

ÔN THI KINH TẾ LƯỢNG 2014 – 2015

Buổi 1. Hồi quy OLS

A. Số liệu trong Kinh tế lượng:

- Số liệu chuỗi thời gian: Số liệu của 1 đơn vị kinh tế theo chuỗi thời gian: ngày, tháng, quý, năm ...
- Số liệu chéo: Số liệu của nhiều đơn vị kinh tế tại 1 thời điểm.
- Số liệu hỗn hợp (số liệu bảng): Số liệu của nhiều đơn vị kinh tế theo chuỗi thời gian.

Số liệu chuỗi thời gian Số liệu chéo

Chi tiêu C và thu nhập Y của anh A từ 2000 đến 2002

Năm	Y	C
2000	5	4
2001	6.5	5
2002	7	5.5

Chi tiêu C và thu nhập Y của A, B, C năm 2000

Người	Y	C
A	5	4
B	6.5	5
C	7	5.5

Số liệu hỗn hợp

Chi tiêu C và thu nhập Y của A, B, C từ 2000 - 2002

	Y			C		
	Y-2000	Y-2001	Y-2002	C-2000	C-2001	C-2002
A	5	6.5	7	4	5	5.5
B	4.5	5	6.5	4	4	4.5
C	5	6	6.5	3.5	5	5

B. Mô hình hồi quy tuyến tính:

Mô hình (hàm) hồi quy tổng thể (PRF):

$$E(Y|X_{2i}, \dots, X_{ki}) = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

$$Y_i = E(Y|X_{2i}, \dots, X_{ki}) + u_i$$

Mô hình (hàm) hồi quy mẫu (SRF):

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

$$Y_i = \hat{Y}_i + e_i$$

Sai số ngẫu nhiên và phần dư: đại diện cho những yếu tố không đưa được vào mô hình.

Sai số ngẫu nhiên (Nhiều)

$$u_i = Y_i - E(Y|X_i)$$

Phần dư (là ước lượng của sai số)

$$e_i = Y_i - \hat{Y}_i$$

Ý nghĩa của hệ số hồi quy:

- β_1 : Trong MH hồi quy tuyến tính, khi tất cả $X_i = 0$ thì giá trị trung bình của biến phụ thuộc là β_1 .
- β_2 :

Dạng hàm	Công thức	Ý nghĩa hệ số góc
Tuyến tính	$Y = \beta_1 + \beta_2 X + u_i$	X tăng 1 đơn vị thì Y trung bình tăng β_2 đơn vị
Log	$\ln(Y) = \beta_1 + \beta_2 \ln(X) + u_i$	X tăng 1% thì Y trung bình tăng $\beta_2\%$
Log - Lin	$\ln(Y) = \beta_1 + \beta_2 X + u_i$	X tăng 1 đơn vị thì Y trung bình tăng $100 \cdot \beta_2\%$
Lin - Log	$Y = \beta_1 + \beta_2 \ln(X) + u_i$	X tăng 1% thì Y trung bình tăng $\beta_2/100$ đơn vị

Lưu ý:

- Mô hình hồi quy tuyến tính là mô hình có các hệ số hồi quy ở dạng bậc nhất (so với dấu "=").
- Mô hình có n quan sát, k hệ số hồi quy thì có số bậc tự do là: $n - k$.
- Khi đơn vị X thay đổi làm giá trị của X tăng m lần, đơn vị Y không đổi, thì hệ số góc β tương ứng với X giảm m lần (hệ số chặn không đổi).
- Nếu mô hình có nhiều biến độc lập, khi nêu ý nghĩa hệ số của biến độc lập này thì phải cố định các biến độc lập còn lại (các biến còn lại không đổi).

Bài luyện tập 1: Mô hình nào sau đây là mô hình hồi quy tuyến tính:

- | | |
|--|--|
| a) $Y = \beta_1 + \frac{\beta_2}{X} + u$ | b) $Y = \frac{X}{\beta_1 + \beta_2 X} + u$ |
| c) $Y^2 = \beta_1 + \beta_2 \sqrt{\ln(X)} + u$ | d) $Y = e^{\beta_1 + \beta_2 X + u}$ |

Bài luyện tập 2: Cho (SRF): $Y = 5 - 0.5X_1 + \ln(X_2) + e_i$. Biết $n = 150$.

- a) Ý nghĩa của hệ số biến X_1 là gì?

- b) Biết Y là lượng cầu cơm rang (đĩa), X_1 là giá 1 đĩa cơm rang (nghìn đồng/đĩa), X_2 là giá 1 bát bún (nghìn đồng/bát). Hỏi khi giá 1 bát bún tăng 2% thì lượng cầu cơm rang thay đổi như thế nào?
- c) Thay đổi đơn vị giá cơm rang từ nghìn đồng/đĩa sang triệu đồng/đĩa. Nêu ý nghĩa hệ số của X_1 ?

C. Phương pháp bình phương nhỏ nhất OLS

Giả thiết OLS:

- X_i đã biết (phi ngẫu nhiên).
- Kỳ vọng u_i bằng 0: $E(u_i) = 0 \forall i$
- Phương sai u_i không đổi: $var(u_i) = \sigma^2 \forall i \neq j$
- Không có sự tương quan giữa các u_i : $cov(u_i, u_j) = 0 \forall i \neq j$
- u_i và X_i không tương quan: $cov(u_i, X_j) = 0 \forall i$
- Nhiễu phân phối chuẩn: $u_i \sim N(0, \sigma^2)$
- Không có đa cộng tuyến hoàn hảo (đối với mô hình hồi quy bội).

Ước lượng OLS:

Ước lượng càng chính xác khi giá trị ước lượng \hat{Y}_i càng gần với giá trị thực Y_i . Phương pháp OLS ước lượng các β sao cho $\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2$ nhỏ nhất. Công thức hệ số ước lượng của mô hình hồi quy đơn:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad \hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{cov(X_i, Y_i)}{var(X_i)}$$

với $x_i = X_i - \bar{X}$ và $y_i = Y_i - \bar{Y}$

Định lý Gauss – Markov: Khi thỏa mãn các giả thiết OLS, các ước lượng OLS được gọi là BLUE (best linear unbiased estimator) hay có tính chất: tuyến tính, không chệch, tốt nhất (có phương sai nhỏ nhất trong tất cả các ước lượng tuyến tính không chệch).

Tính không chệch: $E(\hat{\beta}_i) = \beta_i$. Với mỗi mẫu (cùng số quan sát) được chọn từ tổng thể, chúng ta ước lượng được một bộ các hệ số $\hat{\beta}$ khác nhau và duy nhất. Tính không chệch nghĩa là trung bình các $\hat{\beta}$ bằng β thực của tổng thể ($E(\hat{\beta}_1) = \beta_1; E(\hat{\beta}_2) = \beta_2 \dots$).

Các tính chất của hồi quy mẫu:

- Đường hồi quy mẫu đi qua điểm $(\bar{Y}, \bar{X}_2, \bar{X}_3 \dots \bar{X}_k)$: $\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 + \dots + \hat{\beta}_k \bar{X}_k$
- Trung bình Y thực tế bằng trung bình Y ước lượng: $\bar{Y} = \bar{\hat{Y}}$.
- Tổng phần dư bằng 0: $\sum e_i = 0$
- Phần dư e_i không tương quan với \hat{Y} và X : $\sum e_i \hat{Y}_i = \sum e_i X_i = 0$.

Độ chính xác của hệ số ước lượng: Đánh giá bằng phương sai và sai số tiêu chuẩn.

Sai số tiêu chuẩn trong mô hình hồi quy đơn:

$$se(\hat{\beta}_1) = \hat{\sigma} \sqrt{\frac{\sum X_i^2}{n \sum x_i^2}} \quad se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum x_i^2}} \quad \hat{\sigma}^2 = \frac{\sum e_i^2}{n - 2}$$

Phương sai: $var(\hat{\beta}) = se^2(\hat{\beta})$. Với $\hat{\sigma}^2$ là ước lượng phương sai của u_i (σ^2).

Khoảng tin cậy của hệ số hồi quy β với độ tin cậy $1 - \alpha$:

$$\begin{array}{ll} \text{Hai phía} & (\hat{\beta} - se(\hat{\beta}) \cdot t_{\alpha/2}(n-k); \hat{\beta} + se(\hat{\beta}) \cdot t_{\alpha/2}(n-k)) \\ \text{Phía trái} & (-\infty; \hat{\beta} + se(\hat{\beta}) \cdot t_{\alpha}(n-k)) \\ \text{Phía phải} & (\hat{\beta} - se(\hat{\beta}) \cdot t_{\alpha}(n-k); +\infty) \end{array}$$

Của 2 hệ số hồi quy:

$$(\hat{\beta}^* - se(\hat{\beta}^*) \cdot t_{\alpha/2}(n-k); \hat{\beta}^* + se(\hat{\beta}^*) \cdot t_{\alpha/2}(n-k))$$

Trong đó:

- $\hat{\beta}^* = \hat{\beta}_i \pm \hat{\beta}_j$.
- $t_{\alpha/2}(n-k)$ là giá trị tới hạn Student, mức ý nghĩa $\alpha/2$ bậc $(n-k)$.
- $t_{\alpha}(n-k)$ là giá trị tới hạn Student, mức ý nghĩa α bậc $(n-k)$.

$$-se(\hat{\beta}^*) = \sqrt{var(\hat{\beta}_i \pm \hat{\beta}_j)} = \sqrt{var(\hat{\beta}_i) + var(\hat{\beta}_j) \pm 2cov(\hat{\beta}_i, \hat{\beta}_j)} \text{ với } var(\hat{\beta}) = se^2(\hat{\beta}).$$

Bài luyện tập 3: Cho hồi quy: $Y_i = aX_i + b + e_i$ với mẫu 10 quan sát thu được kết quả:

$$\sum Y_i = 120, \sum X_i = 230, \sum x_i y_i = 2600, \sum x_i^2 = 5300, \sum y_i^2 = 1500 \quad (x_i = X_i - \bar{X}; y_i = Y_i - \bar{Y})$$

Các ước lượng OLS thu được có giá trị:

A. $a = 0.49, b = 7.17$

C. $a = 0.49, b = 0.717$

B. $a = 7.17, b = 0.49$

D. $a = 0.717, b = 0.49$

Bài luyện tập 4: Cho hồi quy:

$$Y_i = 1.2 + X_{2i} - X_{3i} + e_i$$

$$se(\hat{\beta}_i): 2.2 - 0.3 - 1.4$$

$$n = 30; \alpha = 0.05$$

Tìm khoảng tin cậy 2 phía của β_2 . Khi X_{2i} tăng 1 đơn vị (X_{3i} không đổi) thì Y_i thay đổi trong khoảng nào? Tăng tối đa là bao nhiêu? Biết $t_{0.025}(27) = 2.052; t_{0.05}(27) = 1.703$.

D. Kiểm định giả thuyết về các hệ số hồi quy:

Loại kiểm định	Giả thuyết H_0	Giả thuyết H_1	Miền bác bỏ
Hai phía	$\beta = a$	$\beta \neq a$	$ t > t_{\alpha/2}(n - k)$
Phía trái	$\beta \geq a$	$\beta < a$	$t < -t_{\alpha}(n - k)$
Phía phải	$\beta \leq a$	$\beta > a$	$t > t_{\alpha}(n - k)$

– Dùng thống kê t: Tính giá trị quan sát: $t_{qs} = \frac{\hat{\beta} - a}{se(\hat{\beta})}$. Nếu t_{qs} thuộc miền bác bỏ thì bác bỏ H_0 .

– Dùng khoảng tin cậy: Nếu a không nằm trong khoảng tin cậy thì bác bỏ H_0 .

– Dùng giá trị P – value:

+ Kiểm định hai phía: Nếu P – value $< \alpha$ thì bác bỏ H_0 .

+ Kiểm định một phía: Nếu P – value $< 2\alpha$ thì bác bỏ H_0 .

Ngược lại thì chưa có cơ sở bác bỏ giả thuyết H_0 .

Trường hợp 2 hệ số: Đặt $\beta^* = \beta_i \pm \beta_j$. Coi biểu thức giữa 2 hệ số là 1 hệ số mới. Quy tắc bác bỏ tương tự.

Bài luyện tập 5: (tiếp Bài luyện tập 4)

a) Hệ số biến X_{2i} có ý nghĩa thống kê không?

b) Khi X_{2i} không đổi, có thể cho rằng khi X_{3i} tăng 1 đơn vị thì Y_i giảm ít hơn 1.2 đơn vị không?

E. Hệ số xác định R^2 :

– R^2 cho biết các biến độc lập X_i giải thích được bao nhiêu % sự biến động của biến phụ thuộc Y_i .

($1 \geq R^2 \geq 0$) (R^2 có thể âm khi mô hình không có hệ số chặn)

– Các tổng bình phương độ lệch: $TSS = ESS + RSS$ với:

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum y_i^2$$

$$ESS = \sum (\hat{Y}_i - \bar{Y})^2$$

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum e_i^2$$

Giá trị của R^2 :

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

Hệ số xác định đã hiệu chỉnh \bar{R}^2 :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k}$$

Tính chất:

– $1 \geq R^2 \geq \bar{R}^2, \bar{R}^2$ có thể âm.

– Dùng $\overline{R^2}$ để xem xét việc đưa biến mới vào mô hình: Nếu việc thêm biến mới làm $\overline{R^2}$ tăng, hệ số hồi quy của biến được đưa vào có ý nghĩa thì nên đưa biến đó vào mô hình.

Bài luyện tập 6:

OLS, using observations 1-20

Dependent variable: Q

	Coefficient	Std. Error	t-ratio	p-value
const	10.31076	2.586328	3.986638	0.0004
P	- 0.389687	0.118694	3.283121	0.0028
Px	5.33333	3.365153	1.58487	0.0553
Mean dependent var	18.45833		S.D. dependent var	4.613
Sum squared resid	...		S.E. of regression	3.989267
R-squared	...		Adjusted R-squared	...

a) Viết mô hình tổng thể và mẫu.

b) Tính TSS, ESS, RSS, R^2 , $\overline{R^2}$.

(Mean dependent var = \bar{Y} : Trung bình biến phụ thuộc.

Sum squared resid = RSS: Tổng bình phương phần dư.

S.D. dependent var = $\sqrt{\frac{TSS}{n-1}}$: Sai số tiêu chuẩn của biến phụ thuộc.

S.E. of regression = $\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{RSS}{n-k}}$: Sai số tiêu chuẩn của hàm hồi quy.)

c) Hệ số xác định của hàm hồi quy có giá trị bằng bao nhiêu? Ý nghĩa của nó là gì?

F. Kiểm định sự phù hợp của mô hình:

Xét mô hình $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u$

– Cặp giả thuyết: $\begin{cases} H_0: R^2 = 0 \text{ (mô hình không phù hợp)} \\ H_1: R^2 > 0 \text{ (} R^2 \neq 0 \text{) (mô hình phù hợp)} \end{cases}$

Hoặc: $\begin{cases} H_0: \beta_2 = \beta_3 = \dots = \beta_k = 0 \\ H_1: \beta_2^2 + \beta_3^2 + \dots + \beta_k^2 > 0 \text{ (tồn tại ít nhất một hệ số khác 0)} \end{cases}$

– Giá trị quan sát:

$$F_{qs} = \frac{R^2}{1 - R^2} \cdot \frac{n - k}{k - 1}$$

Nếu $F_{qs} > F_{\alpha}^{(k-1, n-k)}$ thì bác bỏ H_0 . Mô hình hồi quy là phù hợp, các biến độc lập giải thích được sự biến động của biến phụ thuộc.

Kiểm định thu hẹp hồi quy (kiểm định nhiều ràng buộc về các hệ số hồi quy):

– Cặp giả thuyết: $\begin{cases} H_0: \text{Hệ số các biến muốn bỏ bằng 0} \\ H_1: \text{tồn tại ít nhất một hệ số khác 0} \end{cases}$

– Giá trị quan sát:

$$F_{qs} = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - k}{m} = \frac{RSS_2 - RSS_1}{RSS_1} \cdot \frac{n - k}{m}$$

Trong đó:

+ n là số quan sát, k là số tham số hồi quy mô hình trước khi bỏ biến, m là số biến muốn bỏ khỏi mô hình.

+ R_1^2, R_2^2 lần lượt là hệ số xác định của mô hình trước và sau khi bỏ biến.

+ RSS_1, RSS_2 lần lượt là tổng bình phương phần dư của mô hình trước và sau khi bỏ biến.

Nếu $F_{qs} > F_{\alpha}^{(m, n-k)}$ thì bác bỏ $H_0 \Rightarrow$ Không bỏ biến khỏi mô hình.

Các kiểm định trên đều có thể dùng p – value để đánh giá. Nếu p – value < mức ý nghĩa α thì bác bỏ H_0 .

Bài luyện tập 7: (tiếp Bài luyện tập 6)

- a) Mô hình có phù hợp không? Biết $F(2, 17) = 3.59$ ở mức ý nghĩa 5%.
 b) Bỏ biến Px, hồi quy lại mô hình thu được $R - squared = 0.156343$. Có nên bỏ biến Px không? Biết $F(1, 17) = 4.48$ ở mức ý nghĩa 5%.

G. Dự báo:

1. Dự báo (Ước lượng) điểm giá trị trung bình khi $X = X_0$:

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$$

2. Dự báo giá trị trung bình khi $X = X_0$:

$$\hat{Y}_0 - se(\hat{Y}_0) \cdot t_{\alpha/2}(n - k) \leq E(Y|X_0) \leq \hat{Y}_0 + se(\hat{Y}_0) \cdot t_{\alpha/2}(n - k)$$

3. Dự báo giá trị cá biệt khi $X = X_0$:

$$\hat{Y}_0 - se(Y_0) \cdot t_{\alpha/2}(n - k) \leq Y_0 \leq \hat{Y}_0 + se(Y_0) \cdot t_{\alpha/2}(n - k)$$

Với:

$$se(\hat{Y}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}; \quad se(Y_0) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}}; \quad x_i = X_i - \bar{X}$$

Một số thuật ngữ trong bảng kết quả của các phần mềm Kinh Tế Lượng

- | | |
|--|---|
| • Model SS | ESS |
| • Residual SS; Sum squared resid | RSS |
| • Total SS | TSS |
| • Number of obs; observations | Số quan sát |
| • Dependent variable | Biến phụ thuộc |
| • Sample 1 12 | Mẫu sử dụng các quan sát từ 1 đến 12 |
| • F(3,20) | Fqs trong kiểm định sự phù hợp của mô hình |
| • Prob > F; P-value(F); Prob (F-statistic) | P-value trong kiểm định sự phù hợp của mô hình |
| • R-squared | R^2 – hệ số xác định |
| • Adj R-squared; Adjusted R-squared | \bar{R}^2 – hệ số xác định đã hiệu chỉnh |
| • Cột Coef.; Coefficient | Hệ số ước lượng (các beta mũ) |
| • Cột Std. Err; Std. Error | Sai số tiêu chuẩn của hệ số ước lượng |
| • Cột t; t-ratio; t-Statistic | Giá trị quan sát t khi kiểm định ý nghĩa của beta |
| • Cột P> t ; p-value; Prob. | P – value trong của kiểm định ý nghĩa của beta |
| • 95% Coef. Interval | Khoảng tin cậy 95% của hệ số hồi quy |
| • _cons; const; C | Các thông số của hệ số chặn |
| • Mean dependent var | \bar{Y} – giá trị trung bình biến phụ thuộc |
| | Sai số (Độ lệch) tiêu chuẩn của biến phụ thuộc: |

$$\sqrt{\frac{TSS}{n - 1}}$$

Sai số (Độ lệch) tiêu chuẩn của hàm hồi quy:

$$\hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n - k}} = \sqrt{\frac{RSS}{n - k}}$$

BÀI TẬP BUỔI I
Bài Tập Trắc Nghiệm

1. Trong ước lượng OLS :

- A. R^2 nhỏ nhất
- B. RSS nhỏ nhất
- C. ESS nhỏ nhất
- D. Cả 3 ý trên

2. Cho phương trình hồi quy tuyến tính : $\ln(Y) = 0,214.\ln(X_1) + 0,112.X_2 + u_i$. Y là doanh thu (trăm triệu \$), X_1 là số giờ lao động (giờ), X_2 là chi phí cho quảng cáo (triệu \$), trả lời cho câu 2, 3.

Ý nghĩa hệ số của X_1 :

- A. Trong điều kiện các yếu tố khác không thay đổi, tăng giờ lao động 1 đơn vị thì doanh thu tăng 0,214%
- B. Trong điều kiện các yếu tố khác không thay đổi, tăng giờ lao động 1% thì doanh thu tăng 0,214%
- C. Trong điều kiện các yếu tố khác không thay đổi, tăng giờ lao động 1% thì doanh thu giảm 0,214%
- D. Tăng giờ lao động 1% thì doanh thu tăng 0,214%

3. Ý nghĩa hệ số của X_2 :

- A. Tăng chi phí quảng cáo 1 triệu \$ thì doanh thu tăng 0.112 triệu \$
- B. Tăng chi phí quảng cáo 1 triệu \$ thì doanh thu tăng 11.2 triệu \$
- C. Không đáp án nào đúng

4. Các ước lượng OLS có tính chất là:

- A. Không chệch
- B. Tuyến tính, không chệch
- C. Tuyến tính, không chệch, tốt nhất
- D. Tuyến tính, hiệu quả

5. Hệ số góc ước lượng của mô hình hồi quy đơn bằng 0, khi đó :

- A. $0 < R^2 < 1$
- B. $R^2 = \bar{Y}$
- C. $R^2 = 0$
- D. $R^2 > ESS/TSS$

6. Trong các đại lượng sau, đại lượng nào luôn nhận giá trị không âm?

- A. T
- B. F
- C. $cov(X, Y)$
- D. \bar{R}^2

7. Phần dư là:

- A. Chênh lệch giữa giá trị thực và giá trị dự đoán
- B. Chênh lệch giữa giá trị thực và giá trị trung bình
- C. Chênh lệch giữa giá trị trung bình và giá trị dự đoán
- D. Không có đáp án nào đúng

8. Sử dụng phương pháp OLS với dữ liệu thu thập được, ta nhận được mô hình sau

$$Y_i = 1,45.X_{1i} - 5,025.X_{2i} + 12,82.X_{3i} + u_i \quad i=1, 2, \dots, 43.$$

Số bậc tự do của mô hình trên bằng:

- A. 38
- B. 39
- C. 40
- D. Không có đáp án nào đúng

9. Cho mô hình sau: $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$ $i=1,2,\dots,n$. Biết rằng $var(X) = 2 cov(X,Y)$, nếu mô hình ước lượng bằng phương pháp OLS thì ước lượng của $\hat{\beta}_2$ bằng:

- A. 2
- B. 0,5
- C. 0,2
- D. 0

10. Cho mô hình sau: $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X$. Hồi quy mô hình bằng phương pháp OLS; $i=1,2,\dots,n$

Ước lượng của $\hat{\beta}_2$ là ước lượng không chệch, tức là:

- A. $E(\hat{\beta}_2) = \beta_2$
- B. $\hat{\beta}_2$ có tính chất BLUE
- C. $E(\beta_2) = \hat{\beta}_2$
- D. $E(\beta_2) = 0$

11. Cho dữ liệu sau:

X	4	5	7	9	12
Y	6	9	12	25	30

Sử dụng phương pháp OLS với mô hình $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$, ta thu được các ước lượng như sau:

- A. $\hat{\beta}_0 = -7,16505; \hat{\beta}_1 = 3,18447$
- B. $\hat{\beta}_0 = 3,18447; \hat{\beta}_1 = -7,16505$
- C. $\hat{\beta}_0 = 2,52312; \hat{\beta}_1 = 0,297371$
- D. $\hat{\beta}_0 = 0,297371; \hat{\beta}_1 = 2,52312$

12. Với dữ kiện của câu 11, RSS, TSS và ESS lần lượt là?

- A. 2,184950; 441,2; 417,8
- B. 21,84950; 441,2; 23,398
- C. 23,39806; 441,2; 41,78
- D. 23,39806; 441,2; 417,8

13. Với dữ kiện của câu 12, $R^2 = ?$

- A. 0,853415
- B. 0,946967
- C. 0,929290
- D. Không đáp án nào đúng

14. Mức ý nghĩa giảm, độ rộng khoảng tin cậy sẽ :

- A. Tăng lên
- B. Giảm đi
- C. Không đổi
- D. Các trường hợp trên đều đúng

15. Cho các phương trình sau :

- (1) $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$
 (2) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + u_i$
 (3) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

Phương trình nào trong các phương trình trên được biểu diễn đúng :

- A. (1)
 B. (2)
 C. (3)
 D. (1) và (3)

16. Giả sử có kiểm định 2 phía không đối xứng sử dụng t-test với mức ý nghĩa α và $\alpha_1 + \alpha_2 = \alpha$, α_1 sử dụng cho đuôi bên trái, α_2 sử dụng cho đuôi bên phải cho phân phối Student. Giả thuyết $H_0: b=0$, $H_1: b \neq 0$ khi t quan sát nhận giá trị âm có khả năng bị bác bỏ lớn hơn khi t quan sát nhận giá trị dương, lúc này so sánh giá trị tới hạn của t:

- A. $|t(a_1)| < |t(a_2)|$
 B. $|t(a_1)| = |t(a_2)|$
 C. $|t(a_1)| > |t(a_2)|$
 D. Chưa thể kết luận gì về các giá trị tới hạn

17. Khẳng định nào sau đây SAI đối với việc dùng p-value vào kiểm định

- A. P-values chỉ được sử dụng đối với kiểm định 2 phía
 B. P-value là mức độ ý nghĩa chính xác của test
 C. P-value là mức độ ý nghĩa cận biên – nơi mà chúng ta sẽ phân vân giữa bác bỏ và không bác bỏ giả thuyết H_0 (null hypothesis)
 D. Nếu biết được p-value, ta có thể suy diễn thống kê mà không cần tra các bảng thống kê (statistical tables)

18. Trong mô hình hồi quy $\log(q_i) = 2,25 - 0,7 \cdot \log(p_i) + 0,02 \cdot y_i$ với p là giá và q là lượng cầu của 1 hàng hóa nào đó, y là thu nhập sẵn có. Ý nghĩa hệ số số 0,7 là gì?

- A. Nếu giá tăng 1% thì lượng cầu sẽ thấp hơn 0,007% về mặt trung bình trong khi các biến khác giữ nguyên
 B. Nếu giá tăng 1% thì lượng cầu sẽ thấp hơn 70% về mặt trung bình trong khi các biến khác giữ nguyên
 C. Nếu giá tăng 1% thì lượng cầu sẽ thấp hơn 0,7% về mặt trung bình trong khi các biến khác giữ nguyên
 D. Cả a b c đều sai

19. Khi mà $R^2 = 1$ thì khẳng định nào sau đây đúng ?

- (i) Tất cả các điểm dữ liệu nằm ở trên 1 đường thẳng
 (ii) Tất cả phần dư bằng 0
 (iii) Tất cả độ biến động của y xung quanh giá trị trung bình được giải thích bởi mô hình
 (iv) Đường dự đoán (fitted line) nằm ngang đối với tất cả biến giải thích

- A. (ii) và (iv)
 B. (i) và (iii)
 C. (i), (ii) và (iii)
 D. Tất cả

20. Cho pt hồi quy $Y_i = aX_i + b + u_i$ với mẫu 10 quan sát thu được kết quả :

$$\sum Y_i = 99855,75, \sum X_i = 112800$$

$$\sum x_i y_i = 50104729, \sum x_i^2 = 64156000$$

$$\sum y_i^2 = 39296085,66$$

Các ước lượng OLS thu được có giá trị :

- A. $a = 1,275, b = -4396,425$
 B. $a = 1,275, b = -43964,25$
 C. $a = 0,781, b = 11758,95$
 D. $a = 0,781, b = 1175,9$

21. Phương sai sai số ngẫu nhiên càng bé khi:

- A. Kích thước mẫu càng bé
 B. Kích thước mẫu càng lớn
 C. Quan hệ giữa các biến độc lập càng lớn
 D. Không đáp án nào đúng

22. Mô hình $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{\beta}_4 X_{4i} + u_i$ có $R_1^2 = 0,881$. Xem xét bỏ biến X_2, X_3 ra khỏi mô hình. Mô hình $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_4 X_{4i} + u_i$ có $R_2^2 = 0,875$. Giả thuyết H_0 cho kiểm định trên là :

- A. $\beta_2 = \beta_3 = 0$
 B. $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
 C. $R^2 = 0$
 D. Không có đáp án đúng.

23. Với mô hình ở câu trên, số quan sát là 25, thống kê F dùng để bỏ biến tính dc là :

- A. 0,168
 B. 0,504
 C. 0,555
 D. 1,223

24. Mức ý nghĩa 5%, giá trị tới hạn

$F(2, 20) = 3,493$ kết luận là :

- A. Có thể bỏ.
 B. Không thể bỏ.
 C. Chưa kết luận được.

25. Mức ý nghĩa là 5%. P – value nhỏ đi (nhỏ hơn 1%) chỉ ra dẫn chứng :

- A. Bác bỏ giả thuyết không.
 B. Giá trị thống kê t là nhỏ hơn 2.58
 C. Chấp nhận giả thuyết không
 D. Tất cả các câu trên đều sai.

B. Bài Tập Với Bảng

Bảng STATA:

Source	SS	df	MS	Number of obs = 24		
Model	41574.703	3	13858.2343	F(3, 20)	= 112.89	
Residual	(1)	20	122.756515	Prob > F	= 0.0000	
Total	44029.8333	23	1914.34058	R-squared	= (2)	
				Adj R-squared	= (3)	
				Root MSE	= 11.08	

Q	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
L	.8227443	(4)	4.31	0.000	.4246169	1.220872
K	(5)	.1959274	1.59	0.127	-.0964928	.7209018
Z	-.0002492	.0003105	-0.80	0.432	(6)	(7)
_cons	-10.77742	16.41637	-0.66	0.519	-45.02137	23.46653

Bảng GRETL:

Model 1: OLS, using observations 1951-1980 (T = 30)

Dependent variable: C

	Coefficient	Std. Error	t-ratio	p-value	
const	25,6728	14,805	1,7341	0,09522	*
I	0,120114	0,137018	0,8766	0,38903	
G	0,0310287	(1)	2,3031	0,02987	**
H	(2)	0,00453969	-2,0059	0,05580	*
A	-0,225253	0,427166	-0,5273	0,60262	

Mean dependent var	45,63967	S.D. dependent var	21,74352
Sum squared resid	(3)	S.E. of regression	5,944055
R-squared	0,935576	Adjusted R-squared	(4)
F(4, 25)	(5)	P-value(F)	1,65e-14
Log-likelihood	-93,30508	Akaike criterion	196,6102
Schwarz criterion	203,6161	Hannan-Quinn	198,8514
rho	0,381501	Durbin-Watson	1,229532

Bảng EIEWS:

Dependent Variable: Q
 Method: Least Squares
 Date: 10/05/14 Time: 12:44
 Sample: 1 12
 Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24463.28	5619.777	(1)	0.0018
DC	84.21205	58.81760	(2)	0.1860
P	(3)	5.698364	-0.566235	0.5851

R-squared	0.969258	Mean dependent var	113916.7
Adjusted R-squared	(4)	S.D. dependent var	32329.15
S.E. of regression	(5)	Akaike info criterion	20.53620
Sum squared resid	3.53E+08	Schwarz criterion	20.65743
Log likelihood	-120.2172	Hannan-Quinn criter.	20.49132
F-statistic	141.8776	Durbin-Watson stat	2.099322
Prob(F-statistic)	0.000000		

B Bài Tập Tự Luận

Bài 1: Hồi quy mô hình với các biến Y, X_1, X_2, X_3 ta được:

$$\hat{Y} = 139 + 0.412 \cdot X_1 + 0.015 \cdot X_2 - 0.083 \cdot X_3$$

s.e. (0,33) (0,094) (0,0011) (0,026)

$n = 141; R^2 = 0.234.$

- Tìm khoảng tin cậy 95% cho hệ số X_1 biết $t_{0.025}(137) = 1.978.$
- Có thể bác bỏ (1) $H_0: \beta_{X_1} = 0.4;$ (2) $H_0: \beta_{X_1} = 1$ trong kiểm định 2 phía ở mức ý nghĩa 5% không?
- Các yếu tố ngoài X_1, X_2, X_3 giải thích được bao nhiêu sự biến động của Y ?
- Hồi quy Y với các biến độc lập $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8$

Kết quả:

$$\hat{Y} = 0.9793273 + 0.4451077 \cdot X_1 + 0.010549 \cdot X_2 - 0.0844848 \cdot X_3 + 0.0166968 \cdot X_4$$

p – value (0.146) (0.000) (0.000) (0.002) (0.485)

$$+ 0.0422996 \cdot X_5 + 0.1514983 \cdot X_6 - 0.0203601 \cdot X_7 + 0.0858214 \cdot X_8$$

(0.496) (0.458) (0.770) (0.150)

R-squared = 0.2532

F – statistic (8.132) = 5.60

P – value for F() = 0.0000

Hệ số ước lượng của biến nào nào không có ý nghĩa? Hãy kiểm định việc bỏ các biến này ra khỏi mô hình. Cho $F(5, 132) = 2.28$ ở mức ý nghĩa 5%.

- Tính hệ số xác định đã hiệu chỉnh của mỗi mô hình.

Bài 2:

Trong các mô hình sau, dấu của các hệ số như thế nào thì phù hợp với lý thuyết kinh tế?

- $CT = \beta_1 + \beta_2 \cdot TN + u.$ CT là chi tiêu, TN là thu nhập.
- $Q = \beta_1 + \beta_2 \cdot P + u.$ Q là lượng bán, P là giá bán HH bình thường. Nếu P là HH bổ sung hoặc thay thế thì sao?
- $TC = \beta_1 + \beta_2 \cdot Q + u.$ TC là tổng chi phí, Q là sản lượng.
- $M = \beta_1 + \beta_2 \cdot R + u.$ M là cầu tiền mặt, R là lãi suất.

Bài 3:

Cho số liệu của 20 đại lí bán hàng, P là giá bán thịt hộp (nghìn đồng/hộp), Q là lượng bán (hộp). Kết quả ước lượng mô hình Q phụ thuộc P:

Model 1: OLS, using observations 1 – 20

Dependent variable: Q

	Coefficient	Std. Error	T – ratio	P – value
const	486.7856	73.02756	6.665779	0,0000
P	- 10.44716	2.857461	- 3.656100	0.0018
Mean dependent var	221.9500	S.E. of regression		41.46273
Sum squared resid	30944.85	F(1, 18)		13.36706
R – squared	0.426150	P – value(F)		0.001807

- Viết phương trình hồi quy tổng thể, mẫu.
- Dấu của các hệ số có phù hợp với lý thuyết không?
- Các yếu tố khác với giá bán giải thích được bao nhiêu % sự biến động của lượng bán thịt hộp?
- Ước lượng điểm lượng bán khi giá 1 hộp là 34 nghìn đồng.
- Ước lượng lượng bán trung bình và cá biệt khi $P = 34.$

Bài 4: (tiếp Bài 3)

Thêm biến PC là giá sản phẩm thay thế vào mô hình (nghìn đồng/ hộp)

Kết quả ước lượng mô hình Q phụ thuộc P, PC:

Model 1: OLS, using observations 1 – 20

Dependent variable: Q

	Coefficient	Std. Error	T – ratio	P – value
const	302.7606	14.52326	20.84660	0,0000
P	- 14.76790	0.517617	- 28.53056	0.0000

PC	11.62597	0.473371	24.55997	0.0000
Mean dependent var	221.9500	Adjusted R-squared		0.982420
Sum squared resid	848.2248	F(2, 17)		531.8780
R – squared	0.984270	P – value(F)		0.0000

- Hàm hồi quy phù hợp không? Các biến độc lập có giải thích được cho lượng bán không?
- Khi P tăng 1 nghìn đồng, PC không đổi thì lượng bán dao động trong khoảng nào? Lượng bán thay đổi tối đa bao nhiêu?
- Kiểm định giả thuyết cho rằng khi P giữ nguyên, PC tăng 1 nghìn thì Q tăng hơn 10 hộp.
- Kiểm định giả thuyết cho rằng khi PC giữ nguyên, P tăng 1 nghìn thì Q tăng 10 hộp.
- Kết hợp với bảng hồi quy ở bài tập 2, kiểm định việc bỏ biến PC ra khỏi mô hình

Bài 5: Cho Y: lượng tiêu thụ thịt gà (đơn vị: kg/năm); X2: thu nhập khả dụng (đơn vị: VND). X3 là giá bán lẻ thịt gà, X4 là giá bán lẻ thịt lợn, X5 là giá bán lẻ thịt bò (VND/kg). Mức ý nghĩa 5%. Bảng kết quả:

Dependent Variable: Y; Included observation: 20

Variable	Coefficient	Std. Err	t – statistic	Prob
X2	0.372	0.083		0.0007
X3	- 0.2931		- 4.55	
X4	0.128	0.099		
X5	0.091		0.904	
C	2.189		14.062	0.0000
Sum squared resid	21.137	S.D. dependent var		
Durbin – Watson stat	0.89	F – statistic	89.928 [.000]	

- Viết phương trình hồi quy tổng thể, mẫu. Các biến độc lập giải thích được bao nhiêu% sự thay đổi của biến phụ thuộc?
- Biến nào không có ý nghĩa thống kê?
- Khi giá thịt gà tăng 1 đơn vị (các yếu tố khác không đổi), lượng tiêu thụ thịt gà tăng tối đa bao nhiêu?
- Khi thu nhập tăng 1 đơn vị (các yếu tố khác không đổi), lượng thịt gà tăng ít hơn 0.3 đơn vị đúng không?
- Có ý kiến cho rằng tác động của thu nhập lên cầu thịt gà mạnh hơn tác động của giá thịt gà. Nêu cặp giả thuyết cho kiểm định này.

ĐÁP ÁN BUỔI I

A Trắc Nghiệm

1. B. RSS nhỏ nhất

$RSS = \sum e_i^2$. PP OLS tìm tham số hồi quy sao cho $\sum e_i^2$ nhỏ nhất, hay RSS nhỏ nhất.

2. B. Trong điều kiện các yếu tố khác không đổi, tăng giờ lao động 1% thì doanh thu tăng 0,214%

Hàm có dạng tuyến tính Log: $\ln Y = a + b \ln X$. Khi X tăng 1% thì Y tăng b%.

3. C. Không đáp án nào đúng

Ý nghĩa đúng là: Tăng chi phí quảng cáo 1 triệu thì doanh thu tăng $0,112 \cdot 100\% = 11,2\%$.

4. C. Tuyến tính, không chệch, tốt nhất

Định lý Gauss – Markov: ƯL OLS là tuyến tính, không chệch, tốt nhất trong các ƯL không chệch của tham số.

5. C. $R^2 = 0$

Hệ số góc bằng 0, khi đó biến độc lập không có ý nghĩa thống kê, không giải thích được cho biến phụ thuộc. Vậy $R^2 = 0$.

6. B. F

$T = \hat{\beta}_i / se(\hat{\beta}_i)$, $cov(X, Y) = E(XY) - E(X)E(Y)$, \bar{R}^2 có thể âm. $F = \frac{R^2}{1-R^2} \frac{n-k}{k-1}$ luôn không âm.

7. A. Chênh lệch giữa giá trị thực và giá trị dự đoán

Phần dư $e_i = Y_i - \hat{Y}_i$

8. C. 40

Số bậc tự do trong mô hình có n quan sát, k tham số hồi quy là $(n - k) = 43 - 3 = 40$.

9. B. 0,5

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{cov(X, Y)}{var(X)} = \frac{1}{2} c$$

10. A. $E(\hat{\beta}_2) = \beta_2$

Ước lượng không chệch là ước lượng thỏa mãn: Kỳ vọng của ước lượng bằng giá trị thực của tham số cần ước lượng.

11. A. $\hat{\beta}_0 = -7,16505$, $\hat{\beta}_1 = 3,18447$

12. D. 23,39806

13. B. 0,946967

	$x_i = X_i - \bar{X}$	$y_i = Y_i - \bar{Y}$	$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_1 + \hat{\beta}_2 X_i)$			
X	4	5	7	9	12	$\bar{X} = 7,4$
Y	6	9	12	25	30	$\bar{Y} = 16,4$
x_i	-3,4	-2,4	-0,4	1,6	4,6	
y_i	-10,4	-7,4	-4,4	8,6	13,6	
$x_i y_i$	35,36	17,76	1,76	13,76	62,56	$\sum x_i y_i = 131,2$
x_i^2	11,56	5,67	0,16	2,56	21,16	$\sum x_i^2 = 41,2$
y_i^2	108,16	54,76	19,36	73,96	184,96	$\sum y_i^2 = 441,2$
e_i	0,42717	0,2427	-3,12624	3,50482	-1,04859	

$$11. \hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{131,2}{41,2} \approx 3,18447.$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} = 16,4 - 3,18447 \cdot 7,4 \approx -7,16505$$

$$12. RSS = \sum e_i^2 \approx 23,39806. TSS = \sum y_i^2 = 441,2. ESS = TSS - RSS = 417,8.$$

$$13. R^2 = \frac{ESS}{TSS} = \frac{417,8}{441,2} \approx 0,947$$

14. A. Tăng lên

Độ rộng khoảng tin cậy phụ thuộc vào $\hat{\beta}$; $se(\hat{\beta})$; t_{α} hoặc $\alpha/2(n - k)$. Khi mức ý nghĩa càng nhỏ thì t tới hạn càng lớn, do đó độ rộng khoảng tin cậy tăng khi mức ý nghĩa giảm.

15. C. (3)

(1) đúng là: $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$

(2) đúng là: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

16. A. $|t(a_1)| < |t(a_2)|$.

Giá trị âm dễ bị bác bỏ hơn chứng tỏ t quan sát âm dễ nằm trong vùng bác bỏ bên trái hơn so với giá trị t quan sát dương tương ứng. Hay miền bác bỏ bên trái lớn hơn miền bác bỏ bên phải. Suy ra t tới hạn bên trái gần mốc 0 hơn t tới hạn bên phải. Vậy $|t(a_1)| < |t(a_2)|$.

17. A. P – value chỉ được sử dụng với kiểm định 2 phía

P – value có thể dùng cho kiểm định 1 phía, nhưng so sánh với 2α thay vì α .

18. C. Nếu giá tăng 1% thì lượng cầu sẽ thấp hơn 0,7% về mặt TB trong khi các biến khác giữ nguyên.

Hàm tuyến tính Log: $\ln Y = a + b \cdot \ln X$.

19. C. (i), (ii) và (iii)

$$R^2 = 1 = 1 - RSS/TSS \Rightarrow RSS = 0 = \sum e_i^2 \Rightarrow e_i = 0 \forall i \text{ và } Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i.$$

\Rightarrow Các điểm dữ liệu nằm trên 1 đường thẳng.

Tất cả độ biến động của Y xung quanh giá trị TB được giải thích bởi mô hình.

(iv) sai vì đường dự đoán nằm ngang chứng tỏ X không tác động gì đến Y trong khi $R^2 = 1$ (X hoàn toàn giải thích được sự biến động của Y).

20. D. a = 0,781; b = 1175,9

$$a = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{50104729}{64156000} \approx 0,78098; b = \bar{Y} - a\bar{X} = \frac{99855,75}{10} - 0,78098 \cdot \frac{112800}{10} \approx 1176,1206$$

21. B. Kích thước mẫu càng lớn

$$\hat{\sigma}^2 = \sqrt{\frac{\sum e_i^2}{n-k}}. n \text{ tăng thì phương sai càng bé.}$$

22. A. $\beta_2 = \beta_3 = 0$

23. B. 0,504

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n-k}{m} = \frac{0,881 - 0,875}{1 - 0,881} \cdot \frac{25-5}{2} \approx 0,504$$

24. A. Có thể bỏ.

$F_{qs} < F(2, 20) \Rightarrow$ chưa có cơ sở để bác bỏ H_0 .

25. A. Bác bỏ giả thuyết không.

P – value nhỏ hơn mức ý nghĩa, bác bỏ giả thuyết không. Câu B sai vì thống kê t còn phụ thuộc vào số quan sát của mô hình.

B Bài Tập Với Bảng

Bảng STATA

$$(1) = RSS = TSS - MSS = 44029,8333 - 41574,703 = 2455,1303$$

(MSS = ESS)

$$(2) = R^2 = \frac{ESS}{TSS} = \frac{41574,703}{44029,8333} \approx 0,94424$$

$$(3) = \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,94424) \frac{24-1}{24-4} = 0,935876$$

$$(4) = se(\hat{\beta}_2) = \frac{\hat{\beta}_2}{t_{\hat{\beta}_2}} = \frac{0,8227443}{4,31} \approx 0,19089$$

$$(5) = \hat{\beta}_3 = t_{\hat{\beta}_3} \cdot se(\hat{\beta}_3) = 1,59 \cdot 0,1959274 \approx 0,31152$$

$$((6); (7)) = (\hat{\beta}_4 - t_{\alpha/2}(n-k) \cdot se(\hat{\beta}_4); \hat{\beta}_4 + t_{\alpha/2}(n-k) \cdot se(\hat{\beta}_4))$$

$$= (-0,0002492 - 2,086 \cdot 0,003105; -0,0002492 + 2,086 \cdot 0,003105)$$

$$\approx (-0,00673; 0,00623)$$

$$t_{\alpha/2}(n-k) = t_{0,025}(20) = 2,086.$$

Bảng GRETL

$$(1) = se(G) = \frac{\hat{\beta}_G}{t_G} = \frac{0,0310287}{2,3031} \approx 0,01347$$

$$(2) = \hat{\beta}_H = t_H \cdot se(H) = -2,0059 \cdot 0,00453969 \approx -0,009106$$

$$S. E. \text{ of regression} = \hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{RSS}{n-k}} = \sqrt{\frac{\text{sum squared resid}}{30-5}}$$

Suy ra: (3) = sum squared resid = S. E. of regression² . 25 ≈ 883.295

$$(4) = \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0,935576) \frac{30-1}{30-5} = 0,92568$$

$$(5) = F(4, 25) = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} = \frac{0,935576}{1-0,935576} \cdot \frac{30-5}{5-1} \approx 90,7635$$

Bảng EVIEWS:

$$(1) = t(\hat{\beta}_1) = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{24463.28}{5619.777} \approx 4.353$$

$$(2) = t(\hat{\beta}_2) = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{84.21205}{58.81760} \approx 1.432$$

$$(3) = \hat{\beta}_3 = t_3 \cdot se(\hat{\beta}_3) = -0.566235 \times 5.698364 \approx -3.2266$$

$$(4) = \bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} = 1 - (1 - 0.969258) \frac{12-1}{12-3} \approx 0.96243$$

$$(5) = \text{S. E. of regression} = \hat{\sigma} = \sqrt{\frac{\sum e_i^2}{n-k}} = \sqrt{\frac{RSS}{n-k}} = \sqrt{\frac{\text{sum squared resid}}{12-3}} \approx 6262.765$$

Việc tính toán giá trị có thể sai lệch với bảng kết quả, các giá trị này đã được máy tính làm tròn.

B Bài Tập Tự Luận

Bài 1:

$$n = 141.; R^2 = 0.234.$$

$$\hat{Y} = 139 + 0.412.X_1 + 0.015.X_2 - 0.083.X_3$$

a) Khoảng tin cậy 95% của β_{X_1} :

$$\left(\hat{\beta}_{X_1} - t_{\alpha/2}(n-k).se(\hat{\beta}_{X_1}); \hat{\beta}_{X_1} + t_{\alpha/2}(n-k).se(\hat{\beta}_{X_1}) \right)$$

$$= (0.412 - 1.978 \times 0.094; 0.412 + 1.978 \times 0.094)$$

$$= (0.226; 0.598)$$

$$t_{\alpha/2}(n-k) = t_{0.025}(137) = 1.987$$

b)

$$(1) \begin{cases} H_0: \beta_{X_1} = 0.4 \\ H_1: \beta_{X_1} \neq 0.4 \end{cases}$$

$$t_{qs} = \frac{\hat{\beta}_{X_1} - 0.4}{se(\hat{\beta}_{X_1})} = \frac{0.412 - 0.4}{0.094} = 0.12766.$$

Miền bác bỏ: $|t| > |t_{\alpha/2}(n-k)| = 1.978.$

$t_{qs} \notin$ Miền bác bỏ \Rightarrow Chưa có cơ sở để bác bỏ H_0 .

$$(2) \begin{cases} H_0: \beta_{X_1} = 1 \\ H_1: \beta_{X_1} \neq 1 \end{cases}$$

$$t_{qs} = \frac{\hat{\beta}_{X_1} - 1}{se(\hat{\beta}_{X_1})} = \frac{0.412 - 1}{0.094} = -6.25532.$$

Miền bác bỏ: $|t| > |t_{\alpha/2}(n-k)| = 1.978.$

$t_{qs} \in$ Miền bác bỏ \Rightarrow Bác bỏ H_0 .

Cách khác: Dùng khoảng ước lượng ở phần a)

c)

$R^2 = 0.234$ cho biết X_1, X_2, X_3 giải thích được 23.4% sự biến động của Y . Vậy các yếu tố ngoài 3 biến này giải thích được $100\% - 23.4\% = 76.6\%$ sự biến động của biến phụ thuộc.

d)

– Kiểm định ý nghĩa hệ số ước lượng: $\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases}$

Các biến mà hệ số ước lượng không có ý nghĩa là: X_4, X_5, X_6, X_7, X_8 . Vì P-value của các biến này > 0.05 , chưa đủ cơ sở bác bỏ H_0 .

– Kiểm định bỏ biến: $\begin{cases} H_0: \text{Tất cả các hệ số bằng 0} \\ H_1: \text{Có ít nhất 1 hệ số khác 0} \end{cases}$

$$F_{qs} = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n-k}{m} = \frac{0.2532 - 0.234}{1 - 0.2532} \cdot \frac{141-9}{5} \approx 0.6787. F_{qs} < F(5, 132) = 2.28$$

(R_1^2 là hệ số xác định của mô hình trước khi bỏ biến. R_2^2 là hệ số xác định của mô hình sau khi bỏ biến, n là số quan sát, k là số tham số hồi quy, m là số biến muốn bỏ khỏi mô hình)

\Rightarrow Chưa có cơ sở để bác bỏ H_0 . Có thể bỏ 5 biến trên khỏi mô hình.

e) Hệ số xác định đã hiệu chỉnh:

$$\text{Mô hình ban đầu: } \bar{R}_2^2 = 1 - (1 - R_2^2) \frac{n-1}{n-k_2} = 1 - (1 - 0.234) \cdot \frac{141-1}{141-4} \approx 0.2172$$

$$\text{Mô hình thêm biến: } \bar{R}_1^2 = 1 - (1 - R_1^2) \frac{n-1}{n-k_1} = 1 - (1 - 0.2532) \cdot \frac{141-1}{141-9} \approx 0.2079$$

Bài 2:

a) $\beta_1 > 0, \beta_2 > 0$: Lý thuyết tiêu dùng của Keynes.

b)

Hàng hóa thông thường	$P \uparrow \Rightarrow Q \downarrow$	$\beta_1 > 0, \beta_2 < 0$
Hàng hóa bổ sung	$P \uparrow \Rightarrow Q \downarrow$	$\beta_1 > 0, \beta_2 < 0$
Hàng hóa thay thế	$P \uparrow \Rightarrow Q \uparrow$	$\beta_1 \leq 0, \beta_2 > 0$

c) $Q \uparrow \Rightarrow TC \uparrow \Rightarrow \beta_2 > 0$.

$Q = 0 \Rightarrow TC = \beta_1 > 0$ (chi phí cố định) $\Rightarrow \beta_1 > 0$.

d) $R \uparrow \Rightarrow M \downarrow \Rightarrow \beta_2 < 0$.

$R = 0 \Rightarrow M = \beta_1 \geq 0$.

Bài 3:

$$n = 20; k = 2.$$

a) $Q = 486,7856 - 10,44716.P + e_i$ hoặc $\hat{Q} = 486,7856 - 10,44716.P$.

$\hat{\beta}_2 = -10,44716$: Khi P tăng 1 đơn vị thì Q giảm 10,44716 đơn vị.

b) Dấu của các hệ số phù hợp Lý thuyết về Cầu: Khi giá tăng thì cầu (ứng với lượng bán được) sẽ giảm. Khi giá bằng 0 (bán miễn phí) thì lượng bán được > 0.

c) R – squared = 0,42615. Các yếu tố khác giá bán giải thích được $100\% - 42,615\% = 57,385\%$ sự biến động của lượng bán.

d) $P = 34 \Rightarrow Q = 486,7856 - 10,44716 \cdot 34 = 131,58 \approx 132$ hộp.

e) Ước lượng giá trị trung bình và cá biệt khi $P_0 = 34$:

- Tính \hat{Q}_0 : $\hat{Q}_0 = \hat{\beta}_1 + \hat{\beta}_2 P_0 = 486,7856 - 10,44716 \cdot 34 = 131,58$

- Tính \bar{P} : Ta có: $\bar{Q} = \hat{\beta}_1 + \hat{\beta}_2 \bar{P} \Rightarrow \bar{P} = \frac{\bar{Q} - \hat{\beta}_1}{\hat{\beta}_2} = \frac{221,9500 - 486,7856}{-10,44716} = 25,35$

- Tính $\sum p_i^2$:

Cách 1: $p_i = P_i - \bar{P}$; $\hat{q}_i = \hat{Q}_i - \bar{Q} \Rightarrow \hat{q}_i = \hat{\beta}_2 \cdot p_i$

$$RSS = 30944,85; TSS = \frac{RSS}{(1 - R^2)} = \frac{30944,85}{(1 - 0,426150)} \Rightarrow ESS = TSS - RSS$$

Lại có: $ESS = \sum \hat{q}_i^2 = \hat{\beta}_2^2 \sum p_i^2 \Rightarrow \sum p_i^2 = \frac{ESS}{\hat{\beta}_2^2} = \frac{ESS}{10,44716^2} \approx 210,55$.

Cách 2:

$$se(\hat{\beta}_2) = \frac{\hat{\sigma}}{\sqrt{\sum p_i^2}} \Rightarrow \sum p_i^2 = \frac{\hat{\sigma}^2}{se^2(\hat{\beta}_2)} = \frac{41,46273^2}{2,857461^2} \approx 210,55$$

Do đó:

$$se(\hat{Q}_0) = \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(P_0 - \bar{P})^2}{\sum p_i^2}} = 41,46273 \cdot \sqrt{\frac{1}{20} + \frac{(34 - 25,35)^2}{210,55}} \approx 26,399$$

$$se(Q_0) = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(P_0 - \bar{P})^2}{\sum p_i^2}} = 41,46273 \cdot \sqrt{1 + \frac{1}{20} + \frac{(34 - 25,35)^2}{210,55}} \approx 49,153$$

1. Dự báo giá trị trung bình khi $P_0 = 34$:

$$\begin{aligned} E(Q|P_0) &\in (\hat{Q}_0 - se(\hat{Q}_0) \cdot t_{\alpha/2}(n-k); \hat{Q}_0 + se(\hat{Q}_0) \cdot t_{\alpha/2}(n-k)) \\ &= (131,58 - 26,399 \cdot t_{0,025}(18); 131,58 + 26,399 \cdot t_{0,025}(18)) \\ &= (76,116; 187,044) \\ t_{\alpha/2}(n-k) &= t_{0,025}(18) = 2,101 \end{aligned}$$

2. Dự báo giá trị cá biệt khi $P_0 = 34$:

$$\begin{aligned} Q_0 &\in (\hat{Q}_0 - se(Q_0) \cdot t_{\alpha/2}(n-k); \hat{Q}_0 + se(Q_0) \cdot t_{\alpha/2}(n-k)) \\ &= (131,58 - 49,153 \cdot t_{0,025}(18); 131,58 + 49,153 \cdot t_{0,025}(18)) \\ &= (28,31; 234,85) \\ t_{\alpha/2}(n-k) &= t_{0,025}(18) = 2,101 \end{aligned}$$

Bài 4:

$n = 20$; $k = 3$.

a) Xét sự phù hợp của hàm hồi quy: $\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{cases}$

$$F_{qs} = F(2, 17) = 531,8780 > F_{\alpha}(k-1, n-k) = F_{0,05}(2, 17) = 3,59$$

\Rightarrow Bác bỏ H_0 , hàm hồi quy là phù hợp.

- P – value của các tham số hồi quy < 0,05. Suy ra các biến độc lập có ý nghĩa, chúng giải thích được lượng bán.

b) Khi P tăng 1 nghìn đồng, PC không đổi thì lượng bán dao động trong khoảng tin cậy của β_2 .

$$\begin{aligned} &(\hat{\beta}_2 - t_{\alpha/2}(n-k) \cdot se(\hat{\beta}_2); \hat{\beta}_2 + t_{\alpha/2}(n-k) \cdot se(\hat{\beta}_2)) \\ &= (-14,76790 - 2,110 \cdot 0,517617; -14,76790 + 2,110 \cdot 0,517617) \\ &= (-15,86; -13,676) \end{aligned}$$

Hay lượng bán giảm từ 13,676 đến 15,86 hộp.

Khoảng tin cậy phía trái:

$$(-\infty; \hat{\beta}_2 + se(\hat{\beta}_2) \cdot t_{\alpha}(n-k)) = (-\infty; -14,76790 + 1,74 \cdot 0,517617)$$

$$= (-\infty; -13,87)$$

Hay, khi P tăng 1 nghìn đồng, PC không đổi thì lượng bán tăng tối đa - 13,87 hay cách nói khác là giảm tối thiểu 13,87 hộp.

Khoảng tin cậy phía phải:

$$\begin{aligned} (\hat{\beta}_2 - se(\hat{\beta}_2) \cdot t_{\alpha}(n-k); +\infty) &= (-14,76790 - 1,74 \cdot 0,517617; +\infty) \\ &= (-15,67; +\infty) \end{aligned}$$

Hay, khi P tăng 1 nghìn đồng, PC không đổi thì lượng bán tăng tối thiểu - 15,67 hay cách nói khác là giảm tối đa 15,67 hộp.

Với $t_{\alpha/2}(n-k) = t_{0,025}(17) = 2,11$ và $t_{\alpha}(n-k) = t_{0,05}(17) = 1,74$.

c) Cặp giả thuyết kiểm định: $\begin{cases} H_0: \beta_3 \leq 10 \\ H_1: \beta_3 > 10 \end{cases}$

$$t_{qs} = \frac{\hat{\beta}_3 - 10}{se(\hat{\beta}_3)} = \frac{11,62597 - 10}{0,473371} \approx 3,4348.$$

Miền bác bỏ: $t > t_{\alpha}(n-k) = t_{0,05}(17) = 1,74$.

$t_{qs} \in$ Miền bác bỏ \Rightarrow Bác bỏ H_0 , chấp nhận H_1 . Có thể cho rằng: Khi P giữ nguyên, PC tăng 1 thì Q tăng hơn 10 hộp.

d) Cặp giả thuyết kiểm định: $\begin{cases} H_0: \beta_2 = 10 \\ H_1: \beta_2 \neq 10 \end{cases}$

$$t_{qs} = \frac{\hat{\beta}_2 - 10}{se(\hat{\beta}_2)} = \frac{-14,7679 - 10}{0,517671} \approx -44,8499.$$

Miền bác bỏ: $|t| > |t_{\alpha/2}(n-k)| = |t_{0,025}(17)| = 2,110$.

$t_{qs} \in$ Miền bác bỏ \Rightarrow Bác bỏ H_0 , chấp nhận H_1 . Không thể cho rằng: Khi PC giữ nguyên, P tăng 1 thì Q tăng 10 hộp.

e) Kiểm định bỏ biến PC: $\begin{cases} H_0: \beta_3 = 0 \\ H_1: \beta_3 \neq 0 \end{cases}$

$$F_{qs} = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n-k}{m} = \frac{0,98427 - 0,42615}{1 - 0,98427} \cdot \frac{20-3}{1} \approx 603,1812 > F_{\alpha}(m, n-k) = F_{0,05}(1, 17) = 4,45.$$

\Rightarrow Bác bỏ H_0 . Không nên loại biến PC ra khỏi mô hình.

Bài 5:

a) Phương trình hồi quy tổng thể: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + u$

Phương trình hồi quy mẫu: $\hat{Y} = 2,189 + 0,372 \cdot X_2 - 0,2931 \cdot X_3 + 0,128 \cdot X_4 + 0,091 \cdot X_5$

Ta có: $F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{k-1} = 89,928 \Rightarrow \frac{R^2}{1-R^2} = 89,928 \cdot \frac{5-1}{20-5} = 23,9808 \Rightarrow R^2 \approx 0,95997$

Vậy các biến độc lập giải thích được khoảng 96% sự thay đổi của biến phụ thuộc.

b)

Ta kiểm định giả thuyết: $\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases}$. Miền bác bỏ là $|t| > t_{\alpha/2}(n-k) = t_{0,025}(15) = 2,131$

Biến X2: p-value < 0,05. Bác bỏ H_0 . X2 có ý nghĩa thống kê.

Biến X3: t = -4,55. Thuộc miền bác bỏ, bác bỏ H_0 . X3 có ý nghĩa thống kê.

Biến X4: t = 0,128/0,099 \approx 1,293. Không thuộc miền bác bỏ. Chưa có cơ sở bác bỏ H_0 . Biến X4 không có ý nghĩa thống kê.

Biến X5: t = 0,904. Không thuộc miền bác bỏ. Chưa có cơ sở bác bỏ H_0 . Biến X5 không có ý nghĩa thống kê.

c) Khi giá thịt gà tăng 1 đơn vị, lượng tiêu thụ thịt gà sẽ tăng 1 lượng là β_3 đơn vị.

Khoảng tin cậy bên trái của β_3 :

$$\begin{aligned} (-\infty; \hat{\beta}_3 + se(\hat{\beta}_3) \cdot t_{\alpha}(n-k)) &= (-\infty; -0,2931 + \frac{-0,2931}{-4,55} \cdot t_{0,05}(15)) \approx (-\infty; -0,18) \\ se(\hat{\beta}_3) &= \frac{-0,2931}{-4,55} = 0,065; t_{0,05}(15) = 1,753 \end{aligned}$$

Vậy lượng tiêu thụ thịt gà sẽ tăng tối đa là - 0,18 kg/năm.

(Một cách diễn giải khác là: khi giá thịt gà tăng 1 đơn vị lượng tiêu thụ thịt gà giảm tối thiểu 0,18kg/năm).

d) Khi thu nhập tăng 1 đơn vị (các yếu tố khác không đổi), lượng thịt gà tăng β_2 đơn vị. Ta kiểm định β_2 có nhỏ hơn 0.3 hay không. Cặp giả thuyết:
$$\begin{cases} H_0: \beta_2 \geq 0.3 \\ H_1: \beta_2 < 0.3 \end{cases}$$

$$t_{qs} = \frac{\hat{\beta}_2 - 0.3}{se(\hat{\beta}_2)} = \frac{0.372 - 0.3}{0.083} \approx 0.867$$

Miền bác bỏ: $t < -t_\alpha(n - k) = -t_{0.05}(15) = -1.753$. Giá trị quan sát không thuộc miền bác bỏ. Suy ra chưa có cơ sở bác bỏ H_0 . Chưa thể cho rằng khi thu nhập tăng 1 đơn vị (các yếu tố khác không đổi), lượng thịt gà tăng ít hơn 0.3 đơn vị.

e) Tác động của thu nhập và giá thịt gà lên cầu thịt gà lần lượt là β_2 ; β_3 .

Khi thu nhập tăng, cầu về hàng hóa thịt gà sẽ tăng lên, dấu của β_2 là dương (biểu thị quan hệ thuận chiều).

Khi giá thịt gà tăng, cầu về hàng hóa thịt gà sẽ giảm, dấu của β_3 là âm (biểu thị quan hệ nghịch chiều). Tác động của thu nhập và giá thịt gà lên cầu thịt gà có chiều ngược nhau. Nên để so sánh chúng, ta sử dụng giá trị tuyệt đối. Tức là so sánh $|\beta_2| = \beta_2$ và $|\beta_3| = -\beta_3$.

Cặp giả thuyết kiểm định tác động của thu nhập lên cầu thịt mạnh hơn tác động của giá lên cầu thịt là:

$$\begin{cases} H_0: \beta_2 \leq -\beta_3 \\ H_1: \beta_2 > -\beta_3 \end{cases}$$

Buổi 2. Khuyết tật mô hình

A. Đa cộng tuyến – Multicollinearity:

Đa cộng tuyến (ĐCT): Hiện tượng tồn tại quan hệ tuyến tính giữa các biến độc lập.

- ĐCT hoàn hảo: $\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} = 0$.
 - ĐCT không hoàn hảo: $\lambda_0 + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + v_i = 0$.
- (với $\lambda_0; \lambda_1 \dots$ là các hằng số không đồng thời bằng 0; v_i là sai số ngẫu nhiên).

Hậu quả:

ĐCT hoàn hảo: Không ước lượng được mô hình.

ĐCT không hoàn hảo: Vẫn ước lượng được mô hình:

- Ước lượng OLS vẫn tuyến tính, không chệch, tốt nhất.

– Phương sai và sai số tiêu chuẩn của ước lượng lớn, làm khoảng tin cậy lớn, thống kê $t = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$ thấp, UL dễ mất ý nghĩa. Không ảnh hưởng đến dự báo.

– Các ước lượng OLS nhạy cảm với sự thay đổi của dữ liệu, dấu của các hệ số có thể bị ước lượng sai.

Nhận biết:

- R^2 cao ($R^2 > 0.8$), t thấp.
- Hệ số tương quan riêng giữa các biến độc lập $r > 0.8$.
- Hồi quy phụ:

Xét mô hình hồi quy k biến: $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$.

Hồi quy phụ là hồi quy mô hình giữa các biến độc lập X_2, X_3, \dots, X_k với nhau.

Giả sử hồi quy X_2 theo các biến còn lại: $X_2 = \alpha_1 + \alpha_2 X_3 + \dots + \alpha_{k-1} X_k + v$ (*). Thu được $R_{phụ}^2$.

Nếu mô hình (*) là phù hợp, ta có X_2 phụ thuộc tuyến tính vào X_3, \dots, X_k , hay mô hình có đa cộng tuyến.

Kiểm định sự phù hợp của (*) có thể dùng thống kê F:

$$F_{qs} = \frac{R_{phụ}^2}{1 - R_{phụ}^2} \cdot \frac{n - (k - 1)}{(k - 1) - 1} = \frac{R_{phụ}^2}{1 - R_{phụ}^2} \cdot \frac{n - k + 1}{k - 2}$$

– Thừa số tăng phương sai (hay Nhân tử phóng đại phương sai): $VIF = \frac{1}{1 - R_{phụ}^2}$ với $R_{phụ}^2$ là hệ số xác định của mô hình hồi quy phụ. Nếu $VIF > 10$ (tương đương $R_{phụ}^2 > 0,9$) thì có ĐCT.

Cách khắc phục:

Bỏ qua ĐCT nếu $VIF < 10$, mô hình dùng để dự báo chứ không phải kiểm định.

- Bổ sung dữ liệu hoặc tìm dữ liệu mới.
- Sử dụng thông tin tiên nghiệm.
- Thay đổi dạng hàm.
- Bỏ bớt biến độc lập.
- Kết hợp số liệu chéo và số liệu chuỗi thời gian.
- Sử dụng sai phân cấp 1:

Thay vì ước lượng $Y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + u_t$, ta ước lượng $\Delta y = \alpha_1 + \alpha_2 \Delta x_2 + \alpha_3 \Delta x_3 + v_t$

với $\Delta y = Y_t - Y_{t-1}$, $\Delta x_2 = X_{2,t} - X_{2,t-1}$, $\Delta x_3 = X_{3,t} - X_{3,t-1}$. Mô hình này có thể không còn đa cộng tuyến hoặc làm giảm đa cộng tuyến.

B. Phương sai sai số thay đổi – Heteroskedasticity:

Phương sai sai số (PSSS) thay đổi: Hiện tượng mỗi sai số u_i có một phương sai riêng:

$$\text{var}(u_i) = \sigma_i^2.$$

Hậu quả:

- Ước lượng OLS vẫn tuyến tính, không chệch, nhưng không tốt nhất (không còn nhỏ nhất).
- Phương sai của ước lượng bị chệch, khoảng tin cậy và kiểm định t không còn giá trị.
- Các dự báo không còn hiệu quả.

Nhận biết:

- Phương pháp định tính: Sử dụng đồ thị của e_i , e_i^2 theo X hoặc \hat{Y}_i . Nếu đồ thị biến đổi có quy luật (không phải đường nằm ngang song song trục hoành) thì có PSSS thay đổi.
- Phương pháp định lượng: Kiểm tra sự phụ thuộc của PSSS σ_i^2 theo 1 biến nào đó.

+ Kiểm định Park:

Giả sử $\ln\sigma_i^2 = \alpha_1 + \alpha_2 \ln X_i + v_i$.

- Ước lượng mô hình gốc thu được e_i .

Thay $\ln\sigma_i^2$ bởi $\ln e_i^2$.

- Ước lượng hồi quy:

$$\ln e_i^2 = \alpha_1 + \alpha_2 \ln X_i + v_i$$

- Xét cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = 0 \\ H_1: \alpha_2 \neq 0 \end{cases}$

Nếu bác bỏ H_0 thì có PSSS thay đổi.

+ Kiểm định Glejser:

Giả sử $\sigma_i = \alpha_1 + \alpha_2 Z_i + v_i$.

(Z_i có thể là: $X_i, \frac{1}{X_i}, \sqrt{X_i}, \frac{1}{\sqrt{X_i}}$)

- Ước lượng mô hình gốc thu được e_i .

Thay σ_i bởi $|e_i|$.

- Ước lượng hồi quy: $|e_i| = \alpha_1 + \alpha_2 Z_i + v_i$

- Xét cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = 0 \\ H_1: \alpha_2 \neq 0 \end{cases}$

Nếu bác bỏ H_0 thì PSSS thay đổi.

+ Kiểm định White:

Xét mô hình hồi quy 3 biến: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$.

Giả sử: $\sigma_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$

- Ước lượng mô hình gốc thu được e_i . Thay σ_i^2 bởi e_i^2 .

– Ước lượng hồi quy: $e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$ (1)

- Xét cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0 \\ H_1: \text{có ít nhất 1 } \alpha \text{ khác 0} \end{cases}$

– Tính $LM = \chi_{qs}^2 = n \cdot R^2$ (R^2 là hệ số xác định của (1), n là số quan sát ở mô hình (1))

Nếu $\chi_{qs}^2 > \chi_{\alpha}^2(k-1)$ thì bác bỏ H_0 , PSSS thay đổi. (k là số hệ số hồi quy của (1)).

+ Kiểm định Breusch – Pagan (BP):

Xét mô hình hồi quy 3 biến: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$.

Giả sử: $\sigma_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i$.

- Ước lượng mô hình gốc thu được e_i . Thay σ_i^2 bởi e_i^2 .

– Ước lượng hồi quy: $e_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i$ (2)

- Xét cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = \alpha_3 = 0 \\ H_1: \text{có ít nhất 1 } \alpha \text{ khác 0} \end{cases}$

– Tính $LM = \chi_{qs}^2 = n \cdot R^2$ (R^2 là hệ số xác định của (2), n là số quan sát ở mô hình (2))

Nếu $\chi_{qs}^2 > \chi_{\alpha}^2(k-1)$ thì bác bỏ H_0 , PSSS thay đổi. (k là số hệ số hồi quy của (2)).

+ Kiểm định Breusch – Pagan – Godfrey (BPG):

Xét mô hình hồi quy k biến: $Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + u$

Giả sử: $\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi} + v_i$ với Z_2, \dots, Z_m là các biến độc lập X_i .

- Ước lượng mô hình gốc thu được e_i .

– Tính: $\tilde{\sigma}^2 = \frac{\sum e_i^2}{n}$

– Tạo biến: $p_i = \frac{e_i^2}{\tilde{\sigma}^2}$

– Ước lượng hồi quy: $p_i = \alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi} + v_i$ (3). thu được ESS.

- Xét cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = \alpha_3 = \dots = \alpha_m = 0 \\ H_1: \text{có ít nhất 1 } \alpha \text{ khác 0} \end{cases}$

– Tính $\theta = \chi_{qs}^2 = \frac{1}{2} ESS$.

Nếu $\theta > \chi_{\alpha}^2(m-1)$ thì bác bỏ H_0 , PSSS thay đổi. (m là số hệ số hồi quy của (3)).

Các kiểm định trên đều có thể dùng p – value để đánh giá: p – value < α thì bác bỏ H_0 .

Cách khắc phục:

Thay đổi dạng hàm hoặc thực hiện theo các trường hợp sau:

- Đã biết σ_i : Phương pháp bình phương nhỏ nhất tổng quát (GLS): Chia cả 2 vế mô hình gốc cho σ_i .
- Chưa biết σ_i : Giả định PSSS tỷ lệ với: biến giải thích, bình phương biến giải thích, hoặc bình phương giá trị kỳ vọng của Y.

VD: Giả sử PSSS tỷ lệ với bình phương biến giải thích: $var(u_i) = c \cdot X_i^2$ với $c = const$.

Chia cả 2 vế của mô hình hồi quy đơn $Y_i = \beta_1 + \beta_2 \cdot X_i + u_i$ cho X_i ta được: $\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i}$

Đặt $Y'_i = \frac{Y_i}{X_i}, X'_i = \frac{1}{X_i}, v_i = \frac{u_i}{X_i} \Rightarrow Y'_i = \beta_2 + \beta_1 \cdot X'_i + v_i$.

Sai số mới là v_i có PS không đổi: $var(v_i) = var\left(\frac{u_i}{X_i}\right) = \frac{1}{X_i^2} var(u_i) = c = const$.

C. Tự tương quan – Autocorrelation:

Tự tương quan (TTQ): Hiện tượng các sai số ngẫu nhiên (nhiều) phụ thuộc lẫn nhau:

$$cov(u_i, u_j) \neq 0 (i \neq j)$$

Tự tương quan bậc p: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$ (ε_t là sai số ngẫu nhiên của mô hình)

Hậu quả:

- ƯL OLS vẫn tuyến tính, không chệch nhưng không tốt nhất (không còn nhỏ nhất)
- Phương sai của ước lượng bị chệch, các kiểm định t và F cùng R^2 không đáng tin cậy.
- Các dự báo không còn hiệu quả.

Nhận biết:

– Phương pháp định tính: Xét đồ thị của u_t theo thời gian t hoặc theo u_{t-1} . Nếu u_t phân phối 1 cách ngẫu nhiên thì không có TTQ và ngược lại.

– Phương pháp định lượng:

+ *Kiểm định Durbin – Watson: Điều kiện sử dụng:*

- Chỉ dùng để kiểm định tự tương quan bậc 1.
- Mô hình ban đầu phải có hệ số chặn, biến X phi ngẫu nhiên.
- Không có biến trễ của biến phụ thuộc, không có quan sát bị thiếu.

Thông kê d Durbin – Watson: $d = \frac{\sum(e_t - e_{t-1})^2}{\sum e_t^2}$. Đặt: $\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_t^2} \in (-1, 1)$

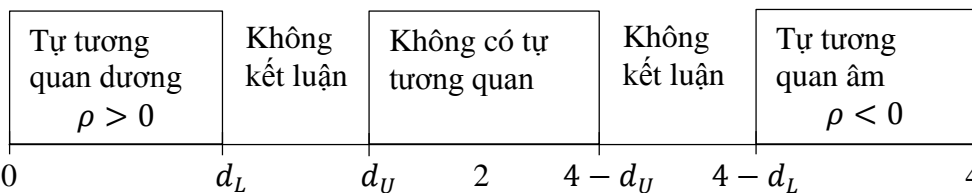
Ta có: $d \approx 2(1 - \hat{\rho}) \in (0, 4)$.

$d = 4 \Leftrightarrow \hat{\rho} = -1$: Tự tương quan hoàn hảo âm.

$d = 2 \Leftrightarrow \hat{\rho} = 0$: Không có tự tương quan.

$d = 0 \Leftrightarrow \hat{\rho} = 1$: Tự tương quan hoàn hảo dương.

Quy tắc kiểm định: Với n quan sát, $k' = k - 1$, với k là số tham số của mô hình, mức ý nghĩa 5%, tra bảng để tìm d_L và d_U . Tính d và so sánh theo sơ đồ sau:



Giả sử hồi quy theo OLS: n = 80 quan sát, k = 2, và giá trị tính toán của d = 1,62. Từ phụ lục bảng D, ta tra cứu được: $d_L = 1,611, d_U = 1,662$. D nằm giữa d_L và d_U nên theo quy tắc trên ta không có kết luận về tính tự tương quan.

+ *Kiểm định Breusch – Godfrey (BG):*

Xét mô hình $Y_t = \beta_1 + \beta_2 X_t + u_t$ và giả sử $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$.

Kiểm định giả thuyết: $\left[\begin{array}{l} H_0: \rho_1 = \rho_2 = \dots = \rho_p = 0 \text{ (không có tự tương quan)} \\ H_1: \text{tồn tại 1 } \rho \text{ nào đó khác 0 (có tự tương quan)} \end{array} \right.$

- Ước lượng mô hình gốc thu được phần dư e_t .
- Từ các giá trị của e_t ước lượng mô hình sau bằng OLS thu được R^2 :

$$e_t = \beta_1 + \beta_2 X_t + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \dots + \rho_p e_{t-p} + v_t$$

- Với n đủ lớn, $LM = (n - p)R^2$ phân phối xấp xỉ $\chi^2(p)$. (n là số quan sát ở mô hình ban đầu)
- Nếu $(n - p)R^2 > \chi^2_\alpha(p)$ thì bác bỏ H_0 , tồn tại tự tương quan ở 1 bậc nào đó.

Các kiểm định trên đều có thể dùng p - value để đánh giá: p - value $< \alpha$ thì bác bỏ H_0 .

Cách khắc phục:

Với tự tương quan bậc 1:

Bước 1: Tính ước lượng $\hat{\rho}$ của ρ : $u_t = \rho u_{t-1} + \varepsilon_t$.

Có thể tính bằng biểu thức $\hat{\rho} \approx 1 - \frac{d}{2}$ hoặc lấy giá trị hệ số góc trong hồi quy: $e_t = \rho e_{t-1} + \varepsilon_t$.

Bước 2: Đổi biến: $Y_t^* = Y_t - \hat{\rho} \cdot Y_{t-1}$, $X_t^* = X_t - \hat{\rho} \cdot X_{t-1}$.

Và ước lượng mô hình mới: $Y_t^* = \beta_1^* + \beta_2^* \cdot X_t^* + v_t$. Mô hình này có thể không còn tự tương quan hoặc làm giảm tự tương quan.

D. Nhiễu không phân phối chuẩn – Normality:

Hậu quả:

- Sai số ngẫu nhiên không tuân theo quy luật chuẩn làm thống kê t không tuân theo quy luật Student, thống kê F không tuân theo quy luật Fisher, không đáng tin cậy khi cỡ mẫu nhỏ, chỉ đáng tin cậy khi cỡ mẫu lớn.

Nhận biết:

- Phương pháp định tính: Xét đồ thị phân phối của phần dư. Nếu đồ thị cân xứng thì có thể cho rằng nhiễu phân phối chuẩn và ngược lại.

- Phương pháp định lượng:

+ Kiểm định Jarque – Bera:

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

Với: S là hệ số bất đối xứng (Skewness), K là độ nhọn (Kurtosis).

- Xét cặp giả thuyết: $\begin{cases} H_0: u_i \text{ phân phối chuẩn} \\ H_1: u_i \text{ không phân phối chuẩn} \end{cases}$

- Nếu $JB > \chi^2_\alpha(2)$ thì bác bỏ H_0 , nhiễu không phân phối chuẩn, ngược lại, nhiễu phân phối chuẩn.

Có thể sử dụng p - value (trong bảng kết quả thống kê) để đánh giá: nếu p - value $< \alpha$ thì bác bỏ H_0 .

E. Mô hình thiếu biến bậc cao:

Kiểm định RESET (hay Ramsey RESET) về thiếu các biến bậc cao ($\hat{Y}^2, \hat{Y}^3 \dots$):

Cặp giả thuyết: $\begin{cases} H_0: \text{Mô hình không thiếu biến} \\ H_1: \text{Mô hình thiếu biến} \end{cases}$

Với:

+ $R_1^2; R_2^2$ lần lượt là hệ số xác định của mô hình khi có biến bậc cao và khi không có biến bậc cao.

+ m là số biến bậc cao muốn khảo sát.

+ $n; k$ lần lượt là số quan sát và số hệ số hồi quy của mô hình khi có biến bậc cao.

(cách làm giống như Kiểm định bỏ biến)

Nếu:

$$F_{qs} = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - k}{m} > F_\alpha^{(m, n-k)}$$

hoặc p - value $< \alpha$ thì bác bỏ H_0 , mô hình thiếu biến.

(cách làm giống như Kiểm định bỏ biến)

Một số thuật ngữ trong bảng kết quả của các phần mềm Kinh Tế Lượng

• VIF; Variance Inflation Factor	Thừa số tăng phương sai
• Heteroskedasticity	Phương sai sai số thay đổi
• Homoskedasticity	Phương sai sai số không đổi
• White's Test White Heteroskedasticity Test	Kiểm định White về PSSS thay đổi
• Heteroskedasticity Test Glejser	Kiểm định Glejser về PSSS thay đổi
• Ramsey RESET Test RESET Specification Test	Kiểm định RESET về mô hình thiếu biến
• Breusch – Godfrey Test	Kiểm định BG về tự tương quan
• Breusch – Pagan	Kiểm định BP về PSSS thay đổi
• Durbin – Watson Statistic	Kiểm định Durbin -Watson
• Chi- squared (n)	Phân phối Khi bình phương bậc n
• Normality Test Jarque – Bera	Kiểm định về phân phối chuẩn của nhiễu
• Residual	Phần dư
• Null hypothesis	Giả thuyết H_0

Lưu ý :

– Trong tất cả các kiểm định về các khuyết tật của mô hình (kiểm định về phương sai của sai số, kiểm định về tự tương quan, kiểm định tính phân phối chuẩn của nhiễu, RESET), giả thuyết H_0 là mô hình không có khuyết tật, H_1 là mô hình có khuyết tật.

BÀI TẬP BUỔI 2

A. Bài Tập Trắc Nghiệm

1. Hiện tượng đa cộng tuyến xảy ra khi:

- A. Mô hình có nhiều biến giải thích
- B. Biến phụ thuộc được biểu diễn bởi 1 tổ hợp tuyến tính của các biến còn lại
- C. 1 biến giải thích được biểu diễn bởi 1 tổ hợp tuyến tính của các biến còn lại
- D. Không có đáp án nào đúng

2. Mô hình có đa cộng tuyến sẽ dẫn tới hậu quả nào sau đây:

- A. Các ước lượng tuyến tính, không chệch, không có phương sai bé nhất trong các ước lượng
- B. Hệ số xác định cao dù giá trị thống kê t có nhiều ý nghĩa
- C. Giá trị thống kê t lớn làm giảm khả năng bác bỏ giả thuyết không
- D. Khoảng tin cậy lớn, ước lượng thiếu chính xác.

3. Cho bảng tương quan giữa các biến độc lập với nhau:

	X1	X2	X3
X1	1		
X2	0.84243	1	
X3	0.94565	0.93457	1

Hỏi phương trình hồi quy mắc lỗi gì?

- A. Đa cộng tuyến
- B. Tự tương quan
- C. PSSS thay đổi

4. Trong White-test thì biến phụ thuộc là gì?

- A. e_i
- B. e_i^2
- C. $|e_i|$
- D. $\ln(e_i^2)$

5. Các điều kiện nào sau đây cần phải có để test Durbin Watson có ý nghĩa?

- (i) Hàm hồi quy phải có hệ số chặn
 - (ii) Các biến hồi quy độc lập không được lấy ngẫu nhiên
 - (iii) Không có biến phụ thuộc trễ trong mô hình
 - (iv) Không có biến độc lập trễ trong mô hình
- (a) Chỉ (i) (ii) (iii)
 - (b) Chỉ (i) và (ii)
 - (c) Tất cả
 - (d) Chỉ (i) (ii) và (iv)

6. Hậu quả của phương sai sai số thay đổi?

- (i) Các ước lượng hệ số chệch
- (ii) Se của hệ số góc sẽ rất nhỏ
- (iii) Các suy diễn thống kê (kiểm định và khoảng tin cậy) có thể không chính xác

- a. (i)
- b. (iii)
- c. Tất cả
- d. Chỉ (i) và (ii)

7. Kiểm định cho lỗi normality của phần dư, thu được Chi-square(2)=31.03 với p-value=0.0000, kết luận:

- a. Error phân phối chuẩn
- b. Error không phân phối chuẩn
- c. Chưa kết luận được

8. Cách nào sau đây không được dùng để chữa bệnh đa cộng tuyến

- a. Bỏ 1 biến giải thích
- b. Biến đổi dữ liệu thành logarith
- c. Biến đổi 2 biến giải thích thành 1 tỉ số
- d. Thu thập thêm số liệu cho tất cả các biến

9. Sử dụng test Durbin watson cho mô hình có hệ số chặn và gồm 3 biến X với 100 quan sát được $d=3,6$. Cho $dL=1,613$; $dU=1,736$. Kết luận nào sau đây đúng?

- a. Có tự tương quan dương
- b. Có tự tương quan âm
- c. Không có tự tương quan
- d. Chưa thể kết luận

10. Nếu xảy ra TTQ âm, giá trị nào sau đây sẽ phù hợp với d của Durbin watson?

- (a) gần 0
- (b) gần 1
- (c) gần 2
- (d) gần 4

11. Sử dụng White test có 3 biến X. Hỏi hàm hồi quy có thể có bao nhiêu tham số?

- (a) 5
- (b) 6
- (c) 10
- (d) 11

12. Kiểm định Durbin Waston có $d=0.039$. Mô hình có kết luận gì?

- A. Phương sai sai số thay đổi
- B. Tự tương quan dương
- C. Tự tương quan âm
- D. Các phương án trên đều sai

13. Cho phương trình :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$$

VIF

X1 16.55

X2 38.86

X3 45.09

Mô hình có bệnh gì không?

- A. Tự tương quan
- B. Đa cộng tuyến
- C. Nhiều không phân phối chuẩn
- D. Không bệnh

14. Cho pt hồi quy $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$ sử dụng kiểm định White phát hiện PSSS thay đổi, pt hồi quy phụ có bao nhiêu tham số?

- A. 4
- B. 5

- C. 6
- D. 7

15. Sử dụng pt ở câu trên, mẫu có 100 quan sát, mức ý nghĩa 5% thì giá trị tới hạn χ^2 là ?

- A. $\chi^2(4)=9.48773$
- B. $\chi^2(5)=11,0705$
- C. $\chi^2(7) = 14.0671$
- D. $\chi^2(8) = 15.5073$

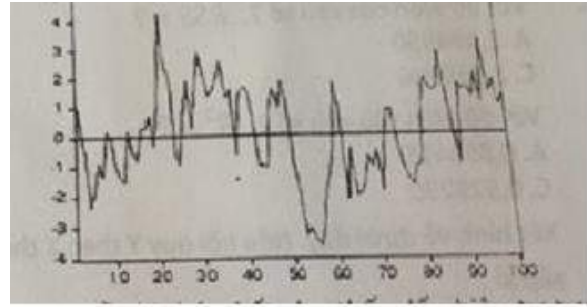
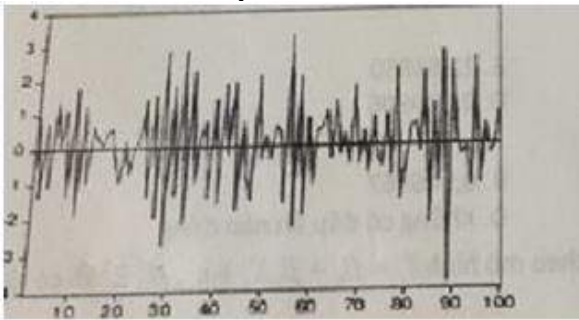
16. Mô hình hồi quy phụ của kiểm định White ở câu trên có $R^2 = 0.7459$. Có kết luận gì về mô hình ban đầu ?

- A. PSSS không đổi.
- B. PSSS thay đổi.
- C. Chưa kết luận được về PSSS.

17. Kiểm định tính chất chuẩn của phần dư cho kết quả thống kê Jacque-Bera=5.66 với p-value =0.0590219. Kết luận ở mức ý nghĩa 5%:

- A. Chuẩn.
- B. Không chuẩn.
- C. Chưa kết luận dc.

18. Cho 2 đồ thị biểu diễn u theo t như sau



- A. Đồ thị thứ nhất cho thấy dấu hiệu tự tương quan âm, đồ thị thứ hai cho thấy dấu hiệu tự tương quan dương
- B. Đồ thị thứ nhất cho thấy dấu hiệu tự tương quan dương, đồ thị thứ hai cho thấy dấu hiệu tự tương quan âm
- C. Cả 2 đồ thị cho thấy dấu hiệu tự tương quan dương
- D. Cả 2 đồ thị cho thấy dấu hiệu của tự tương quan âm

B. Bài Tập Tự Luận

Bài 1:

Hồi quy chi tiêu (Y) theo thu nhập (X) có hệ số chặn thu được kết quả:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.548556	1.800076	2.526869	0.0354
X	0.544619	0.119108	4.572465	0.0018
R-squared	0.723255	Prob(F-statistic)		0.001820

a. Dùng kiểm định White thu được kết quả sau:

White Heteroskedasticity Test:				
F-statistic	1.202720	Probability		0.355644
Obs*R-squared	2.557499	Probability		0.278385
Test Equation: Dependent Variable: RESID^2				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.490826	7.280192	-1.303651	0.2336
X	1.599577	1.058988	1.510478	0.1747
X^2	-0.053123	0.036489	-1.455883	0.1888
R-squared	0.255750	F-statistic		1.202720
Sum squared resid	18.00701	Prob(F-statistic)		0.355644

- Viết phương trình hồi quy của kiểm định White.
- Có kết luận gì về hiện tượng phương sai sai số thay đổi của mô hình?

b. Kiểm định về tính phân phối chuẩn của phần dư cho kết quả như sau. Kết luận?

Normality Test:			
Jarque Bera	0.637444	Probability	0.727077

Bài 2:

Xác định mô hình sau bị mắc khuyết tật gì? Giải thích.

Dependent Variable: Y		Included observations: 100		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-485.9608	95.85601	-5.069695	0.0000
K	1.292811	0.044404	29.11470	0.0000
L	2.214092	0.050943	43.46253	0.0000
R-squared	0.964118	F-statistic		1303.136
Sum squared resid	7221985.	Prob(F-statistic)		0.000000
Ramsey RESET Test: 1 fitted term				
F-statistic	0.006172	Probability		0.937544
White Heteroskedasticity Test: No cross term				
F-statistic	24.53252	Probability		0.000000
Obs*R-squared	50.81036	Probability		0.000000
Normality Test:				
Jarque Bera	9.753215	Probability		0.007623

Bài 3:

Xét các yếu tố ảnh hưởng lên sự thay đổi trong thu nhập trung bình tháng (wggr - %) của người dân quốc gia A bao gồm: tỷ lệ thất nghiệp (unemp - %); chỉ số giá - cpi; tỷ lệ lạm phát (infl - %). Ta có bảng kết quả sau:

Model 1: OLS, using observations 1959-1995 (T = 37)
Dependent variable: wggr

	<i>Coefficient</i>	<i>Std. Error</i>	<i>t-ratio</i>	<i>p-value</i>	
const	3.28358	0.882177	3.7221	0.00073	***
unemp	0.132955	0.160573	0.8280	0.41362	
cpi	-0.0202139	0.0054485	-3.7100	0.00076	***
infl	0.436779	0.0683829		<0.00001	***
Mean dependent var	4.605405	S.D. dependent var		1.976381	
Sum squared resid		S.E. of regression		1.242674	
R-squared	0.637603	Adjusted R-squared		0.604658	
F(3, 33)	19.35346	P-value(F)		2.03e-07	
Log-likelihood	-58.42295	Akaike criterion		124.8459	
Schwarz criterion	131.2896	Hannan-Quinn		127.1176	
rho	0.148440	Durbin-Watson		1.673652	

- White's test for heteroskedasticity -

Null hypothesis: heteroskedasticity not present

Test statistic: LM = 10.5747

with p-value = $P(\text{Chi-square}(9) > 10.5747) = 0.305989$

- Variance Inflation Factors

Minimum possible value = 1.0; Values > 10.0 may indicate a collinearity problem

unemp 1.315

cpi 1.239

infl 1.068

$VIF(j) = 1/(1 - R(j)^2)$, where $R(j)$ is the multiple correlation coefficient between variable j and the other independent variables

a) Viết hàm hồi quy tổng thể và mẫu.

b) Mô hình trên mắc khuyết tật gì? Giải thích.

Bài 4:

Q là lượng bán gas. PG là giá bình gas. PE là giá điện sinh hoạt. PC là giá bếp gas. Mức ý nghĩa 5%.

a) Hồi quy Q theo PG có hệ số chặn. Có xảy ra hiện tượng đa cộng tuyến không?

b) Mô hình 1: Hồi quy Q theo PG, PC, PE thu được:

[1] Observation: 27

Dependent Variable: Q

	Coef.	Std.Error
const	1053.6	123.052
PG	-6.9435	0.626036
PC	0.001737	0.001815
PE	338.15	128.23
R - squared		0.99406

Nghi ngờ mô hình này có hiện tượng đa cộng tuyến vì R - squared cao mà giá trị thống kê t của biến PC thấp. Nêu 1 cách kiểm tra hiện tượng đó.

c) Mô hình 2: Hồi quy PC theo PE và PG thu được:

[2] Observation: 27
Dependent Variable: PC

	Coef.	Std.Error
const	555.7082	50.9517
PE	- 7.3608	3.6730
PG	0.34168	0.020910
R – squared		0.93617

Biến PC có phụ thuộc tuyến tính vào biến PE và PG không. Kết hợp câu b) mô hình này có thể rút ra điều gì?

d) Các ước lượng của mô hình ban đầu có phải là tốt nhất không?

Bài 5:

EX là tổng giá trị xuất khẩu, GAP là tổng sản phẩm nông nghiệp. GIP là tổng sản phẩm công nghiệp của 1 nước A. Mức ý nghĩa 5%.

Mô hình 1: Hồi quy EX theo GAP và GIP thu được phần dư RESID và kết quả:

[1] Observation: 21
Dependent Variable: EX

	Coef.	Std.Error
const	616.0880	1517.432
GAP	- 1.150342	0.610231
GIP	2.254334	0.272353
R – squared		0.986036

Thực hiện Mô hình 2:

[1] Observation: 21
Dependent Variable: log_resid^2

	Coef.	Std.Error
const	8.653524	5.052607
log_GAP	0.532334	0.588999
R – squared		0.636308

Kết quả này dùng để làm gì? Mô hình ban đầu có thể mắc khuyết tật gì?

Bài 6: Kiểm tra một mô hình thu được các kết quả sau:

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance
Variables: fitted values of cigs

chi2(1) = 2.44
Prob > chi2 = 0.1183

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
ei	810	0.0000	0.0000	.	0.0000

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	31.435	1	0.0000
2	31.518	2	0.0000
3	31.529	3	0.0000
4	31.717	4	0.0000

H0: no serial correlation

Mô hình trên mắc những khuyết tật nào? Không mắc khuyết tật nào? Giải thích.

ĐÁP ÁN BUỔI II

A. Bài Tập Trắc Nghiệm

1. D. Không có đáp án nào đúng

Đa cộng tuyến xảy ra khi 1 biến giải thích được biểu diễn bởi 1 tổ hợp tuyến tính của các biến giải thích còn lại.

2. D. Khoảng tin cậy lớn, ước lượng thiếu chính xác.

Xem phần Hậu quả ĐCT.

3. A. Đa cộng tuyến

Dấu hiệu: Hệ số tương quan riêng giữa các biến X_i lớn hơn 0,8.

4. B. e_i^2

Phương trình kiểm định White: $e_i^2 = \alpha_1 + \alpha_2 X_1 + \dots$

5. A. Chỉ (i) (ii) (iii)

Xem phần Điều kiện áp dụng kiểm định Durbin – Watson.

6. B. (iii)

7. B. Error không phân phối chuẩn

Kiểm định lỗi Normality (Nhiều phân phối chuẩn). Giả thuyết H_0 là nhiều phân phối chuẩn.
 $p - \text{value} = 0,0000 < \alpha = 0,05 \Rightarrow$ Bác bỏ H_0 . Nhiều không phân phối chuẩn.

8. C. Biến đổi 2 biến giải thích thành 1 tỷ số.

Cách này dùng để khắc phục PSSS thay đổi.

Xem phần Khắc phục Đa cộng tuyến.

9. B. Có tự tương quan âm

$d_L = 1,613$; $d_U = 1,736 \Rightarrow 4 - d_U = 2,264$; $4 - d_L = 2,387$

$d = 3,6$ nằm trong khoảng từ $4 - d_L$ đến $4 \Rightarrow$ Có tự tương quan âm.

10. D. gần 4.

Xem đồ thị phân quy tắc kiểm định Durbin – Watson. d càng gần 4, có thể kết luận là TTQ âm.

11. C. 10

PT: $e_i^2 = \alpha_1 + \alpha_2 X_1 + \alpha_3 X_2 + \alpha_4 X_3 + \alpha_5 X_1 X_2 + \alpha_6 X_2 X_3 + \alpha_7 X_3 X_1 + \alpha_8 X_1^2 + \alpha_9 X_2^2 + \alpha_{10} X_3^2 + v_t$, có 10 tham số.

12. B. Tự tương quan dương

Tra bảng thấy giá trị nhỏ nhất của d_L là 0,203. $d = 0,039$ thuộc khoảng từ 0 đến d_L , có hiện tượng tự tương quan dương.

13. B. Đa cộng tuyến

Dấu hiệu: Thừa số tăng phương sai VIF > 10 .

14. C. 6

PT: $e_i^2 = \alpha_1 + \alpha_2 X_1 + \alpha_3 X_2 + \alpha_4 X_1 X_2 + \alpha_5 X_1^2 + \alpha_6 X_2^2 + v_t$, có 6 tham số.

15. B. $\chi^2(5) = 11,0705$

Hồi quy White có k tham số thì giá trị tới hạn là $\chi^2(k - 1)$, ở đây $k = 6$.

16. B. PSSS thay đổi

Kiểm định PSSS thay đổi có giả thuyết H_0 : PSSS không đổi.

Ta có: $\chi_{qs}^2 = n \cdot R^2 = 100 \cdot 0,7459 = 74,59 > \chi^2(5) = 11,0705$

\Rightarrow Bác bỏ H_0 , hay PSSS thay đổi.

17. A. Chuẩn

Kiểm định Jarque – Bera: H_0 : Nhiều phân phối chuẩn.

$p - \text{value} = 0,0590219 > \alpha = 0,05 \Rightarrow$ Không có cơ sở bác bỏ H_0 , nhiều là phân phối chuẩn.

18. A.

Trong đồ thị 1, các phần dư đổi dấu liên tục, chứng tỏ tương quan giữa u_t và u_{t-1} là ngược chiều, hay tự tương quan âm.

Trong đồ thị 2, phần dư không đổi dấu liên tục, các giá trị lớn đi cùng các giá trị lớn, các giá trị nhỏ đi cùng các giá trị nhỏ, cho thấy tự tương quan dương.

B. Bài Tập Tự Luận

Bài 1:

a) Phương trình hồi quy phụ của KĐ White:

$$e_i^2 = -9,490826 + 1,599577X_i - 0,053123X_i^2 + v_i$$

Kiểm định White có $P - \text{value} = 0,278685 > \alpha = 0,05 \Rightarrow$ Không bác bỏ giả thuyết H_0 PSSS không đổi.

b) Giả thuyết H_0 : Nhiều phân phối chuẩn. $P - \text{value} = 0,727077 > \alpha = 0,05 \Rightarrow$ Không bác bỏ H_0 , nhiều phân phối chuẩn.

Bài 2:

Lưu ý: H_0 là giả thuyết mô hình không có khuyết tật ở kiểm định tương ứng, $\alpha = 0,05$.

$$Y = -485,9608 + 1,292811.K + 2,214092.L + u_i$$

Kiểm định Ramsey RESET: $p - \text{value} = 0,937544 > \alpha$, không bác bỏ H_0 , mô hình không thiếu biến.

Kiểm định White: $p - \text{value} = 0 < \alpha$, bác bỏ H_0 , PSSS thay đổi.

Kiểm định Jarque – Bera: $p - \text{value} = 0,007623 < \alpha$, bác bỏ H_0 , Nhiều không phân phối chuẩn.

Bài 3:

a) Hàm hồi quy tổng thể:

$$wggr = \beta_1 + \beta_2 unemp + \beta_3 cpi + \beta_4 infl + u_i$$

Hàm hồi quy mẫu:

$$\widehat{wggr} = 3.38358 + 0.132955unemp - 0.0202139cpi + 0.436779infl$$

b) Từ kết quả hồi quy, ta xét:

+ Kiểm định White:

$P - \text{value} = 0.305989 >$ mức ý nghĩa 5%. Suy ra chưa có cơ sở bác bỏ H_0 . Có thể cho rằng mô hình không mắc PSSS thay đổi.

+ Thừa số tăng phương sai VIF:

Các giá trị VIF tương ứng với các biến độc lập đều nhỏ hơn 10. Có thể cho rằng mô hình không mắc đa cộng tuyến.

+ Kiểm định Durbin – Watson:

Giá trị thống kê $d = 1.673652$. Với $n = 37$ quan sát, $k = 4$ tham số hồi quy, mức ý nghĩa 5%, tra bảng ta có:

$$d_L = 1.307; d_U = 1.655$$

Dựa vào quy tắc kiểm định ta thấy thống kê d thuộc khoảng d_U đến $(4 - d_U)$. Vậy mô hình không có tự tương quan.

Lưu ý: H_0 là giả thuyết mô hình không có khuyết tật ở kiểm định tương ứng, $\alpha = 0,05$.

Bài 4:

a) Mô hình hồi quy đơn không có hiện tượng đa cộng tuyến. (ít nhất phải có 2 biến độc lập).

b) Cách kiểm tra phát hiện đa cộng tuyến (chọn 1 cách):

– Hệ số tương quan riêng giữa các biến PE, PC, PG.

– Thừa số tăng phương sai VIF

– Hồi quy phụ giữa các biến độc lập PE, PC, PG.

c) Ta kiểm định ý nghĩa thống kê của các biến PE và PG trong mô hình 2.

Cặp giả thuyết: $\begin{cases} H_0: \beta_i = 0 \\ H_1: \beta_i \neq 0 \end{cases}$. Miền bác bỏ: $|t| > t_{\alpha/2}(n - k) = t_{0.025}(27 - 3) = 2.064$

Biến PE: thống kê $t = -7.3608/3.6730 \approx -2.004$. Không thuộc miền bác bỏ, nên chưa có cơ sở bác bỏ H_0 . Biến PC có thể không phụ thuộc vào PE.

Biến PG: thống kê $t = 0.34168/0.020910 \approx 16.34$. Thuộc miền bác bỏ, nên bác bỏ H_0 . Biến PC phụ thuộc tuyến tính vào PG.

Mô hình 2 chính là mô hình hồi quy phụ giữa các biến độc lập của mô hình 1. Mô hình 2 (hay mô hình hồi quy phụ) có hệ số xác định lớn hơn 0.9, vậy mô hình ban đầu mắc khuyết tật đa cộng tuyến.

d) Các ước lượng của mô hình ban đầu vẫn là tốt nhất. (đa cộng tuyến không ảnh hưởng đến tính chất BLUE).

Bài 5:

Mô hình 2 là hồi quy của logarit phần dư bình phương với logarit của 1 biến độc lập – dạng phương trình của kiểm định Park về PSSS thay đổi. Phương trình:

$$\log(\widehat{resid}^2) = 8.653524 + 0.532334 \cdot \log(GAP) \quad (1)$$

Để xét xem mô hình ban đầu có mắc PSSS thay đổi hay không bằng kiểm định Park, ta kiểm định cặp giả thuyết cho mô hình (1). Cặp giả thuyết: $\begin{cases} H_0: \alpha_2 = 0 \\ H_1: \alpha_2 \neq 0 \end{cases}$

Thông kê $t = 0.532334/0.588999 \approx 0.9038$

Miền bác bỏ: $|t| > t_{\alpha/2}(n - k) = t_{0.05}(19) = 2.069$.

Giá trị quan sát không thuộc miền bác bỏ, nên chưa có cơ sở bác bỏ H_0 , hay phần dư không phụ thuộc vào biến GAP.

Theo kiểm định Park, mô hình có thể không mắc PSSS thay đổi.

Bài 6:

Lưu ý: H_0 là giả thuyết mô hình không có khuyết tật ở kiểm định tương ứng, $\alpha = 0,05$.

+ Kiểm định Breusch – Pagan:

P – value = 0.1183 > mức ý nghĩa 5%. Suy ra chưa có cơ sở bác bỏ H_0 . Mô hình có thể không mắc PSSS thay đổi.

+ Kiểm định Jarque – Bera:

P – value = 0.000 < mức ý nghĩa 5%. Suy ra bác bỏ H_0 . Nhiễu của mô hình không phân phối chuẩn.

+ Kiểm định Breusch – Godfrey:

Các giá trị p – value đều nhỏ hơn mức ý nghĩa 5%. Suy ra bác bỏ H_0 . Mô hình mắc tự tương quan bậc 4 (chọn bậc cao nhất).

Buổi 3. Biến giả

A. Biến định tính được đưa vào mô hình nhằm lượng hóa các thuộc tính về chất. Sử dụng biến giả (thường ký hiệu là D (Dummy)) để đưa biến định tính vào mô hình.

Các biến về chất có thể là: Sự ra quyết định (Có – Không), Giới tính (Nam – Nữ), Môn học (Toán – Tiếng Anh – Kinh tế lượng)...

Một biến giả D sẽ có 2 giá trị: 0 và 1, tương ứng với 2 thuộc tính, nhiều hơn 2 thuộc tính sẽ có 2 biến giả trở lên. Thuộc tính tương ứng với các giá trị biến giả bằng 0 gọi là thuộc tính cơ sở, nhóm quan sát chứa thuộc tính này gọi là nhóm cơ sở. VD:

– Giới tính:

$$D = \begin{cases} 1 & \text{nếu là Nam} \\ 0 & \text{nếu là Nữ} \end{cases}$$

Thuộc tính cơ sở là Nữ.

Nhóm cơ sở là những người giới tính Nữ.

$$(D = 0)$$

– Môn học:

$$D_1 = \begin{cases} 1 & \text{nếu là môn Toán} \\ 0 & \text{nếu là môn khác} \end{cases}$$
$$D_2 = \begin{cases} 1 & \text{nếu là môn T. Anh} \\ 0 & \text{nếu là môn khác} \end{cases}$$

Thuộc tính cơ sở là Kinh tế lượng.

Nhóm cơ sở là những người học Kinh tế lượng.

$$(D_1 = D_2 = 0)$$

Tổng quát:

– Biến định tính có a thuộc tính thì số biến giả là $(a - 1)$.

– Có N biến định tính (chẳng hạn muốn xét nhiều yếu tố tác động 1 lúc), mỗi biến có $a; b; \dots; z$ thuộc tính, số biến giả là:

$$(a - 1) + (b - 1) + \dots + (z - 1) = (a + b + \dots + z) - N$$

B. Ý nghĩa hệ số hồi quy của biến giả trong mô hình:

Xét biến giả có 2 thuộc tính là A, B: $D = \begin{cases} 1: & \text{thuộc tính A} \\ 0: & \text{thuộc tính B} \end{cases}$

Biến XD gọi là biến tương tác của X và D . Ta có nhóm B là nhóm cơ sở.

Dạng 1: $Y = \beta_1 + \beta_2 \cdot X + \beta_3 \cdot D + u$

Nhóm A ($D = 1$): $Y = (\beta_1 + \beta_3) + \beta_2 \cdot X + u$

Nhóm B ($D = 0$): $Y = \beta_1 + \beta_2 \cdot X + u$

$\beta_1 + \beta_3$: Y trung bình của nhóm A khi $X = 0$

β_1 : Y trung bình nhóm B khi $X = 0$

β_3 : Chênh lệch Y trung bình hai nhóm khi cùng giá trị X

β_2 : Lượng thay đổi của Y trung bình khi X thay đổi 1 đơn vị (ở cả 2 nhóm)

Dạng 2: $Y = \beta_1 + \beta_2 \cdot X + \beta_4 \cdot XD$

Nhóm A ($D = 1$): $Y = \beta_1 + (\beta_2 + \beta_4) \cdot X + u$

Nhóm B ($D = 0$): $Y = \beta_1 + \beta_2 \cdot X + u$

β_1 : Y trung bình của mỗi nhóm khi $X = 0$

$\beta_2 + \beta_4$: Lượng thay đổi Y trung bình nhóm A khi X thay đổi 1 đơn vị

β_2 : Lượng thay đổi của Y trung bình nhóm B khi X thay đổi 1 đơn vị

β_4 : Chênh lệch lượng thay đổi của Y trung bình hai nhóm khi X thay đổi 1 đơn vị.

Dạng 3: $Y = \beta_1 + \beta_2 \cdot X + \beta_3 \cdot D + \beta_4 \cdot XD$

Nhóm A ($D = 1$): $Y = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) \cdot X + u$

Nhóm B ($D = 0$): $Y = \beta_1 + \beta_2 \cdot X + u$

$\beta_1 + \beta_3$: Y trung bình của nhóm A khi $X = 0$

β_1 : Y trung bình nhóm B khi $X = 0$

β_3 : Chênh lệch Y trung bình hai nhóm khi $X = 0$

$\beta_2 + \beta_4$: Lượng thay đổi Y trung bình nhóm A khi X thay đổi 1 đơn vị

β_2 : Lượng thay đổi của Y trung bình nhóm B khi X thay đổi 1 đơn vị

β_4 : Chênh lệch lượng thay đổi của Y trung bình hai nhóm khi X thay đổi 1 đơn vị.

Lưu ý:

- Đối với bài tập biến giả, nên viết riêng phương trình hồi quy của mỗi thuộc tính để dễ suy luận. (cho $D = 0,1$).
- Hệ số biến giả dùng để so sánh hệ số chặn của 1 nhóm nào đó với nhóm cơ sở.
- Hệ số biến tương tác dùng để so sánh hệ số góc của 1 nhóm nào đó với nhóm cơ sở.
- Kiểm định giả thuyết hồi quy về các hệ số trong mô hình có biến giả thực hiện tương tự với hệ số hồi quy thông thường.

Bài luyện tập 1:

Biết CT là chi tiêu, TN là thu nhập (đơn vị: triệu đồng). Biến giả D nhận giá trị = 1 nếu là Nam và = 0 nếu là Nữ. Kết quả hồi quy:

Dependent Variable: CT				
Included observations: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	27.09551	1.510313	17.94033	0.0000
D	5.092130	0.739796	6.883152	0.0000
TN	0.555216	0.013928	39.86394	0.0000
R-squared	0.980889	Mean dependent var	89.85000	
Adjusted R-squared	0.979856	F-statistic	949.5470	
Sum squared resid	187.4964	Prob(F-statistic)	0.000000	

Cho hiệp phương sai ước lượng hệ số chặn và hệ số biến D bằng 0,04

- Viết phương trình hồi quy đối với nam và nữ.
- Tìm ƯL điểm chi tiêu trung bình của nam, của nữ khi thu nhập là 100 triệu đồng.
- Tiêu dùng tự định của nam và nữ có khác nhau không?
- Có thể nói tiêu dùng tự định của nam cao hơn của nữ không? Nếu có thì cao hơn trong khoảng nào?
- Tìm ƯL khoảng cho tiêu dùng tự định của nam; của nữ.
- Trong số 40 người tham gia điều tra, có một số người đã có gia đình và một số người chưa có gia đình. Có ý kiến cho rằng khi thu nhập tăng thêm cùng 1 lượng thì những người đã có gia đình có khuynh hướng chi tiêu nhiều hơn những người chưa có gia đình. Nêu cách kiểm tra ý kiến trên.

Bài Tập Tổng Hợp:

Cho kết quả hồi quy sau với GIP là tổng sản phẩm ngành công nghiệp, DI là đầu tư khu vực trong nước, GE là chi tiêu chính phủ. Số liệu của 52 nước đang phát triển, đơn vị triệu USD.

Dependent Variable: GIP				
Method: Least Squares				
Included observations: 52				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5160.358	4141.907	1.245889	0.2187
DI	0.571653	0.068992	8.285835	0.0000
GE	0.682839	0.074374	9.181150	0.0000
R-squared	0.959679	Mean dependent var	64198.19	
Adjusted R-squared	0.958033	S.D. dependent var	132485.3	
S.E. of regression	27140.69	F-statistic	583.1222	
Sum squared resid	3.61E+10	Prob(F-statistic)	0.000000	

- Viết phương trình hồi quy. Giải thích ý nghĩa của các hệ số hồi quy.
- Hàm hồi quy phù hợp không?
- Khi Đầu tư trong nc tăng 1tr USD, Chi tiêu Chính phủ không đổi thì GIP thay đổi trong khoảng nào?
- Khi Chi tiêu chính phủ tăng 1tr USD, Đầu tư trong nước không đổi thì GIP có tăng hơn 0.5tr USD không?
- Kiểm định ý kiến cho rằng tác động của DI và GE đến GIP là bằng nhau, biết cov của ước lượng 2 hệ số góc gần bằng 0.
- Thêm biến đầu tư trực tiếp nước ngoài FDI vào mô hình, thu được kết quả sau:

Dependent Variable: GIP		Included observations: 52		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2200.373	3582.184	0.614255	0.5419
DI	0.516150	0.059942	8.610833	0.0000
GE	0.317418	0.103559	3.065095	0.0036
FDI	4.755274	1.067499	4.454594	0.0001
R-squared	0.971472	F-statistic	544.8594	
Sum squared resid	2.55E+10	Prob(F-statistic)	0.000000	
Ramsey RESET Test:				
F-statistic	1.963698	Probability	0.584441	
White Heteroskedasticity Test:				
F-statistic	3.079677	Probability	0.257954	
Obs*R-squared	4.181637	Probability	0.312584	
Normality Test:				
Jarque Bera	3.905786	Probability	0.295843	

- Viết phương trình hồi quy mới, giải thích ý nghĩa của hệ số của biến FDI.
- Có nên thêm biến FDI vào mô hình hay không?
- Có thể bỏ biến FDI ra khỏi mô hình (tức giữ nguyên mô hình ban đầu) không?
- Khi các yếu tố khác không đổi, FDI tăng 1tr USD thì GIP tăng tối thiểu bao nhiêu?
- Phân tích kết quả của các kiểm định đã được thực hiện. Các ước lượng của mô hình có phải là ước lượng tốt nhất không?

ĐÁP ÁN BUỔI III

Bài luyện tập 1:

- a) Viết phương trình hồi quy: $Y_i = \beta_1 + \beta_2 X_i + \beta_D D + u_i$.
b) Áp dụng công thức ước lượng điểm: $\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0$ cho 2 trường hợp $D = 0; 1$.
c) Chênh lệch tiêu dùng tự định của nam và nữ do biến giả D gây nên.

Kiểm định cặp giả thuyết:
$$\begin{cases} H_0: \beta_D = 0 \\ H_1: \beta_D \neq 0 \end{cases}$$

d) Kiểm định cặp giả thuyết:
$$\begin{cases} H_0: \beta_D \leq 0 \\ H_1: \beta_D > 0 \end{cases}$$

Tìm khoảng tin cậy cho β_D .

e) Tìm khoảng tin cậy của β_1 (tiêu dùng tự định của nữ) và $\beta_1 + \beta_D$ (tiêu dùng tự định của nam).

f) Thêm biến giả D_2 vào mô hình. Thu nhập tăng cùng 1 lượng thì người có gia đình có khuynh hướng chi tiêu nhiều hơn những người chưa có gia đình, chứng tỏ biến giả tác động lên hệ số góc. (Khiến X thay đổi 1 lượng thì Y cũng thay đổi theo)

Bài Tập Tổng Hợp

a) Phương trình hồi quy: $GIP = 5160,358 + 0,571653 \cdot DI + 0,682839 \cdot GE + e_i$

Ý nghĩa của các hệ số hồi quy:

+ $\hat{\beta}_1 = 5160,358$: Khi đầu tư trong nước DI và chi tiêu chính phủ GE bằng 0 thì tổng sản phẩm ngành công nghiệp trung bình là 5160,358 triệu USD.

+ $\hat{\beta}_{DI} = 0,571653$: Khi chi tiêu chính phủ GE giữ nguyên, DI tăng 1 triệu USD thì GIP trung bình tăng 0,571653 triệu USD.

+ $\hat{\beta}_{GE} = 0,682839$: Khi đầu tư trong nước DI giữ nguyên, GE tăng 1 triệu USD thì GIP trung bình tăng 0,682839 triệu USD.

b) Kiểm định sự phù hợp của hàm hồi quy:
$$\begin{cases} H_0: R^2 = 0 \\ H_1: R^2 > 0 \end{cases}$$

Ta có $\text{Prob}(F - \text{statistic}) = 0,000000 < \alpha = 0,05 \Rightarrow$ bác bỏ H_0 , hàm hồi quy là phù hợp.

c) Ta tìm khoảng tin cậy cho β_{DI} :

$$\begin{aligned} & (\hat{\beta}_{DI} - \text{se}(\hat{\beta}_{DI}) \cdot t_{\alpha/2}(n-k); \hat{\beta}_{DI} + \text{se}(\hat{\beta}_{DI}) \cdot t_{\alpha/2}(n-k)) \\ &= (0,571653 - 0,068992 \cdot 2,009575; 0,571653 + 0,068992 \cdot 2,009575) \\ &= (0,433008; 0,710298). \end{aligned}$$

$$(t_{\alpha/2}(n-k) = t_{0,025}(49) = 2,009575)$$

Khi DI tăng 1 triệu USD thì GIP thay đổi trong khoảng $(0,433008; 0,710298)$.

d) Ta kiểm định giả thuyết:
$$\begin{cases} H_0: \beta_{GE} \leq 0,5 \\ H_1: \beta_{GE} > 0,5 \end{cases}$$

$$t_{qs} = \frac{\hat{\beta}_{GE} - 0,5}{\text{se}(\hat{\beta}_{GE})} = \frac{0,682839 - 0,5}{0,074374} = 2,458373. \text{ Miền bác bỏ: } t > t_{\alpha}(n-k) = t_{0,05}(49) = 1,6766$$

$t_{qs} \in$ Miền bác bỏ \Rightarrow Bác bỏ H_0 , hay có thể cho rằng khi DI không đổi, GE tăng 1 triệu USD thì GIP có tăng hơn 0,5 triệu USD.

e) Cặp giả thuyết:
$$\begin{cases} H_0: \beta_{DI} - \beta_{GE} = 0 \\ H_1: \beta_{DI} - \beta_{GE} \neq 0 \end{cases} \text{ với } t_{qs} = \frac{\hat{\beta}_{DI} - \hat{\beta}_{GE}}{\text{se}(\hat{\beta}_{DI} - \hat{\beta}_{GE})}$$

$$\text{Ta có: } \text{se}(\hat{\beta}_{DI} - \hat{\beta}_{GE}) = \sqrt{\text{var}(\hat{\beta}_{DI} - \hat{\beta}_{GE})} = \sqrt{\text{var}(\hat{\beta}_{DI}) + \text{var}(\hat{\beta}_{GE}) - 2\text{cov}(\hat{\beta}_{DI}, \hat{\beta}_{GE})}$$

$$= \sqrt{\text{se}^2(\hat{\beta}_{DI}) + \text{se}^2(\hat{\beta}_{GE})} \text{ (vì } \text{cov}(\hat{\beta}_{DI}, \hat{\beta}_{GE}) \approx 0)$$

$$= \sqrt{0,068992^2 + 0,074374^2} \approx 0,10145 \Rightarrow t_{qs} = \frac{0,571653 - 0,682839}{0,10145} = -1,09597.$$

Miền bác bỏ: $|t| > t_{\alpha/2}(n-k) = t_{0,025}(49) = 2,009575$.

$t_{qs} \notin$ Miền bác bỏ \Rightarrow Không có cơ sở bác bỏ H_0 , hay có thể cho rằng tác động của DI và GE lên GIP là như nhau.

f)

– $GIP = 2200,373 + 0,516150 \cdot DI + 0,317418 \cdot GE + 4,755274 \cdot FDI + e_i$

Ý nghĩa hệ số hồi quy biến FDI:

+ $\hat{\beta}_{FDI} = 4,755274$: Khi DI và GE giữ nguyên, FDI tăng 1 triệu USD thì GIP trung bình tăng 4,755274 triệu USD.

– Xét thêm biến vào mô hình dùng $\overline{R^2}$: nếu hệ số biến thêm vào có ý nghĩa thống kê và $\overline{R^2}$ tăng thì nên thêm biến:

+ Kiểm định ý nghĩa thống kê của biến FDI với giả thuyết không là không có ý nghĩa thống kê: Ta thấy p-value của biến FDI là 0.0001 nhỏ hơn mức ý nghĩa 5%. Suy ra bác bỏ giả thuyết không hay biến FDI có ý nghĩa thống kê.

+ $\overline{R^2}$ mô hình ban đầu: $\overline{R^2} = 0.958003$

+ $\overline{R^2}$ mô hình sau khi thêm biến FDI:

$$\overline{R^2} = 1 - (1 - R^2) \cdot \frac{n-1}{n-k} = 1 - (1 - 0.971472) \cdot \frac{52-1}{52-4} = 0.969689$$

Suy ra sau khi thêm biến FDI thì $\overline{R^2}$ tăng.

Kết hợp cả 2 điều kiện trên ta suy ra nên thêm biến FDI vào mô hình.

– Kiểm định bỏ biến FDI: $\begin{cases} H_0: \beta_{FDI} = 0 \\ H_1: \beta_{FDI} \neq 0 \end{cases}$

$$F_{qs} = \frac{R_L^2 - R_N^2}{1 - R_L^2} \cdot \frac{n-k}{m} = \frac{0,971472 - 0,959679}{1 - 0,971672} \cdot \frac{52-4}{1} \approx 19,8424 > F(1, 48) = 4,08 \Rightarrow \text{Bác bỏ } H_0.$$

Không nên bỏ biến FDI ra khỏi mô hình.

– Khi các yếu tố khác không đổi, FDI tăng 1 triệu thì GIP tăng tối thiểu bao nhiêu:

Ta tìm khoảng tin cậy phía phải:

$$(\hat{\beta}_{FDI} - se(\hat{\beta}_{FDI}) \cdot t_{\alpha}(n-k); +\infty) = (4,755274 - 1,067499 \cdot 1,677224; +\infty)$$

$$= (2,964839; +\infty). ((t_{\alpha}(n-k) = t_{0,05}(48) = 1,677224)$$

Vậy GIP tăng tối thiểu 2,964839 triệu USD.

– Phân tích:

Kiểm định Ramsey RESET: p-value = 0,584441 > α , không bác bỏ H_0 , mô hình không thiếu biến.

Kiểm định White: p-value = 0,312584 > α , không bác bỏ H_0 , PSSS không đổi.

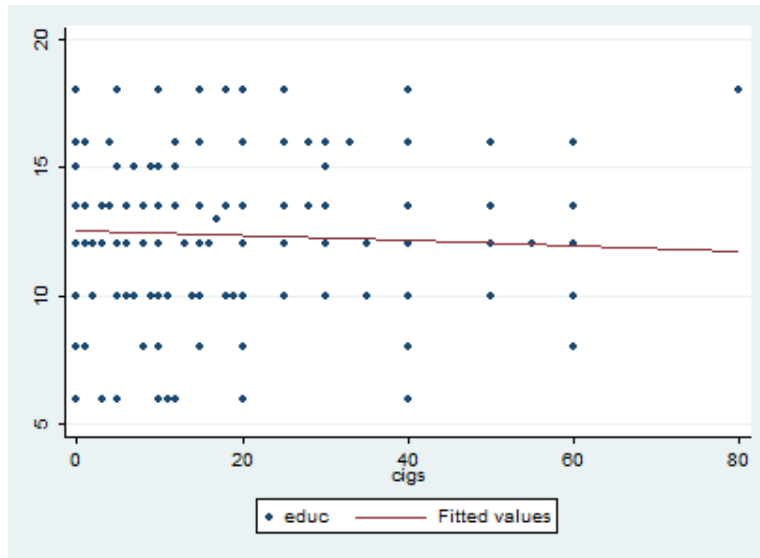
Kiểm định Jarque – Bera: p-value = 0,295843 > α , không bác bỏ H_0 , Nhiễu phân phối chuẩn.

Vì không mắc các khuyết tật được kiểm tra nên các ước lượng của mô hình có thể là tốt nhất.

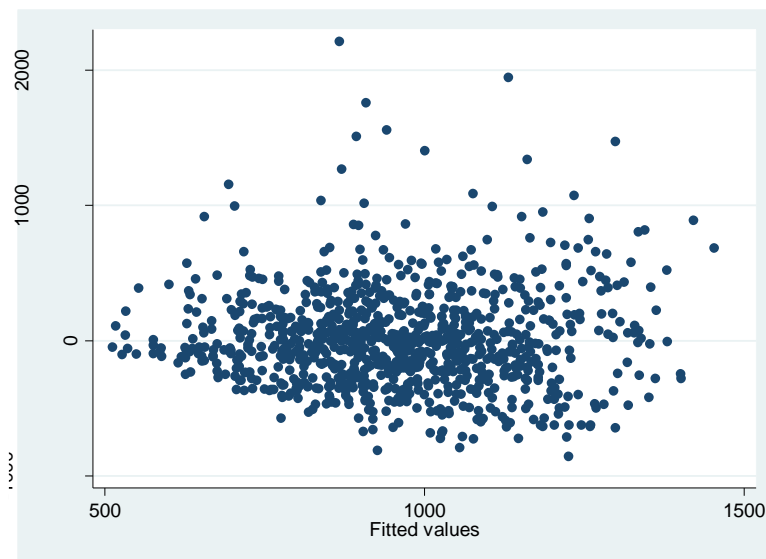
Phụ Lục

ĐỒ THỊ

Đồ thị hàm hồi quy mẫu

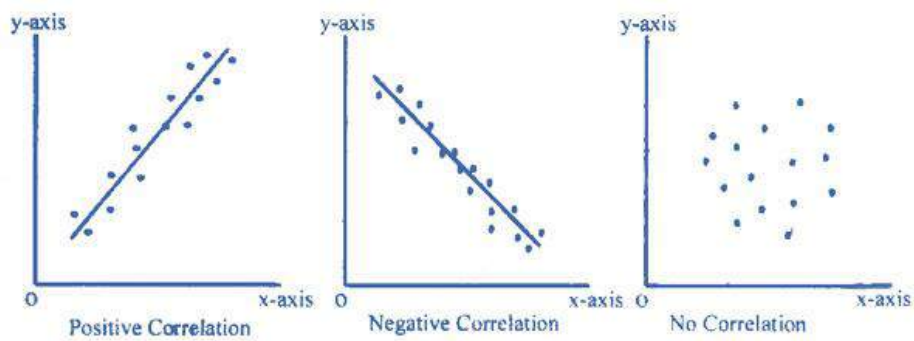


Phương sai sai số thay đổi

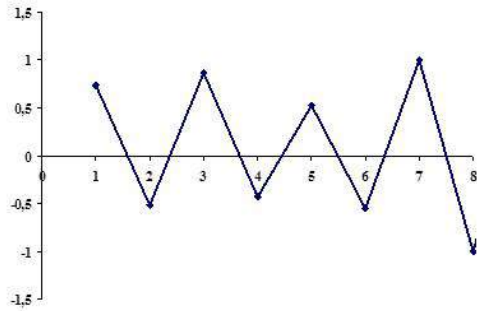


Tự tương quan

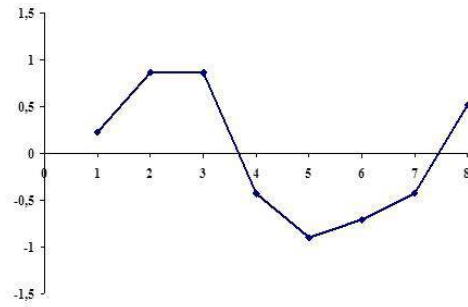
Đồ thị e_t theo e_{t-1}



Đồ thị e_t theo thời gian t



Tự tương quan âm



Tự tương quan dương

STATA

Hồi quy OLS:

Source	SS	df	MS
Model	1562.64873	5	312.529746
Residual	150346.572	804	186.998224
Total	151909.221	809	187.774068

Number of obs	=	810
F(5, 804)	=	1.67
Prob > F	=	0.1390
R-squared	=	0.0103
Adj R-squared	=	0.0041
Root MSE	=	13.675

cigs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
educ	-.3719665	.1695535	-2.19	0.029	-.7047863 - .0391466
cigpric	-.0261197	.10186	-0.26	0.798	-.2260626 .1738231
white	.2311376	1.470453	0.16	0.875	-2.655243 3.117518
age	-.0405122	.0287614	-1.41	0.159	-.0969685 .0159441
income	.0001169	.0000559	2.09	0.037	7.22e-06 .0002266
_cons	14.11225	6.749512	2.09	0.037	.8635045 27.36099

Kiểm định bỏ biến:

- (1) educ = 0
- (2) white = 0

F(2, 804)	=	2.42
Prob > F	=	0.0895

Hệ số tương quan:

	cigs	educ	cigpric	white	age	income
cigs	1.0000					
educ	-0.0480	1.0000				
cigpric	-0.0098	0.0312	1.0000			
white	0.0067	-0.0067	-0.0607	1.0000		
age	-0.0408	-0.1798	0.0294	-0.0298	1.0000	
income	0.0529	0.3353	0.0471	-0.0184	-0.0645	1.0000

Thừa số tăng phương sai VIF:

Variable	VIF	1/VIF
educ	1.16	0.861999
income	1.13	0.885996
age	1.04	0.965565
cigpric	1.01	0.992790
white	1.00	0.995211
Mean VIF	1.07	

Kiểm định White:

White's test for H_0 : homoskedasticity
against H_a : unrestricted heteroskedasticity

chi2(19) = 32.06
Prob > chi2 = 0.0308

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	32.06	19	0.0308
Skewness	56.66	5	0.0000
Kurtosis	8.23	1	0.0041
Total	96.95	25	0.0000

Kiểm định BP:

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
 H_0 : Constant variance
Variables: fitted values of cigs

chi2(1) = 2.44
Prob > chi2 = 0.1183

Kiểm định Durbin – Watson:

Durbin-Watson d-statistic(2, 44) = .2136836

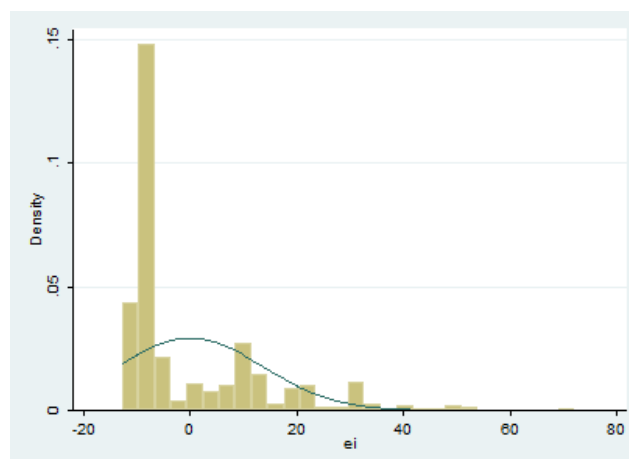
Kiểm định BG:

Breusch-Godfrey LM test for autocorrelation

lags (p)	chi2	df	Prob > chi2
1	31.435	1	0.0000
2	31.518	2	0.0000
3	31.529	3	0.0000
4	31.717	4	0.0000

H_0 : no serial correlation

Đồ thị phân phối của phần dư:



Kiểm định phân phối chuẩn của nhiễu:

Skewness/Kurtosis tests for Normality

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	adj chi2(2)	joint Prob>chi2
ei	810	0.0000	0.0000	.	0.0000

Kiểm định Ramsey RESET:

Ramsey RESET test using powers of the fitted values of *cigs*
Ho: model has no omitted variables
F(3, 801) = 0.66
Prob > F = 0.5751

GRET

Hồi quy OLS:

Model 1: OLS, using observations 1-51
Dependent variable: *crime*

	coefficient	std. error	t-ratio	p-value	
const	1638.80	774.680	2.115	0.0397	**
pov	83.2998	35.4091	2.352	0.0229	**
metro	31.1747	7.35529	4.238	0.0001	***
popdens	0.456305	0.127950	3.566	0.0008	***

Mean dependent var	5098.333	S.D. dependent var	1544.348
Sum squared resid	52171634	S.E. of regression	1053.582
R-squared	0.562504	Adjusted R-squared	0.534579
F(3, 47)	20.14318	P-value (F)	1.55e-08
Log-likelihood	-425.2406	Akaike criterion	858.4811
Schwarz criterion	866.2084	Hannan-Quinn	861.4340

Khoảng tin cậy của hệ số hồi quy:

$t(47, 0.025) = 2.012$

VARIABLE	COEFFICIENT	95% CONFIDENCE INTERVAL	
const	1638.80	80.3490	3197.26
pov	83.2998	12.0658	154.534
metro	31.1747	16.3777	45.9716
popdens	0.456305	0.198902	0.713707

Kiểm định bỏ biến:

Null hypothesis: the regression parameter is zero for *pov*
Test statistic: $F(1, 47) = 5.53423$, p-value 0.0228868

Hệ số tương quan:

Correlation Coefficients, using the observations 1 - 51
5% critical value (two-tailed) = 0.2759 for $n = 51$

crime	pov	metro	popdens	
1.0000	0.3385	0.5440	0.6119	crime
	1.0000	-0.0830	0.3300	pov
		1.0000	0.3159	metro
			1.0000	popdens

Thừa số tăng phương sai VIF:

Variance Inflation Factors

Minimum possible value = 1.0

Values > 10.0 may indicate a collinearity problem

pov	1.174
metro	1.162
popdens	1.295

$VIF(j) = 1/(1 - R(j)^2)$, where $R(j)$ is the multiple correlation coefficient between variable j and the other independent variables

Properties of matrix $X'X$:

1-norm = 96162259
Determinant = 8.5040873e+016
Reciprocal condition number = 1.8453186e-008

Kiểm định White:

White's test for heteroskedasticity

OLS, using observations 1-51

Dependent variable: uhat²

	coefficient	std. error	t-ratio	p-value
const	818170	4.27211e+06	0.1915	0.8491
pov	2688.25	392727	0.006845	0.9946
metro	10610.8	98782.9	0.1074	0.9150
popdens	-6878.55	20001.8	-0.3439	0.7327
sq_pov	2810.25	10148.4	0.2769	0.7832
X2_X3	-2156.11	3291.00	-0.6552	0.5160
X2_X4	593.026	686.731	0.8635	0.3929
sq_metro	110.476	742.821	0.1487	0.8825
X3_X4	21.4834	170.861	0.1257	0.9006
sq_popdens	-1.14204	1.29182	-0.8841	0.3818

Unadjusted R-squared = 0.051736

Test statistic: $TR^2 = 2.638527$,
with p-value = $P(\text{Chi-square}(9) > 2.638527) = 0.976923$

Kiểm định BP:

Breusch-Pagan test for heteroskedasticity

OLS, using observations 1-51

Dependent variable: scaled uhat²

	coefficient	std. error	t-ratio	p-value
const	0.556322	0.994654	0.5593	0.5786
pov	0.0119858	0.0454637	0.2636	0.7932
metro	0.00453606	0.00944385	0.4803	0.6332
popdens	-9.74861e-05	0.000164282	-0.5934	0.5558

Explained sum of squares = 0.791974

Test statistic: LM = 0.395987,
with p-value = $P(\text{Chi-square}(3) > 0.395987) = 0.941070$

Kiểm định Durbin – Watson:

Durbin-Watson statistic = 0.213684
p-value = 8.35738e-011

Kiểm định BG:

Breusch-Godfrey test for autocorrelation up to order 3
OLS, using observations 1959-2002 (T = 44)
Dependent variable: uhat

	coefficient	std. error	t-ratio	p-value
const	0.457595	0.808343	0.5661	0.5746
Product	-0.00638620	0.00948377	-0.6734	0.5047
uhat_1	0.954391	0.172746	5.525	2.37e-06 ***
uhat_2	-0.117923	0.234597	-0.5027	0.6180
uhat_3	0.0327493	0.173242	0.1890	0.8510

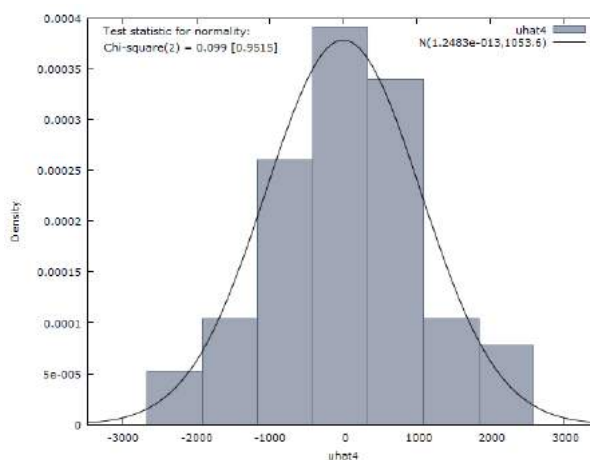
Unadjusted R-squared = 0.716573

Test statistic: LMF = 32.867206,
with p-value = $P(F(3,39) > 32.8672) = 9.12e-011$

Alternative statistic: $TR^2 = 31.529216$,
with p-value = $P(\text{Chi-square}(3) > 31.5292) = 6.58e-007$

Ljung-Box Q' = 69.5761,
with p-value = $P(\text{Chi-square}(3) > 69.5761) = 5.26e-015$

Đồ thị phân phối của phần dư:



Frequency distribution for uhat7, obs 1-51
number of bins = 7, mean = 1.24833e-013, sd = 1053.58

interval	midpt	frequency	rel.	cum.
< -1917.0	-2292.7	2	3.92%	3.92% *
-1917.0 - -1165.8	-1541.4	4	7.84%	11.76% **
-1165.8 - -414.62	-790.23	10	19.61%	31.37% *****
-414.62 - 336.60	-39.012	15	29.41%	60.78% *****
336.60 - 1087.8	712.20	13	25.49%	86.27% *****
1087.8 - 1839.0	1463.4	4	7.84%	94.12% **
>= 1839.0	2214.6	3	5.88%	100.00% **

Test for null hypothesis of normal distribution:
Chi-square(2) = 0.099 with p-value 0.95154

Kiểm định phân phối chuẩn của nhiễu:

Test for normality of e_i :

Doornik-Hansen test = 0.0993532, with p-value 0.951537
Shapiro-Wilk W = 0.989353, with p-value 0.925577
Lilliefors test = 0.0712944, with p-value \approx 0.74
Jarque-Bera test = 0.256649, with p-value 0.879568

Kiểm định Ramsey RESET:

RESET test for specification (squares and cubes)
Test statistic: $F = 11.506702$,
with p-value = $P(F(2, 45) > 11.5067) = 9.2e-005$

RESET test for specification (squares only)
Test statistic: $F = 22.274305$,
with p-value = $P(F(1, 46) > 22.2743) = 2.24e-005$

RESET test for specification (cubes only)
Test statistic: $F = 23.455683$,
with p-value = $P(F(1, 46) > 23.4557) = 1.49e-005$

Hoặc:

Auxiliary regression for RESET specification test
OLS, using observations 1-51
Dependent variable: crime

	coefficient	std. error	t-ratio	p-value	
const	5601.15	1463.32	3.828	0.0004	***
pov	-253.816	342.524	-0.7410	0.4625	
metro	-58.7475	128.053	-0.4588	0.6486	
popdens	-4.63186	2.00044	-2.315	0.0252	**
yhat ²	0.000141071	0.000666575	0.2116	0.8333	
yhat ³	3.11193e-08	3.42774e-08	0.9079	0.3688	

Test statistic: $F = 11.506702$,
with p-value = $P(F(2, 45) > 11.5067) = 9.2e-005$

EIEWS

Hồi quy OLS:

Dependent Variable: Q
 Method: Least Squares
 Date: 10/05/14 Time: 12:44
 Sample: 1 12
 Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24463.28	5619.777	4.353070	0.0018
DC	84.21205	58.81760	1.431749	0.1860
P	-3.226614	5.698364	-0.566235	0.5851
R-squared	0.969258	Mean dependent var	113916.7	
Adjusted R-squared	0.962426	S.D. dependent var	32329.15	
S.E. of regression	6266.693	Akaike info criterion	20.53620	
Sum squared resid	3.53E+08	Schwarz criterion	20.65743	
Log likelihood	-120.2172	Hannan-Quinn criter.	20.49132	
F-statistic	141.8776	Durbin-Watson stat	2.099322	
Prob(F-statistic)	0.000000			

Kiểm định White:

Heteroskedasticity Test: White

F-statistic	0.863252	Prob. F(5,6)	0.5546
Obs*R-squared	5.020728	Prob. Chi-Square(5)	0.4134
Scaled explained SS	1.908834	Prob. Chi-Square(5)	0.8616

Test Equation:
 Dependent Variable: RESID^2
 Method: Least Squares
 Date: 10/05/14 Time: 12:45
 Sample: 1 12
 Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.43E+08	1.50E+08	1.619166	0.1565
DC	-1725132.	1783504.	-0.967272	0.3708
DC^2	-22073.69	12183.10	-1.811828	0.1200
DC*P	4414.439	2419.654	1.824409	0.1179
P	144669.1	161146.3	0.897751	0.4039
P^2	-219.9212	120.0447	-1.831994	0.1167
R-squared	0.418394	Mean dependent var	29453577	
Adjusted R-squared	-0.066278	S.D. dependent var	35767355	
S.E. of regression	36933629	Akaike info criterion	37.99400	
Sum squared resid	8.18E+15	Schwarz criterion	38.23645	
Log likelihood	-221.9640	Hannan-Quinn criter.	37.90423	
F-statistic	0.863252	Durbin-Watson stat	1.998998	
Prob(F-statistic)	0.554559			

Kiểm định Glejser:

Heteroskedasticity Test: Glejser

F-statistic	0.100982	Prob. F(2,9)	0.9050
Obs*R-squared	0.263375	Prob. Chi-Square(2)	0.8766
Scaled explained SS	0.189510	Prob. Chi-Square(2)	0.9096

Test Equation:

Dependent Variable: ARESID

Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5377.232	3281.553	1.638624	0.1357
DC	9.070092	34.34533	0.264085	0.7977
P	-0.939821	3.327442	-0.282445	0.7840

R-squared	0.021948	Mean dependent var	4380.105
Adjusted R-squared	-0.195397	S.D. dependent var	3346.900
S.E. of regression	3659.306	Akaike info criterion	19.46025
Sum squared resid	1.21E+08	Schwarz criterion	19.58148
Log likelihood	-113.7615	Hannan-Quinn criter.	19.41537
F-statistic	0.100982	Durbin-Watson stat	2.334082
Prob(F-statistic)	0.904959		

Kiểm định BP

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.204107	Prob. F(2,9)	0.8190
Obs*R-squared	0.520668	Prob. Chi-Square(2)	0.7708
Scaled explained SS	0.197953	Prob. Chi-Square(2)	0.9058

Test Equation:

Dependent Variable: RESID^2

Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	46711397	34682493	1.346829	0.2110
DC	102457.6	362993.2	0.282258	0.7841
P	-10950.34	35167.49	-0.311377	0.7626

R-squared	0.043389	Mean dependent var	29453577
Adjusted R-squared	-0.169191	S.D. dependent var	35767355
S.E. of regression	38674935	Akaike info criterion	37.99160
Sum squared resid	1.35E+16	Schwarz criterion	38.11283
Log likelihood	-224.9496	Hannan-Quinn criter.	37.94672
F-statistic	0.204107	Durbin-Watson stat	1.972476
Prob(F-statistic)	0.819048		

Kiểm định BG:

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.824215	Prob. F(1,8)	0.2138
Obs*R-squared	2.228227	Prob. Chi-Square(1)	0.1355

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 10/05/14 Time: 12:46

Sample: 1 12

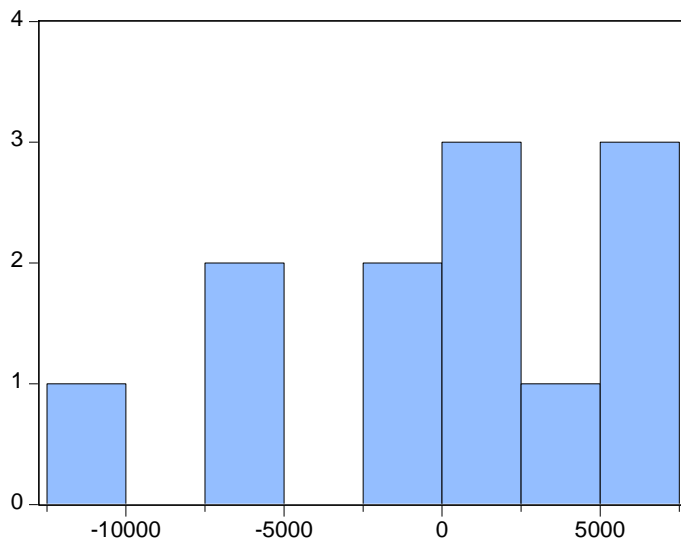
Included observations: 12

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1095.874	5439.727	-0.201457	0.8454
DC	-118.3904	104.1765	-1.136441	0.2887
P	11.59787	10.17268	1.140101	0.2872
RESID(-1)	-0.816408	0.604462	-1.350635	0.2138

R-squared	0.185686	Mean dependent var	6.65E-12
Adjusted R-squared	-0.119682	S.D. dependent var	5668.437
S.E. of regression	5998.059	Akaike info criterion	20.49746
Sum squared resid	2.88E+08	Schwarz criterion	20.65910
Log likelihood	-118.9848	Hannan-Quinn criter.	20.43762
F-statistic	0.608072	Durbin-Watson stat	2.060299
Prob(F-statistic)	0.628254		

Đồ thị phân phối của phần dư và kiểm định Jarque – Bera:



Series: Residuals	
Sample 1 12	
Observations 12	
Mean	6.65e-12
Median	1166.688
Maximum	6995.824
Minimum	-11118.81
Std. Dev.	5668.437
Skewness	-0.604646
Kurtosis	2.351789
Jarque-Bera	0.941281
Probability	0.624602