thiểu có trọng số (WLS) là hà m ước lượng có phương sai tối thiểu.

Giả sử rằng phương sai của sai số có tính chất sau :

$$\sigma_i^2 = w_i^2 \sigma^2$$

Hã y biến mô hình ban đầu bằng cách nhân với $\frac{1}{w_i}$:

$$\frac{Y_i}{w_i} = \beta_1 \frac{1}{w_i} + \beta_2 \frac{X_i}{w_i} + \frac{\varepsilon_i}{w_i}$$

Mô hình nà y cũ ng có thể được thể hiện như sau:

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \varepsilon_i^*$$

Hã y ghi nhận 2 điểm: mô hình nà y không cò n có tung độ gốc và thà nh phần sai số mới nà y thỏa mã n các giả định cổ điển vì :

$$VAR\left[\varepsilon_i^*\right] = \frac{1}{w_i^2} VAR\left[\varepsilon_i\right] = \frac{1}{w_i^2} w_i^2 \sigma^2 = \sigma^2$$

Vì thế, nếu chúng ta áp dụng OLS cho mô hình đã biến đổi, thì định lý Gauss-Markov bảo đảm với chúng ta rằng chúng ta sẽ thu được một hà m ước lượng không chệch tuyến tính tốt nhất. Hà m ước lượng của mô hình đã biến đổi nà y trông ra sao? Sau khi là m việc một chút , chúng ta tìm ra :

$$\hat{\beta}_{2,WLS} = \frac{\sum X_{2i}^* Y_i^* \sum X_{1i}^{*2} - \sum X_{1i}^* Y_i^* \sum X_{1i}^* X_{2i}^*}{\sum X_{1i}^{*2} \sum X_{2i}^{*2} - \left(\sum X_{1i}^* X_{2i}^*\right)^2}$$

$$\hat{\beta}_{2,WLS} = \frac{\sum\left(\frac{X_i}{w_i}\right)\left(\frac{Y_i}{w_i}\right)\sum\left(\frac{1}{w_i^2}\right) - \sum\left(\frac{1}{w_i}\right)\left(\frac{Y_i}{w_i}\right)\sum\left(\frac{1}{w_i}\right)\left(\frac{X_i}{w_i}\right)}{\sum\left(\frac{1}{w_i^2}\right)\sum\left(\frac{X_i^2}{w_i^2}\right) - \left(\sum\left(\frac{1}{w_i}\right)\left(\frac{X_i}{w_i}\right)\right)^2}$$

$$\hat{\beta}_{2,WLS} = \frac{\sum\left(\frac{X_i Y_i}{w_i^2}\right)\sum\left(\frac{1}{w_i^2}\right) - \sum\left(\frac{Y_i}{w_i^2}\right)\sum\left(\frac{X_i}{w_i^2}\right)}{\sum\left(\frac{X_i^2}{w_i^2}\right)\sum\left(\frac{1}{w_i^2}\right) - \left(\sum\left(\frac{X_i}{w_i^2}\right)\right)^2}$$

Chúng ta có thể thấy rằng hà m nà y khác với hà m ước lượng OLS cho mô hình ban đầu. Do hà m ước lượng nà y là hiệu quả, nên hà m ước lượng OLS phải là không hiệu quả.

Chúng ta cũ ng nên nghiên cứu hà m ước lượng của phương sai mà phương pháp OLS tạo ra .

Chúng ta biết rằng công thức cho phương sai của hà m ước lượng OLS khi hiện hữ u tính không nhất quán của phương sai là công thức nà y :

$$VAR\left[\hat{\beta}_2\right] = \frac{\sum x_i^2 \, VAR\left[\varepsilon_i\right]}{\left(\sum x_i^2\right)^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

Nhưng khi phần mềm máy vi tính của chúng ta tính các hà m ước lượng OLS thì nó sử dụng công thức sau để tính phương sai:

$$VAR\left[\hat{\beta}_2\right] = \frac{\sigma^2}{\sum x_i^2}$$

Công thức nà y là không chính xác, vì thế các sai số chuẩn và trị thống kê t mà máy vi tính tìm ra là không chính xác : chúng là vô dụng.

Kết luận : Sự hiện diện của tính không đồng nhất của phương sai là m cho hà m ước lượng OLS là không chệch, nhưng không hiệu quả. Ngoà i ra, nhữ ng sai số chuẩn của các ước lượng và nhữ ng trị thống kê t được tính bởi thuật toán OLS là không chính xác.

Có hai khả năng xảy ra : (1) Chúng ta có thể sử dụng WLS để thu được các hà m ước lượng hiệu quả không chệch; hoặc là (2) chúng ta có thể sử dụng OLS và tìm ra một cách tính các sai số chuẩn và trị thống kê t chính xác. Mỗi phương pháp đều có phần được mà phần mất .

Bình phương tối thiểu có trọng số

Khả năng thứ nhất dựa trên việc tìm ra một hệ số tỉ lệ w_i sao cho

$$\sigma_i^2 = w_i^2\sigma^2$$

Trong một số trườ ng hợp đặc biệt , việc nà y được thực hiện dễ dà ng (như là trong các mô hình dữ liệu dạng bảng), nhưng không thông thườ ng. Trong đa số trườ ng hợp , chúng ta cố gắng tìm ra một mô hình hợp lý đối với tính không đồng nhất của phương sai dựa trên các phần dư OLS và sau đó ước lượng một số trọng số phù hợp. Quá trình nà y cũ ng cung cấp cơ sở cho việc kiểm định tính không đồng nhất của phương sai.

Các trườ ng hợp sau là nhữ ng trườ ng hợp xuất hiện thườ ng xuyên nhất trong tà i liệu. Mỗi một trườ ng hợp có thể được coi là sự chính thức hoá của một phương pháp đồ thị nhất định.

Mô hình hồi qui đa biến :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_K X_{Ki} + \varepsilon_i$$

Ước lượng phép hồi qui nà y và giữ lại những phần dư nà y. Vẽ đồ thị các bình phương phần dư , giá trị tuyệt đối của những phần dư nà y, hay là logarit của những bình phương phần dư nà y với một biến mà với nó phương sai của sai số có thể có tương quan . Hã y kiểm tra hình thái bằng mắt.

Chính thức hơn, hã y chỉ ra một trong các phép hồi qui phụ sau, trong đó biến phụ thuộc được dựa trên những phần dư từ phép hồi qui ban đầu

$$e_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_p Z_{pi} + v_i$$

$$|e_i| = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_p Z_{pi} + v_i$$

$$\log\left(e_i^2\right) = \alpha_1 + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \cdots + \alpha_p Z_{pi} + v_i$$

Những biến hồi qui trong phép hồi qui phụ nà y có thể giống hoặc không với những biến trong phép hồi qui ban đầu.

Kiểm định *Breusch-Pagan* được dựa trên sự hình thà nh thứ nhất, kiểm định *Glejser* dựa trên sự hình thà nh thứ hai, và kiểm định *Harvey-Godfrey* dựa trên sự hình thà nh thứ ba .

Giả thuyết "không" rằng không tồn tại tính không đồng nhất của phương sai là giả thuyết nà y:

$$H_0 : \alpha_2 = \alpha_3 = \cdots = \alpha_p = 0$$

Trong mỗi trường hợp , dạng mẫu lớn của trị thống kê kiểm định nà y là : nR^2 trong đó n là kích cỡ của mẫu được sử dụng để tính phép hồi qui phụ nà y và R^2 là hệ số xác định từ phép hồi qui phụ nà y.

Trong mỗi trường hợp , trị thống kê kiểm định đều có phân phối xác suất tiệm cận với Chi bình phương với (p - 1) bậc tự do. Qui tắc quyết định là :

Nếu $\chi^2_{((1-\alpha),p-1)} \leq nR^2$ thì bác bỏ.

Dạng giá trị p của qui tắc quyết định là :

Nếu $P\left(nR^2 < \chi^2_{p-1}\right) < \alpha$ thì bác bỏ.

Nếu giả thuyết nà y không bị bác bỏ, thì bằng chứng về tính không đồng nhất của phương sai không tồn tại và ước lượng OLS là phù hợp.

Nếu giả thuyết nà y bị bác bỏ , thì hà m ước lượng WLS *khả thi* có thể được triển khai bởi qui trình sau. Sử dụng từ *khả thi* chỉ ra rằng các trọng số thực (nhưng chưa biết) được thay thế bởi các trọng số ước lượng.

Từ mô hình hồi qui phụ cho ta những kết quả tốt nhất, hã y tính các trọng số bình phương "thích hợp " :

$$\hat{w}_i^2 = \hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \hat{\alpha}_3 Z_{3i} + \cdots + \hat{\alpha}_p Z_{pi}$$

$$\hat{w}_i^2 = \left(\hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \hat{\alpha}_3 Z_{3i} + \cdots + \hat{\alpha}_p Z_{pi}\right)^2$$

$$\hat{w}_i^2 = \exp\left(\hat{\alpha}_1 + \hat{\alpha}_2 Z_{2i} + \hat{\alpha}_3 Z_{3i} + \cdots + \hat{\alpha}_p Z_{pi}\right)$$

Hã y sử dụng các căn bậc hai của những trọng số bình phương thích hợp nà y để biến đổi dữ liệu ban đầu và sau đó thực hiện OLS trên mô hình đã biến đổi.

Đa số các phần mềm tự động thực hiện WLS khả thi nếu Anh/Chị có một biến chứa những trọng số nà y. Hã y cẩn thận ghi nhận xem liệu có phải phần mềm nà y yêu cầu biến nà y phải được xác định như là w_i hay $1/w_i$.

Trong Bà i tập Set 14 chúng ta sẽ là m việc nà y cù ng với EViews.

Kiểm định của White

Halbert White (1983) đã ghi nhận rằng nếu tính không đồng nhất của phương sai không tồn tại, thì những biểu thức sau sẽ như y như vậy :

$$VAR\left[\hat{\beta}_2\right] = \frac{\sum x_i^2\, VAR[\varepsilon_i]}{\left(\sum x_i^2\right)^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

$$VAR\left[\hat{\beta}_2\right] = \frac{\sigma^2}{\sum x_i^2}$$

Nói cách khác , nếu tính không đồng nhất của phương sai có mặt , thì nhữ ng biểu thức nà y sẽ khác. White đã phát triển một kiểm định đối với tính không đồng nhất của phương sai dựa trên các khác biệt giữ a nhữ ng hà m ước lượng nhất quán của hai biểu thức nà y. Kiểm định nà y là một kiểm định mẫu lớn.

Sau nà y, ngườ i ta ghi nhận rằng kiểm định của White có thể được tính toán theo cách tương tự như nhữ ng kiểm định dựa trên nhữ ng phép hồi qui phụ đã thảo luận trên đây. Đối với kiểm định White, phép hồi qui phụ nà y có chứa từ ng biến giải thích , bình phương của từ ng biến giải thích, và tổng của từ ng cặp các biến giải thích. Vì thế, đối với mô hình có hai biến hồi qui:

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

Phép hồi qui phụ nà y là :

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

Một lần nữ a, trị thống kê kiểm định là nR^2 dựa trên phép hồi qui phụ nà y. Kiểm định thống kê nà y có phân phối xác suất xấp xỉ Chi bình phương với các bậc tự do bằng số các thà nh phần trong phép hồi qui phụ nà y (trừ hằng số).

Cần cẩn thận khi xác định phép hồi qui phụ nà y nếu mô hình ban đầu có bất cứ biến giả nà o. Bình phương của một biến giả bằng chính biến giả ban đầu, và việc đưa cả hai và o trong phép hồi qui phụ nà y tạo ra tính đa cộng tuyến hoà n hảo.

May mắn thay, EViews thực hiện kiểm định của White với ba động tác nhấp chuột, như Anh/Chị sẽ thấy trong Bà i tập 14.

Như đã nêu trên đây, một phương pháp thay thế cho WLS khả thi sử dụng OLS nhưng tính các sai số chuẩn và các tỉ lệ t chính xác . Phương pháp để là mviệc nà y cũ ng do White (1983) đưa ra và được gọi là ước lượng ma trận tích sai nhất quán với tính không đồng nhất của phương sai (HCCM).

Đối với mô hình hồi qui đơn biến, HCC của White cho phương sai của hệ số độ dốc là hà m ước lượng nhất quán của

$$\mathrm{VAR}\left[\hat{\beta}_2\right] = \frac{\sum x_i^2\ \mathrm{VAR}[\ \varepsilon_i\ ]}{\left(\sum x_i^2\right)^2} = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$$

HCCM của White là m việc tốt trên các mẫu lớn thuộc loại dạng mà chúng ta vừ a ý trong nhiều công việc nghiên cứu chéo. Anh/Chị sẽ có một cơ hội sử dụng EViews để tạo ra HCCM của White.

Đôi khi, các sinh viên nghĩ một cách sai lầm rằng phương pháp White tạo ra các hà m ước lượng WLS. Nó không tạo ra đâu ! Nó tạo ra các hà m ước lượng cho các hệ số OLS và các sai số chuẩn ước lượng một cách nhất quán cho chúng.