



CHƯƠNG 5 – ĐA CỘNG TUYẾN

Nội dung

5.1. Bản chất của Đa cộng tuyến

5.2. Hậu quả của Đa cộng tuyến

5.3. Phát hiện Đa cộng tuyến

5. 4. Các biện pháp khắc phục

5.1. Bản chất của Đa cộng tuyến

5.1.1. Đa cộng tuyến

Xét mô hình hồi quy k biến:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i, k > 2$$

Khi giả thiết số 5 của phương pháp bình phương nhỏ nhất trong mô hình hồi quy bội bị vi phạm: giữa các biến giải thích $\{X_{2i}, X_{3i}, \dots, X_{ki}\}$ có mối quan hệ phụ thuộc tuyến tính, khi đó mô hình đã cho mắc khuyết tật Đa cộng tuyến.

5.1. Bản chất của Đa cộng tuyến

5.1.2. Đa cộng tuyến hoàn hảo

Đa cộng tuyến hoàn hảo (perfect multi) là hiện tượng giữa các biến giải thích $\{X_{2i}, X_{3i}, \dots, X_{ki}\}$ có quan hệ cộng tuyến hoàn toàn với nhau, biểu diễn mối quan hệ giữa các biến này dưới dạng đẳng thức:

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} = 0$$

Trong đó tồn tại ít nhất một $\lambda_j \neq 0 (j = \overline{2, k})$, Khi đó ta có:

$$X_{ji} = -\frac{\lambda_3}{\lambda_j} X_{2i} - \dots - \frac{\lambda_{j-1}}{\lambda_j} X_{j-1i} - \frac{\lambda_{j+1}}{\lambda_j} X_{j+1i} - \dots - \frac{\lambda_k}{\lambda_j} X_{ki}$$

5.1. Bản chất của Đa cộng tuyến

5.1.3. Đa cộng tuyến không hoàn hảo

Đa cộng tuyến không hoàn hảo (inperfect multi) là hiện tượng giữa các biến giải thích $\{X_{2i}, X_{3i}, \dots, X_{ki}\}$ có quan hệ cộng tuyến không hoàn toàn với nhau, biểu diễn mối quan hệ giữa các biến này dưới dạng đẳng thức:

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} + \dots + \lambda_k X_{ki} + V_i = 0$$

Trong đó tồn tại ít nhất một $\lambda_j \neq 0 (j = \overline{2, k})$, khi đó ta có:

$$X_{ji} = -\frac{\lambda_3}{\lambda_j} X_{2i} - \dots - \frac{\lambda_{j-1}}{\lambda_j} X_{j-1i} - \frac{\lambda_{j+1}}{\lambda_j} X_{j+1i} - \dots - \frac{\lambda_k}{\lambda_j} X_{ki} + V_i$$

5.1. Bản chất của Đa cộng tuyến

5.1.4. Nguyên nhân

- Do bản chất kinh tế xã hội các biến kinh tế ít nhiều có quan hệ tuyến tính với nhau.
- Do số liệu mẫu không ngẫu nhiên, hoặc do kích thước mẫu không đủ lớn nên không đại diện tốt nhất cho tổng thể.
- Do quá trình xử lý, tính toán số liệu.
- Do chỉ định sai dạng hàm và một số nguyên nhân khác.

5.2. Hậu quả của Đa cộng tuyến

5.2.1. Hậu quả khi có Đa cộng tuyến hoàn hảo

Xét mô hình 3 biến: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$

Các hệ số hồi quy ước lượng được xác định bằng công thức:

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

5.2. Hậu quả của Đa cộng tuyến

- Nếu mô hình có đa cộng tuyến hoàn hảo thì không thể ước lượng được các hệ số hồi quy và phương sai của chúng là vô hạn.

$$\text{Giả sử: } X_{2i} = \lambda X_{3i} \Leftrightarrow x_{2i} = \lambda x_{3i} \quad (\lambda \neq 0) \Rightarrow r_{23} = r_{32} = 1$$

Khi đó ta có:

$$\hat{\beta}_2 = \frac{(\sum_{i=1}^n x_{2i} y_i)(\lambda^2 \sum_{i=1}^n x_{2i}^2) - (\lambda \sum_{i=1}^n x_{2i} y_i)(\lambda^2 \sum_{i=1}^n x_{2i}^2)}{(\sum_{i=1}^n x_{2i}^2)(\lambda^2 \sum_{i=1}^n x_{2i}^2) - \lambda^2 (\sum_{i=1}^n x_{2i}^2)^2} = \frac{0}{0}$$

$$\hat{\beta}_3 = \frac{(\sum_{i=1}^n x_{3i} y_i)(\lambda^2 \sum_{i=1}^n x_{3i}^2) - (\sum_{i=1}^n x_{3i} y_i)(\lambda^2 \sum_{i=1}^n x_{3i}^2)}{(\sum_{i=1}^n x_{3i}^2)(\lambda^2 \sum_{i=1}^n x_{3i}^2) - \lambda^2 (\sum_{i=1}^n x_{3i}^2)^2} = \frac{0}{0}$$

5.2. Hậu quả của Đa cộng tuyến

5.2.2. Hậu quả khi có Đa cộng tuyến không hoàn hảo

- Nếu mô hình có đa cộng tuyến không hoàn hảo thì vẫn có thể ước lượng được các hệ số hồi quy.

Giả sử: $X_{2i} = \lambda X_{3i} + V_i \Leftrightarrow x_{2i} = \lambda x_{3i} + v_i \quad (\lambda \neq 0)$

Khi đó ta có:

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\lambda^2 \sum x_{2i}^2 + \sum v_i^2\right) - \left(\lambda \sum y_i x_{2i} + \sum y_i v_i\right)\left(\lambda \sum x_{2i}^2\right)}{\left(\sum x_{2i}^2\right)\left(\lambda^2 \sum x_{2i}^2 + \sum v_i^2\right) - \lambda^2 \left(\sum x_{2i}^2\right)^2}$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\lambda^2 \sum x_{3i}^2 + \sum v_i^2\right) - \left(\lambda \sum y_i x_{3i} + \sum y_i v_i\right)\left(\lambda \sum x_{3i}^2\right)}{\left(\sum x_{3i}^2\right)\left(\lambda^2 \sum x_{3i}^2 + \sum v_i^2\right) - \lambda^2 \left(\sum x_{3i}^2\right)^2}$$

5.2. Hậu quả của Đa cộng tuyến

- Phương sai của các ước lượng OLS tiến dần ra vô hạn:

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum_{i=1}^n x_{2i}^2 (1 - r_{23}^2)} \rightarrow +\infty, Var(\hat{\beta}_3) = \frac{\sigma^2}{\sum_{i=1}^n x_{3i}^2 (1 - r_{23}^2)} \rightarrow +\infty$$

- Khoảng tin cậy của các hệ số hồi quy rộng
- Thống kê T mất ý nghĩa khi đó các hệ số hồi quy không có ý nghĩa thống kê.
- Hệ số R^2 có thể khá cao nhưng các giá trị thống kê của kiểm định T khá nhỏ, điều này tạo ra sự mâu thuẫn về ý nghĩa của hệ số R^2 . Khi đó các kiểm định T và F có thể cho kết luận mâu thuẫn vì kiểm định F luôn có xu hướng bác bỏ H_0 .

5.2. Hậu quả của Đa cộng tuyến

- Dấu của các hệ số hồi quy ước lượng được có thể không phù hợp với lý thuyết kinh tế.
- Các ước lượng và sai số chuẩn rất nhạy với sự thay đổi của số liệu.
- Ước lượng của các hệ số hồi quy có thể thay đổi đáng kể về dấu và số lượng khi thêm bớt các biến cộng tuyến.

5.2. Hậu quả của Đa cộng tuyến

Lưu ý:

- ✓ Thực tế đa cộng tuyến hoàn hảo gần như không xảy ra vì sự phụ thuộc hàm số giữa các biến độc lập chỉ tồn tại về mặt lý thuyết. Do đó khi nói đến đa cộng tuyến thì chúng ta hiểu đó là hiện tượng đa cộng tuyến không hoàn hảo.
- ✓ Đối với mô hình hồi quy bội người ta không xem xét có hay không có đa cộng tuyến mà xem xét mức độ nghiêm trọng của đa cộng tuyến.

5.3. Phát hiện Đa cộng tuyến

5.3.1. So sánh R^2 và tỷ số t

- Trong trường hợp R^2 có giá trị cao, đồng thời giá trị tuyệt đối của tỷ số t thấp thì có thể là dấu hiệu của hiện tượng Đa cộng tuyến.
- Ví dụ 1: Hồi quy sản lượng (Y – nghìn tấn) theo vốn kinh doanh của doanh nghiệp (X_2 - tỷ đồng) và số lượng lao động (X_3 – trăm người). Sử dụng báo cáo 1, trang 137 giáo trình.

Kết quả hồi quy cho thấy $R^2 = 0.98$, trong khi đó các tỷ số t có trị số rất nhỏ. Đây là dấu hiệu để nhận biết mô hình ban đầu có thể có đa cộng tuyến.

5.3. Phát hiện Đa cộng tuyến

5.3.2. Xét tương quan cặp giữa các biến giải thích

Nếu trị số tuyệt đối của hệ số tương quan cặp (r_{ij}) giữa các biến giải thích khá cao ($|r_{ij}| > 0,8$) thì có khả năng tồn tại đa cộng tuyến. Tuy nhiên, điều này có thể không hoàn toàn chính xác. Trong nhiều trường hợp giá trị này nhỏ vẫn xảy ra hiện tượng Đa cộng tuyến.

5.3. Phát hiện Đa cộng tuyến

5.3.3. Tương quan riêng

Xét mô hình 3 biến: $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + e_i$

Ký hiệu:

- $r_{12,3}$ là hệ số tương quan riêng giữa Y và X_2 trong khi X_3 không đổi
- $r_{13,2}$ là hệ số tương quan riêng của Y và X_3 trong khi X_2 không đổi

$$r_{123} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

trong đó:

- r_{12} là hệ số tương quan giữa Y và X_2
- r_{13} là hệ số tương quan giữa Y và X_3
- r_{23} là hệ số tương quan giữa X_2 và X_3

5.3. Phát hiện Đa cộng tuyến

Nếu hệ số tương quan giữa Y với tất cả các biến giải thích cao, nhưng hệ số tương quan riêng giữa Y với từng biến giải thích tương đối thấp. Thì điều đó có thể gợi ý rằng các biến X_2, X_3 có tương quan cao và có ít nhất một trong các biến này là thừa (mô hình có đa cộng tuyến).

5.3. Phát hiện Đa cộng tuyến

5.3.4. Hồi quy phụ

- Hồi quy phụ là phương pháp hồi quy một biến giải thích theo các biến giải thích còn lại.

Xét mô hình hồi quy k biến:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i, k > 2$$

Thủ tục kiểm định:

- Bước 1: Hồi quy mô hình hồi quy phụ có dạng:

$$X_{ji} = \lambda_1 + \lambda_2 X_{2i} + \dots + \lambda_{j-1} X_{j-1i} + \lambda_{j+1} X_{j+1i} + \dots + \lambda_k X_{ki} + V_i \Rightarrow R_j^2 (\forall j = 2 \div k)$$

5.3. Phát hiện Đa cộng tuyến

- Bước 2: Kiểm định cặp giả thuyết:

H_0 : X_j không có quan hệ tuyến tính với các biến còn lại

H_1 : X_j có quan hệ tuyến tính với các biến còn lại

Tiêu chuẩn kiểm định:

$$F_j = \frac{R_j^2 / ((k - 1) - 1)}{(1 - R_j^2) / (n - (k - 1))} \sim F(k - 2, n - k + 1)$$

Miền bác bỏ với mức ý nghĩa α cho trước:

$$W_\alpha = \{F_j : F_j > F(k - 2, n - k + 1)\}$$

5.3. Phát hiện Đa cộng tuyến

5.3.5. Nhân tử phóng đại phương sai

Dựa vào kết quả hồi quy phụ, ta có thể tính toán thêm một chỉ số để nhận biết vấn đề đa cộng tuyến đó là nhân tử phóng đại phương sai (Variance inflating factor – VIF).

- Bước 1: Hồi quy biến X_j với các biến độc lập còn lại tìm được R_j^2
- Bước 2: Xác định đại lượng: $VIF(X_j) = \frac{1}{1 - R_j^2} (j = 2 \div k)$
- Nếu $VIF(X_j) > 10$ thì X_j có thể cộng tuyến với các biến độc lập còn lại.

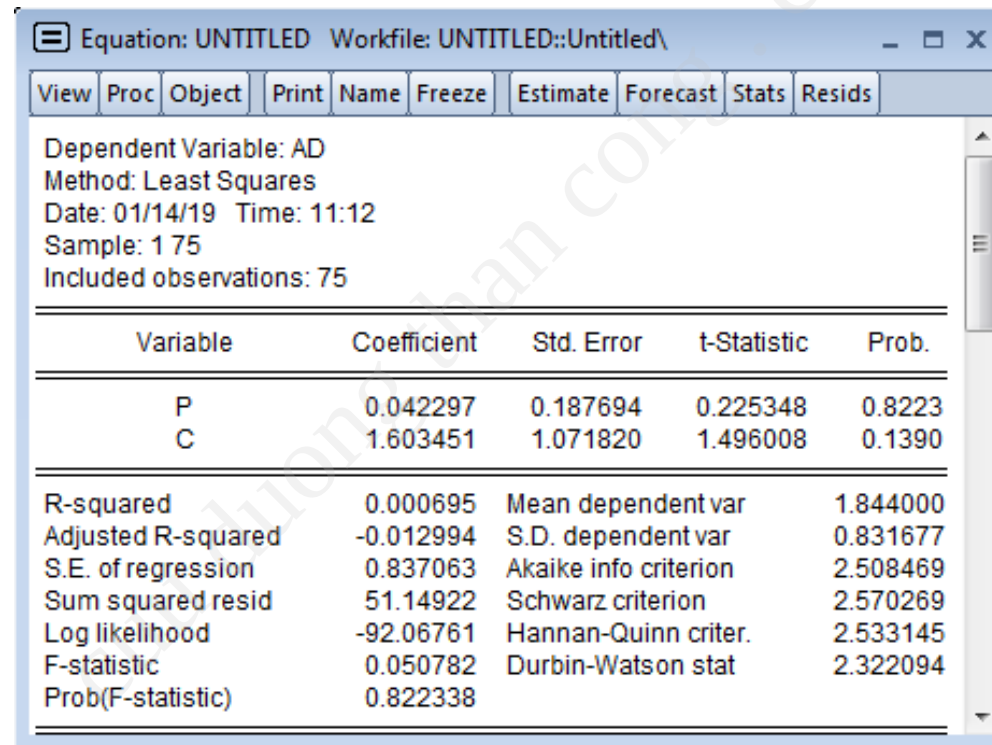
5.3. Phát hiện Đa cộng tuyến

- **Ví dụ 2:** Nghiên cứu mối quan hệ giữa doanh thu bán hàng - S (nghìn USD/tháng) theo giá bán của sản phẩm - P (USD/sản phẩm) và chi phí quảng cáo - AD (nghìn USD/tháng). Sử dụng số liệu của một cửa hàng kinh doanh bánh ngọt theo dõi doanh thu trong 75 tháng.
 - Mô hình hồi quy ban đầu:

$$DT_i = \beta_1 + \beta_2 AD_i + \beta_3 P_i + U_i$$

5.3. Phát hiện Đa cộng tuyến

- Sử dụng số liệu của các biến số trong mô hình ban đầu, hồi quy mô hình hồi quy phụ thu được:



Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: AD
Method: Least Squares
Date: 01/14/19 Time: 11:12
Sample: 1 75
Included observations: 75

Variable	Coefficient	Std. Error	t-Statistic	Prob.
P	0.042297	0.187694	0.225348	0.8223
C	1.603451	1.071820	1.496008	0.1390

R-squared	0.000695	Mean dependent var	1.844000
Adjusted R-squared	-0.012994	S.D. dependent var	0.831677
S.E. of regression	0.837063	Akaike info criterion	2.508469
Sum squared resid	51.14922	Schwarz criterion	2.570269
Log likelihood	-92.06761	Hannan-Quinn criter.	2.533145
F-statistic	0.050782	Durbin-Watson stat	2.322094
Prob(F-statistic)	0.822338		

- Dựa vào kết quả hồi quy, cho kết luận gì về mô hình ban đầu?

5.3. Phát hiện Đa cộng tuyến

5.3.6. Độ đo Theil

Xét mô hình hồi quy k biến:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + U_i, k > 2$$

- Bước 1: Hồi quy mô hình gốc thu được R^2
- Bước 2: Lần lượt hồi quy các mô hình sau:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_{j-1} X_{j-1i} + \beta_{j+1} X_{j+1i} + \dots + \beta_k X_{ki} + V_i \rightarrow R_{-j}^2 (\forall j = 2 \div k)$$

5.3. Phát hiện Đa cộng tuyến

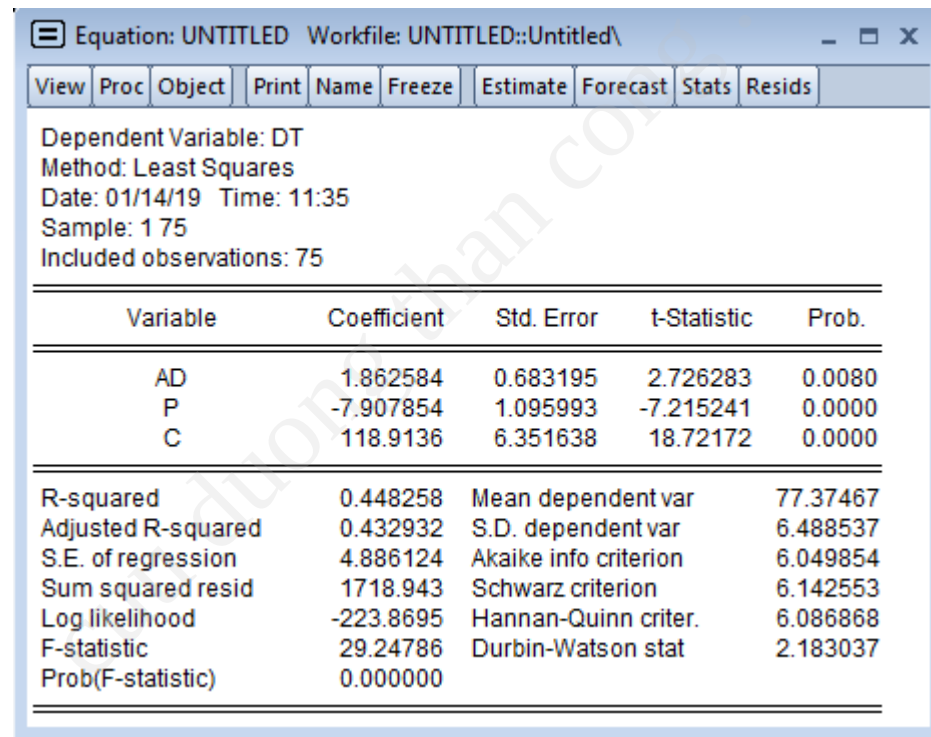
- Bước 3: Tìm độ đo Theil theo công thức sau:

$$m = R^2 - \sum_{j=2}^k (R^2 - R_{-j}^2)$$

- Bước 4: Kết luận
 - Nếu $m \approx 0$ thì mô hình không có đa cộng tuyến
 - Nếu $m \approx 1$ thì mô hình có đa cộng tuyến gần hoàn hảo
 - m càng cao thì mô hình có đa cộng tuyến càng cao.

5.3. Phát hiện Đa cộng tuyến

- **Ví dụ 3:** Sử dụng mối quan hệ của các biến trong ví dụ 2, cùng với bộ dữ liệu của các biến. Hồi quy mô hình ban đầu thu được:



Equation: UNTITLED Workfile: UNTITLED::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: DT
Method: Least Squares
Date: 01/14/19 Time: 11:35
Sample: 1 75
Included observations: 75

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AD	1.862584	0.683195	2.726283	0.0080
P	-7.907854	1.095993	-7.215241	0.0000
C	118.9136	6.351638	18.72172	0.0000

R-squared	0.448258	Mean dependent var	77.37467
Adjusted R-squared	0.432932	S.D. dependent var	6.488537
S.E. of regression	4.886124	Akaike info criterion	6.049854
Sum squared resid	1718.943	Schwarz criterion	6.142553
Log likelihood	-223.8695	Hannan-Quinn criter.	6.086868
F-statistic	29.24786	Durbin-Watson stat	2.183037
Prob(F-statistic)	0.000000		

5.3. Phát hiện Đa cộng tuyến

- Câu hỏi nghiên cứu:

Lần lượt hồi quy các mô hình:

$$DT_i = \alpha_1 + \alpha_2 AD_i + U_i \Rightarrow R_1^2 = 0.0493$$

$$DT_i = \lambda_1 + \lambda_2 P_i + U_i \Rightarrow R_2^2 = 0.391$$

Kết quả hồi quy cho kết luận gì về mô hình ban đầu?

5.4. Các biện pháp khắc phục

5.4.1. Sử dụng thông tin tiên nghiệm

Các thông tin tiên nghiệm được khai thác từ kinh nghiệm thông qua quan sát thực tế hoặc sử dụng các kết luận của kinh tế học.

- Ví dụ 2: $Q_i = \beta_1 K_i^{\beta_2} L_i^{\beta_3} e^{U_i} \Leftrightarrow \ln Q_i = \ln \beta_1 + \beta_2 \ln K_i + \beta_3 \ln L_i + U_i$
- Nếu có cơ sở cho rằng hàm sản xuất này có hiệu suất không đổi theo quy mô tức là $\beta_2 + \beta_3 = 1$. Khi đó biến đổi mô hình ban đầu thành:

$$Q_i = \beta_1 K_i^{\beta_2} L_i^{1-\beta_2} e^{U_i} \Leftrightarrow \frac{Q_i}{L_i} = \beta_1 \left(\frac{K_i}{L_i}\right)^{\beta_2} e^{U_i} \Leftrightarrow \ln\left(\frac{Q_i}{L_i}\right) = \ln \beta_1 + \beta_2 \ln\left(\frac{K_i}{L_i}\right) + U_i$$

- Mô hình này có thể khắc phục vấn đề đa cộng tuyến.

5.4. Các biện pháp khắc phục

5.4.2. Thu thập thêm số liệu mới

- Nếu đa cộng tuyến do đặc trưng của mẫu thì khi chọn mẫu khác liên quan đến các biến trong mẫu ban đầu mức độ đa cộng tuyến có thể không nghiêm trọng nữa. Phương án này có thể sử dụng khi chi phí cho việc lấy mẫu khác ở mức độ chấp nhận được.
- Đôi khi chỉ cần thu thập thêm số liệu, tăng cỡ mẫu có thể làm giảm tính nghiêm trọng của đa cộng tuyến.

5.4. Các biện pháp khắc phục

5.4.3. Bỏ biến

- Khi mô hình có hiện tượng đa cộng tuyến nghiêm trọng thì cách “đơn giản nhất” là bỏ biến ra khỏi mô hình.
- Có 2 cách để chọn biến loại khỏi mô hình:
 - *Cách 1*: Loại khỏi mô hình biến có tỷ số t thấp nhất
 - *Cách 2*: Lần lượt bỏ từng biến, hồi quy mô hình và chọn mô hình có hệ số R^2 cao nhất.
- Hai cách này cho kết quả như nhau.

5.4. Các biện pháp khắc phục

5.4.4. Sử dụng sai phân cấp 1

Xét mô hình 3 biến với số liệu theo thời gian: $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + U_t$

- Tại thời điểm (t-1) ta có mô hình: $Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + U_{t-1}$

- Trừ vế với vế của hai mô hình trên, khi đó ta có:

$$Y_t - Y_{t-1} = \beta_2 (X_{2t} - X_{2t-1}) + \beta_3 (X_{3t} - X_{3t-1}) + (U_t - U_{t-1})$$

$$\Rightarrow Y_t^* = \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + V_t$$

- Mô hình này gọi là mô hình sai phân cấp 1.
- Thực tế cho thấy mô hình sai phân cấp 1 đã giảm được đáng kể mức độ đa cộng tuyến của mô hình gốc.

5.4. Các biện pháp khắc phục

- Hạn chế của phương pháp sai phân cấp 1:
 - Chỉ dùng được với những số liệu theo thời gian
 - Không ước lượng được hệ số chặn
 - Mất đi một quan sát đầu tiên
 - U_t có thể thoả mãn mọi giả thiết của OLS, nhưng V_t thì có thể vi phạm.

5.4. Các biện pháp khắc phục

5.4.5. Các biện pháp khác

- Hồi quy thành phần chính.
- Sử dụng các ước lượng từ bên ngoài.