

Chương 5: KIỂM ĐỊNH VÀ LỰA CHỌN MÔ HÌNH

1. KÌ VỌNG CỦA SAI SỐ NGẪU NHIÊN KHÁC KHÔNG
2. PHƯƠNG SAI SAI SỐ THAY ĐỔI
3. SAI SỐ NGẪU NHIÊN KHÔNG TUÂN THEO QUY LUẬT PHÂN PHỐI CHUẨN
4. VẤN ĐỀ ĐA CỘNG TUYẾN
5. MÔ HÌNH CHỨA BIẾN KHÔNG THÍCH HỢP

5.1 KỲ VỌNG CỦA SAI SỐ NGẪU NHIÊN KHÁC KHÔNG

5.1.1 Nguyên nhân.

GT2: Kỳ vọng có điều kiện của sai số ngẫu nhiên bằng 0 tại mọi giá trị của biến độc lập:

$$E(U|X_{2i}, \dots, X_{ki}) = 0 \text{ với mọi } i.$$

GT 2 thỏa kéo theo :

$$\left\{ \begin{array}{ll} E(U) = 0 & \text{(i)} \\ \text{và } \text{Cov}(X_j, U) = 0. \text{ với mọi } j & \text{(ii)} \end{array} \right.$$

Vì thế nếu (i) hoặc (ii) không thỏa thì GT 2 bị vi phạm.

5.1 KỲ VỌNG CỦA SAI SỐ NGẪU NHIÊN KHÁC KHÔNG

5.1.1 Nguyên nhân.

**Một số
nguyên
nhân**

- 1.** Mô hình thiếu biến quan trọng
- 2.** Dạng hàm hồi quy sai
- 3.** Tính tác động đồng thời của số liệu
- 4.** Sai số đo lường các biến độc lập

5.1 KỲ VỌNG CỦA SAI SỐ NGẪU NHIÊN KHÁC KHÔNG

1) Mô hình thiếu biến quan trọng.

Xét mô hình $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$ (*)

Giả sử Z là một biến số nào đó không thuộc danh sách các biến X_j ($j = 2, \dots, k$). Mô hình (*) được gọi là **thiếu biến quan trọng Z** nếu:

- i. Biến Z có tác động đến biến phụ thuộc Y .
- ii. Và biến Z có tương quan với một biến độc lập X_{j_0} nào đó trong mô hình (*).



Z là một thành phần của U và $\text{Cov}(U, X_{j_0}) \neq 0$.

2) Dạng hàm sai

- Giả sử kỳ vọng của biến phụ thuộc Y không phải là hàm số tuyến tính của các biến số X_j . Tuy nhiên, ta lại ước lượng mô hình hồi quy kỳ vọng Y tuyến tính theo các biến số.

Ví dụ:

Giả sử $E(Y|X) = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 (X_3)^2$

Tuy nhiên, ta lại thực hiện hồi quy: $E(Y|X) = \beta_1 + \beta_2 X_2 + \beta_3 X_3$

5.1.2 Hậu quả.

Hậu quả của việc vi phạm GT2	Các ước lượng OLS bị chệch.
	Các suy diễn thống kê không còn đáng tin cậy

5.1.3 Cách phát hiện

- a) Kiểm định mô hình bỏ sót biến quan trọng (đã biết).
 - b) Kiểm định mô hình có dạng hàm sai
- Sử dụng kiểm định Ramsey Reset

* **Kiểm định Ramsey** (1969) dùng để phát hiện dạng hàm sai – trường hợp mô hình thiếu biến là hàm dạng đa thức của các biến có sẵn trong mô hình.

B1: Ước lượng mô hình hồi quy:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u \quad (1)$$

Thu được giá trị ước lượng $\hat{Y}_i$.

B2: Ước lượng mô hình hồi quy:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \beta_{k+1} (\hat{Y}_i)^2 + \beta_{k+2} (\hat{Y}_i)^3 + v \quad (2)$$

B3: Kiểm định cặp giả thuyết:

$$H_0: \beta_{k+1} = \beta_{k+2} = 0; H_1: \beta_{k+1}^2 + \beta_{k+2}^2 \neq 0$$

Nếu bác bỏ H_0 thì mô hình (1) có dạng hàm sai.

5.1.4 Khắc phục vấn đề kỳ vọng của sai số khác không

TH1: Nguyên nhân là thiếu biến Z (đã biết) thì ...

TH2: Nguyên nhân là dạng hàm sai, được phát hiện từ kiểm định Ramsey thì ...

TH3: Nguyên nhân là thiếu biến không quan sát được thì có thể:

- + Sử dụng biến đại diện (proxy variable)
- + Sử dụng biến công cụ (instrumental variable).

Ví dụ : Xét mô hình :

$$NS = \beta_1 + \beta_2 KN + \beta_3 HV + \beta_4 NL + u$$

- Trong đó: NL: năng lực bẩm sinh cá nhân.
- NL không thể đo lường được và thường có tương quan

dương với trình độ học vấn.

- Thường chọn chỉ số IQ để **đại diện** cho năng lực bẩm sinh cá nhân.

5.2 PHƯƠNG SAI SAI SỐ THAY ĐỔI (Heteroscedasticity)

5.2.1 Nguyên nhân PSSSTĐ

GT3: *Phương sai có điều kiện của sai số ngẫu nhiên bằng nhau tại mọi giá trị của biến độc lập:*

$$\text{Var}(U|X_{2i}, \dots, X_{ki}) = \sigma^2 \text{ với mọi } i.$$

- Nếu GT2 và GT3 thỏa thì $\text{Var}(Y|X_{2i}, \dots, X_{ki}) = \sigma^2$ với mọi i .
- GT3 vi phạm: $\text{Var}(U|X_{2i}, \dots, X_{ki}) = \sigma_i^2$

Nghĩa là phương sai của U phụ thuộc vào giá trị của biến độc lập.

5.2 PHƯƠNG SAI SAI SỐ THAY ĐỔI (Heteroscedasticity)

5.2.1 Nguyên nhân PSSSTĐ

-
1. Do bản chất số liệu.
 2. Do việc thu thập dữ liệu chưa chuẩn xác.
 3. Dạng hàm hồi quy sai.
-
4. Do mô hình thiếu biến quan trọng.

5.2.2 Hậu quả:

Các ước lượng OLS vẫn là ước lượng không chệch.

- Phương sai của hệ số ước lượng là chệch.
- Khoảng tin cậy và kiểm định giả thuyết thống kê về các hệ số không có giá trị.
- Các hệ số ước lượng không còn tốt nhất.

5.2.3 Phát hiện :

Giả sử các GT1, 2, 4, 5 thỏa mãn, GT 3 bị vi phạm.

Khi đó:

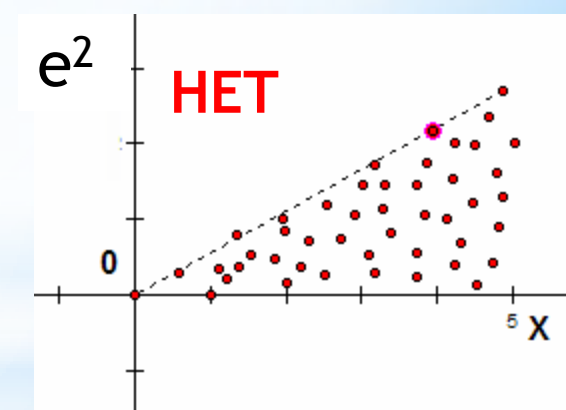
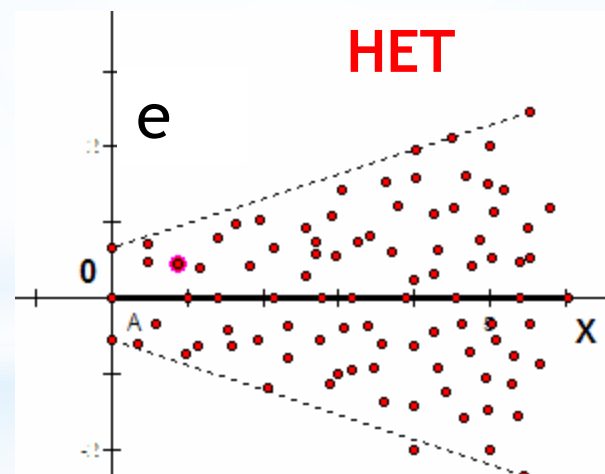
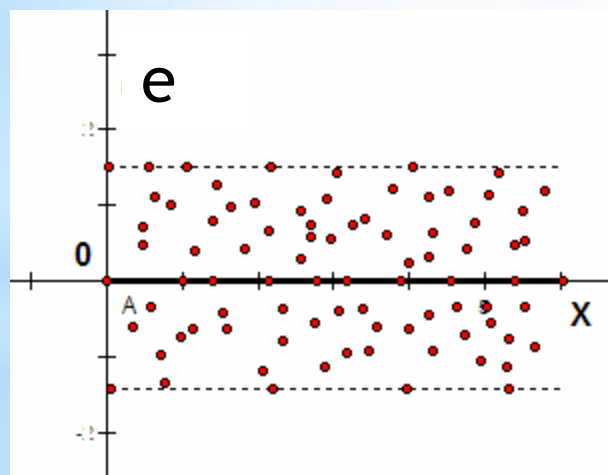
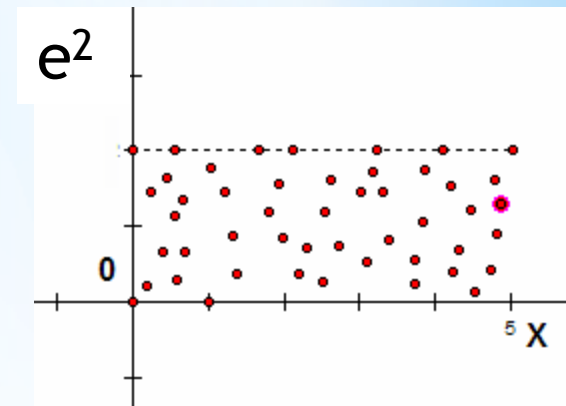
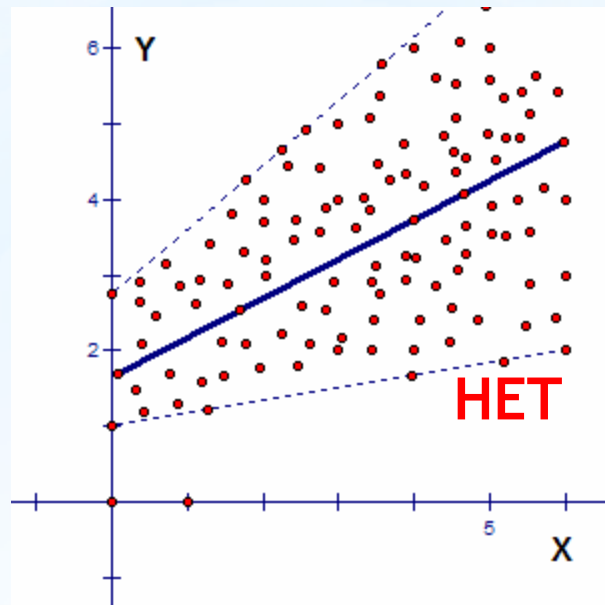
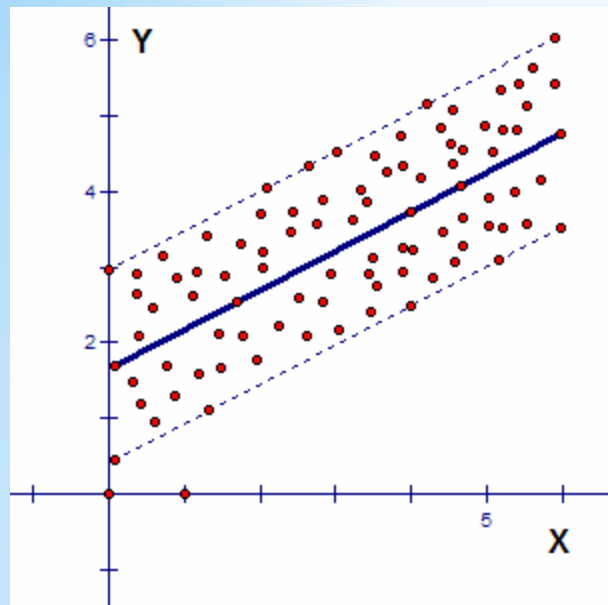
$$\begin{aligned} \text{Var}(U | X_{2i}, \dots, X_{ki}) &= E(U^2 | X_{2i}, \dots, X_{ki}) - [E(U | X_{2i}, \dots, X_{ki})]^2 \\ &= E(U^2 | X_{2i}, \dots, X_{ki}) = \sigma_i^2 \end{aligned}$$

GT3 vi phạm $\leftrightarrow$ Kỳ vọng của bình phương U phụ thuộc vào giá trị của biến độc lập.

5.2.3 Phát hiện :

***a)** Sử dụng đồ thị

- Vẽ đồ thị phần dư e_i hoặc e_i^2 theo một biến X_j nào đó.
- Hoặc vẽ đồ thị phân tán giữa (X_i, Y_i) .
- Có thể thay thế X_i bằng $\hat{Y}_i$ với mô hình đa biến.



5.2.3 Phát hiện :

b) Thực hiện kiểm định

B1: Thực hiện hồi quy $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + U$ (*) thu được các phần dư e_i .

B2: Thực hiện hồi quy phụ theo một trong những cách sau:

i) *Breusch-Pagan(1979)*: $e_i^2 = b_1 + b_2 X_{2i} + \dots + b_k X_{ki} + w_i$

ii) *Park (1986)*: $\ln(e_i^2) = b_1 + b_2 \ln(X_{2i}) + \dots + b_k \ln(X_{ki}) + w_i$

iii) *White*: $e_i^2 = b_1 + \sum_{j=1}^k b_j X_{ji} + \sum_{j=1}^k b_{k+j} X_{ji}^2 + \sum c_{jh} X_{ji} X_{hi} + w_i$

hoặc $e_i^2 = b_1 + b_2 \hat{Y}_j + b_3 \hat{Y}_j^2 + w_i$

Thu được hệ số xác định R_e^2

5.2.3 Phát hiện :

b) Thực hiện kiểm định

B3: Thực hiện kiểm định cặp giả thuyết:

H_0 : tất cả các hệ số hồi quy b_j (trừ b_1) = 0.

H_1 : ít nhất một trong các hệ số b_j khác 0.

Có thể sử dụng kiểm định F hoặc kiểm định LM (Lagrange Multiplier Test) để cho kết luận.

- **Kiểm định LM:**

Nếu $LM = n.R_e^2 > \chi^2_{\alpha}(m)$ thì bác bỏ H_0 , với m là số ràng buộc trong giả thuyết H_0 .

5.2.4 Khắc phục vấn đề phương sai sai số thay đổi

TH1: Nguyên nhân là thiếu biến Z (đã biết) thì ...

TH2: Nguyên nhân là dạng hàm sai, thì ...

TH3: Phương pháp bình phương bé nhất tổng quát
(GLS – Generalized Least Squares)

Giả sử đã biết dạng thay đổi của phương sai sai số.

Ví dụ: $\sigma_i^2 = \sigma^2 X_{2i}^2$

Khi đó, chia hai vế của phương trình hồi quy (*) cho X_{2i} .

$$\frac{Y_i}{X_{2i}} = \frac{\beta_1}{X_{2i}} + \beta_2 + \beta_3 \frac{X_{3i}}{X_{2i}} + \dots + \beta_k \frac{X_{ki}}{X_{2i}} + w_i \quad (**)$$

Mô hình hồi quy (**) không có hiện tượng phương sai thay đổi.

TH4: Phương pháp sai số chuẩn vững (robust standard error).

5.3 SAI SỐ NGẪU NHIÊN KHÔNG TUÂN THEO QUY LUẬT PHÂN PHỐI CHUẨN

5.3.1 Hậu quả

- Khi sai số ngẫu nhiên không tuân theo quy luật phân phối chuẩn thì các thống kê t và F không tuân theo quy luật Student và Fisher (tương ứng).
- Khi kích thước mẫu nhỏ thì các suy diễn thống kê không đáng tin cậy.
- Khi kích thước mẫu lớn thì các suy diễn thống kê vẫn có giá trị.

5.3.2 Phát hiện

- i) Dùng đồ thị phần dư
- ii) Kiểm định Jacque – Bera (1987)

H_0 : u tuân theo luật phân phối chuẩn.

H_1 : u không tuân theo luật phân phối chuẩn

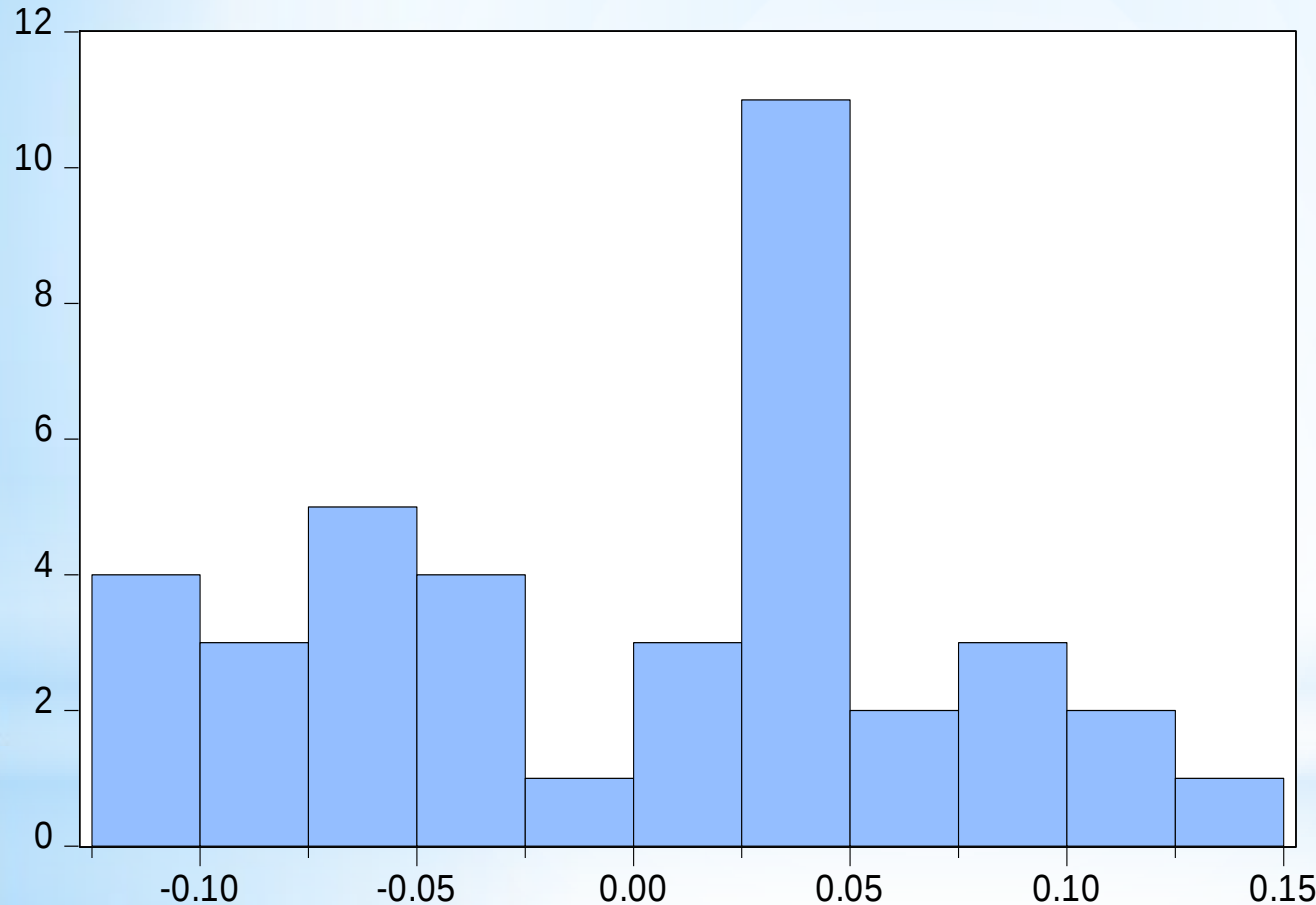
$$JB = n \left(\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right)$$

Với S là độ bất đối xứng (Skewness), K là độ nhọn (Kurtosis)

Nếu $JB > \chi^2_{\alpha}(2)$ thì bác bỏ giả thuyết H_0 .

H_0 : u tuân theo luật phân phối chuẩn.

H_1 : u không tuân theo luật phân phối chuẩn.



Series: Residuals

Sample 1 39

Observations 39

Mean -7.47e-15

Median 0.020419

Maximum 0.128223

Minimum -0.108989

Std. Dev. 0.068728

Skewness -0.068771

Kurtosis 1.802354

Jarque-Bera 2.361568

Probability 0.307038

5.4 VẤN ĐỀ ĐA CỘNG TUYẾN (multicollinearity)

5.4.1 Khái niệm đa cộng tuyến trong mô hình hồi quy

- Mô hình hồi quy với các biến độc lập $X_2, X_3, \dots, X_k$ được gọi là có hiện tượng **đa cộng tuyến hoàn hảo** khi tồn tại các hằng số $\lambda_2, \dots, \lambda_k$ không đồng thời bằng 0 sao cho

$$\lambda_2 X_2 + \dots + \lambda_k X_k = 0. \text{ (theo Ragnar Frisch, 1934)}$$

- Ngày nay, hiện tượng đa cộng tuyến được hiểu các biến độc lập có tương quan với nhau theo nghĩa:

$$\lambda_2 X_{i2} + \dots + \lambda_k X_{ik} + V_i = 0 \quad (V_i \text{ là sai số ngẫu nhiên})$$

5.4.2 Nguyên nhân và hậu quả

Nguyên nhân

Do bản chất mối quan hệ giữa các biến.

Mô hình có dạng đa thức.

Việc chọn mẫu không mang tính đại diện

Hậu quả

**TH đa cộng tuyến hoàn hảo:* Các hệ số hồi quy là không xác định bằng phương pháp OLS.

** TH đa cộng tuyến cao:* Khoảng tin cậy của các β_j lớn, ước lượng khoảng kém chính xác.

Bài toán kiểm định giả thuyết thống kê cho kết luận sai.

Một sự thay đổi dù bé trong mẫu có thể gây ra sự thay đổi lớn trong kết quả ước lượng.

Xét phương pháp ước lượng OLS

Khi các GT1 – 4 thỏa mãn thì:

$$\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{1 - R_j^2 \sum_{i=1}^n x_{ji}^2} ;$$

$$\text{se}(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}$$

Trong đó:

- R_j^2 là hệ số xác định của mô hình hồi quy biến X_j theo hệ số chặn và các biến độc lập còn lại.
- σ^2 chưa biết, có ước lượng (không chệch) là $\hat{\sigma}^2$, với:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n - k}$$

5.4.3 Phát hiện đa cộng tuyến:

Tồn
tại

Điều
kiện
đủ

Điều
kiện
cần
và đủ

$$VIF(X_j) = 1 / (1 - R_j^2)$$

5.4.4 Khắc phục vấn đề đa cộng tuyến

Nếu đa cộng tuyến cao nhưng **không** gây ra **hậu quả nghiêm trọng** thì **không cần** đến các biện pháp khắc phục.

Ví dụ:

- 1) Có đa cộng tuyến cao nhưng các $se(\hat{\beta}_j)$ không quá lớn so với $\hat{\beta}_j$ (giá trị $|t| > 2$).
- 2) Có đa cộng tuyến cao nhưng $se(\hat{\beta}_{j0})$ của biến X_{j0} mà ta quan tâm nghiên cứu không lớn.
- 3) Mô hình hồi quy nhằm phục vụ việc dự báo chứ không phải kiểm định.
- 4) R^2 của mô hình cao hơn R_j^2 của mô hình hồi quy phụ.

5.2.4 Khắc phục vấn đề đa cộng tuyến

Nếu đa cộng tuyến cao **gây ra hậu quả nghiêm trọng**, một số biện pháp có thể làm giảm nhẹ mức độ ảnh hưởng.

- 1) Gia tăng kích thước mẫu.
- 2) Sử dụng thông tin tiên nghiệm hoặc thông tin từ các nguồn khác.

Ví dụ:

- 3) Bỏ bớt biến độc lập.

Ví dụ

Theo các chuyên gia kinh tế, khi nền kinh tế của một quốc gia có tỷ lệ lạm phát cao thì lãi suất trên thị trường tiền tệ cũng có xu hướng cao.

a) Mô hình nghiên cứu giá cổ phiếu dưới tác động của các nhân tố lãi suất, tỷ giá hối đoái, cung tiền, lạm phát tại Việt Nam giai đoạn 2010 – 2011 (giai đoạn lạm phát cao) có thể gặp vấn đề gì? Vấn đề này có thể gây ra hậu quả gì?

b) Có thể sử dụng mô hình trên để nghiên cứu tác động của tỷ giá hối đoái đến giá cổ phiếu tại Việt Nam trong giai đoạn 2010 – 2011 hay không?

5.5 MÔ HÌNH CHỨA BIẾN KHÔNG THÍCH HỢP (irrelevant variable)

5.5.1 Hậu quả

Biến không thích hợp là biến không có tác động riêng phần lên biến phụ thuộc.

Nếu mô hình hồi quy chứa biến không thích hợp thì

Các ước lượng vẫn không chệch.

Phương sai các hệ số ước lượng lớn.

→ Khoảng tin cậy cho các hệ số hồi quy rất rộng.

→ Các tỉ số t bé, bài toán kiểm định cho kết luận không đúng.

5.5.2 Phát hiện biến không thích hợp

- + Sử dụng kiểm định t để phát hiện sự phù hợp của 1 biến.
- + Sử dụng kiểm định F để phát hiện sự phù hợp của nhiều biến (≥ 2).
- + Khi thấy các biến không thích hợp thì loại các biến ra khỏi mô hình.