1. Mô hình hồi quy tuyến tính

Xem xét sự phụ thuộc của *Y* (***biến phụ thuộc***) vào các ***biến độc lập*** *X2*, *X3*,…, *Xk*, ta có

Hàm hồi quy ***tổng thể***

$$E(Y/X2,X3,...Xk) = \beta_1 + \beta_2 X2 + ... + \beta_k Xk$$

Mô hình hồi quy ***tổng thể***

$$Y = \beta_1 + \beta_2 X2 + ... + \beta_k Xk + U$$

Sử dụng thông tin từ mẫu ta xây dựng được

Hàm hồi quy ***mẫu***

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X2 + ... + \hat{\beta}_k Xk$$

Mô hình hồi quy ***mẫu***

$$Y = \hat{\beta}_1 + \hat{\beta}_2 X2 + ... + \hat{\beta}_k Xk + e$$

$\beta_j \, (j = \overline{1,k})$ gọi là các ***hệ số hồi quy***

$\hat{\beta}_j \, (j = \overline{1,k})$ là ***ước lượng điểm của các hệ số hồi quy***

U : ***sai số ngẫu nhiên*** (sai số giữa giá trị cá biệt của *Y* và giá trị trung bình của nó E(*Y/X2,X3,...Xk*) trong ***tổng thể***)

e : ***phần dư*** (residual – sai số giữa giá trị cá biệt/thực tế của Y và giá trị ước lượng của nó trong hồi quy $\hat{Y}$ trong ***mẫu*** quan sát)

(+) Ý nghĩa của các hệ số:

β_1 là hệ số chặn, ***nó là giá trị trung bình của biến phụ thuộc khi các biến độc lập trong mô hình nhận giá trị bằng 0.***

$\beta_j \, (j = \overline{2,k})$ là các hệ số hồi quy riêng (các hệ số góc). Nó phản ánh tác động của biến độc lập X_j tới biến phụ thuộc *Y*. ***Nếu các yếu tố khác không đổi, X_j tăng 1 đơn vị thì trung bình của Y sẽ tăng là β_j đơn vị và ngược lại (điều kiện các yếu tố khác không đổi)***.

(+) Dấu của β_j sẽ thể hiện chiều của mối quan hệ

$\beta_j > 0$: X_j tăng làm *Y* tăng và ngược lại (ảnh hưởng cùng chiều)

$\beta_j < 0$: X_j tăng làm *Y* giảm và ngược lại (ảnh hưởng ngược chiều)

$\beta_j = 0$: X_j thay đổi không làm *Y* thay đổi (*Y* không có quan hệ phụ thuộc tuyến tính vào X_j)

(+) Để ước lượng 1 hồi quy mẫu tuyến tính với 1 mẫu quan sát cụ thể, phương pháp được sử dụng phổ biến nhất hiện nay là phương pháp bình phương nhỏ nhất OLS với tiêu chuẩn ước lượng:

$$\sum_{i=1}^{n} e_i^2 \rightarrow \min$$

Giá trị này được gọi là Tổng bình phương phần dư (***Residual Sum of Squares – RSS*** hoặc ***Sum squared residual***)

Báo cáo OLS do phần mềm EVIEWS cung cấp:
Mô hình hồi quy tuyến tính:

$$Y = \beta_1 + \beta_2 K + \beta_3 L + U$$

Dependent Variable: Y *(Biến phụ thuộc là Y)*
Method: Least Squares *(Phương pháp bình phương nhỏ nhất OLS)*
Date: 12/19/12 Time: 09:11
Sample: 1 20 *(Kích thước mẫu: 20 quan sát)*
Included observations: 20 *(Số quan sát bao gồm: 20)*

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C (β_1)	$\hat{\beta}_1$= -21717.59	S.E($\hat{\beta}_1$) = 22180.83	$\frac{\hat{\beta}_1}{\text{S.E}(\hat{\beta}_1)}$ = -0.979116	0.3413
K (β_2)	$\hat{\beta}_2$=10751.92	S.E($\hat{\beta}_2$) = 2165.515	$\frac{\hat{\beta}_2}{\text{S.E}(\hat{\beta}_2)}$ = 4.965061	0.0001
L (β_3)	$\hat{\beta}_3$=17662.45	S.E($\hat{\beta}_3$) = 4533.201	$\frac{\hat{\beta}_3}{\text{S.E}(\hat{\beta}_3)}$ = 3.896242	0.0012

R-squared	R^2 = 0.715471	Mean dependent var	109468.7
Adjusted R-squared	$\bar{R}^2$ = 0.681997	S.D. dependent var	57734.42
S.E. of regression	32557.46	Akaike info criterion	23.75688
Sum squared resid. *(Tổng bình phương phần dư)*	*1.80E+10*	Schwarz criterion	23.90624
Log likelihood	-234.5688	F-statistic	*21.37391*
Durbin-Watson stat	*2.289076*	Prob(F-statistic)	*0.000023*

Breusch-Godfrey Serial Correlation LM Test: *(Kiểm tra hiện tượng tự tương quan)*

F-statistic	F_{qs} = 0.656872	Probability	0.429557
Obs*R-squared	χ^2_{qs} = 0.788709	Probability	0.374491

Ramsey RESET Test: *(Kiểm tra dạng hàm sai)*

F-statistic	F_{qs} = 0.160628	Probability	0.693880
Log likelihood ratio *(Không sử dụng)*	0.199784	Probability	0.654895

White Heteroskedasticity Test: cross terms *(Kiểm tra phương sai sai số thay đổi (có hệ số chéo))*

F-statistic	F_{qs} = 5.228787	Probability	0.006478
Obs*R-squared	χ^2_{qs} = 13.02510	Probability	0.023145

White Heteroskedasticity Test: no cross terms *(Kiểm tra phương sai sai số thay đổi (không có hệ số chéo))*

F-statistic	F_{qs} = 7.001717	Probability	0.002182
Obs*R-squared	χ^2_{qs} = 13.02437	Probability	0.011157

Trong báo cáo trên thì số hệ số của hồi quy là $k = 3$: β_1, β_2 và β_3

Mô hình hồi quy tuyến tính với các biến logarith:

$$\ln(Y) = \beta_1 + \beta_2 \ln(K) + \beta_3 \ln(L) + U$$

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 12/19/12 Time: 11:50
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.770251	0.228568	42.74543	0.0000
LOG(K)	0.523699	0.093755	5.585820	0.0000
LOG(L)	0.693005	0.140540	4.931025	0.0001
R-squared	0.781422	Mean dependent var		11.45945
Adjusted R-squared	0.755707	S.D. dependent var		0.570617
S.E. of regression	0.282033	Akaike info criterion		0.443897
Sum squared resid	1.352226	Schwarz criterion		0.593257
Log likelihood	-1.438970	F-statistic		30.38777
Durbin-Watson stat	1.833099	Prob(F-statistic)		0.000002

(+) $\beta_j\,(j=\overline{2,k})$ vẫn là các hệ số hồi quy riêng (các hệ số góc). Trong dạng hàm này, nó phản ánh tác động tương đối của biến độc lập X_j tới biến phụ thuộc Y. ***Nếu các yếu tố khác không đổi, X_j tăng 1 % thì trung bình của Y sẽ tăng là β_j % và ngược lại (điều kiện các yếu tố khác không đổi)***. Trong kinh tế học thì các hệ số góc của dạng hàm hồi quy này được gọi là hệ số co dãn của biến phụ thuộc Y theo biến độc lập X_j

(+) Dấu của β_j sẽ thể hiện chiều của mối quan hệ

$\beta_j > 0$: X_j tăng làm Y tăng và ngược lại (ảnh hưởng cùng chiều)

$\beta_j < 0$: X_j tăng làm Y giảm và ngược lại (ảnh hưởng ngược chiều)

$\beta_j = 0$: X_j thay đổi không làm Y thay đổi (Y không có quan hệ phụ thuộc tuyến tính vào X_j)

(+) Theo kết quả hồi quy ta có $\hat{\beta}_2 = 0.523699$ cho biết khi biến vốn (K) tăng 1% thì biến sản lượng (Y) tăng 0.523699% và ngược lại (trong điều kiện các yếu tố khác không đổi)

Tương tự, $\hat{\beta}_3 = 0.693005$ cho biết khi biến lao động (L) tăng 1% thì biến sản lượng (Y) tăng 0.693005% và ngược lại (trong điều kiện các yếu tố khác không đổi)

(+) Các câu hỏi phân tích hồi quy với dạng hàm này chỉ khác với dạng hàm tuyến tính thông thường ở đơn vị của các biến.

Ví dụ: Trong dạng hàm tuyến tính thông thường, nếu hỏi X (biến độc lập) tăng 1 đơn vị thì Y (biến phụ thuộc) tăng 2 đơn vị, nhận xét ý kiến này → cần kiểm định cặp giả thuyết:

H_0: $\beta_2 = 2$ (tương đương với nhận xét ý kiến đầu bài là đúng)

H_0: $\beta_2 \neq 2$ (tương đương với nhận xét ý kiến đầu bài là sai)

Còn trong dạng hàm tuyến tính với các biến dưới dạng loga Nepe này thì cách hỏi sẽ thay đổi → hỏi X (biến độc lập) tăng 1 % thì Y (biến phụ thuộc) tăng 2 %, nhận xét ý kiến này → ta vẫn cần kiểm định cặp giả thuyết:

H_0: $\beta_2 = 2$ (tương đương với nhận xét ý kiến đầu bài là đúng)

H_0: $\beta_2 \neq 2$ (tương đương với nhận xét ý kiến đầu bài là sai)

2. Công thức khoảng tin cậy của các hệ số hồi quy

(+) Với độ tin cậy (1 - α) cho trước, khoảng tin cậy của các hệ số β_j

KTC đối xứng : $\hat{\beta}_j - SE(\hat{\beta}_j)t_{\alpha/2}(n-k) < \beta_j < \hat{\beta}_j + SE(\hat{\beta}_j)t_{\alpha/2}(n-k)$

KTC bên phải : $\hat{\beta}_j - SE(\hat{\beta}_j)t_{\alpha}(n-k) < \beta_j$ *(k là số hệ số của mô hình)*

KTC bên trái : $\beta_j < \hat{\beta}_j + SE(\hat{\beta}_j)t_{\alpha}(n-k)$

<u>Chú ý cách sử dụng:</u>

- Nếu hỏi lượng thay đổi trung bình của biến phụ thuộc nằm trong khoảng nào (khi biến độc lập thay đổi) ta sử dụng khoảng tin cậy đối xứng.
- Khi mối quan hệ xem xét là thuận chiều ($\beta_j > 0$), nếu hỏi lượng thay đổi tối đa của biến phụ thuộc thì dùng KTC tối đa, và ngược lại.
- Khi mối quan hệ là ngược chiều ($\beta_j < 0$), nếu hỏi lượng thay đổi tối đa của biến phụ thuộc ta sử dụng KTC tối thiểu và ngược lại. Sau đó đổi dấu giá trị tìm được để có kết quả cuối cùng.

(+) Với độ tin cậy (1 - α) cho trước, khoảng tin cậy của **a.β_j + b.β_s**

KTC đối xứng : $a.\hat{\beta}_j + b.\hat{\beta}_s - Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t^{(n-k)}_{\alpha/2} < \beta_j < a.\hat{\beta}_j + b.\hat{\beta}_s + Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t^{(n-k)}_{\alpha/2}$

KTC bên phải : $a.\hat{\beta}_j + b.\hat{\beta}_s - Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t^{(n-k)}_{\alpha} < \beta_j < +\infty$ *(k là số hệ số của mô hình)*

KTC bên trái : $-\infty < \beta_j < a.\hat{\beta}_j + b.\hat{\beta}_s + Se(a.\hat{\beta}_j + b.\hat{\beta}_s).t^{(n-k)}_{\alpha}$

Trong đó:

$$Se(a.\hat{\beta}_j + b.\hat{\beta}_s) = \sqrt{a^2.[Se(\hat{\beta}_j)]^2 + b^2.[Se(\hat{\beta}_s)]^2 + 2a.b.\text{cov}(\hat{\beta}_j, \hat{\beta}_s)}$$

3. Quy tắc kiểm định giả thuyết đối với các hệ số hồi quy

a1. Cặp giả thuyết 1 $\begin{cases} H_0 : \beta_j = \beta_j^* \\ H_1 : \beta_j \neq \beta_j^* \end{cases}$ Tiêu chuẩn kiểm định : $T = \dfrac{\hat{\beta}_j - \beta_j^*}{Se(\hat{\beta}_j)}$

Với kết quả ước lượng, ta có:

$$T_{qs} = \frac{\hat{\beta}_j - \beta_j^*}{Se(\hat{\beta}_j)}$$

Với α cho trước, miền bác bỏ H_0:

$$W_\alpha = \left\{ T : |T| > t^{(n-k)}_{\alpha/2} \right\}$$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

b1. Cặp giả thuyết 2 $\begin{cases} \mathbf{H}_0 : \beta_j = \beta_j^* \\ \mathbf{H}_1 : \beta_j > \beta_j^* \end{cases}$

Với α cho trước, miền bác bỏ H_0:

$W_\alpha = \{T : T > t_\alpha^{(n-k)}\}$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

c1. Cặp giả thuyết 3 $\begin{cases} \mathbf{H}_0 : \beta_j = \beta_j^* \\ \mathbf{H}_1 : \beta_j < \beta_j^* \end{cases}$

Với α cho trước, miền bác bỏ H_0:

$W_\alpha = \{T : T < -t_\alpha^{(n-k)}\}$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

(+) Trường hợp đặc biệt khi $\beta_j^* = 0$ $\rightarrow$ $T_{qs} = \dfrac{\hat{\beta}_j}{Se(\hat{\beta}_j)}$ = ***T- Statistic***

Khi hỏi X (biến độc lập) tăng có làm Y (biến phụ thuộc) thay đổi hay không → cần kiểm định cặp giả thuyết:

$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$

Khi hỏi X (biến độc lập) tăng (giảm) có làm Y (biến phụ thuộc) tăng (giảm) hay không → cần kiểm định cặp giả thuyết:

$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j > 0 \end{cases}$

Khi hỏi X (biến độc lập) tăng (giảm) có làm Y (biến phụ thuộc) giảm (tăng) hay không → cần kiểm định cặp giả thuyết:

$\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j < 0 \end{cases}$

(+) Khi kiểm định cặp giả thuyết $\begin{cases} H_0 : \beta_j = 0 \\ H_1 : \beta_j \neq 0 \end{cases}$ có thể sử dụng quy tắc p-value (***Prob - Probability***) như sau :

Nếu p-value = hoặc < α → bác bỏ H_0

Nếu p-value > α → chấp nhận H_0

(+) Kiểm định biểu thức giữa các hệ số hồi quy:

a2. Cặp giả thuyết 1 $\begin{cases} H_0 : a.\beta_j + b.\beta_s = a^* \\ H_1 : a.\beta_j + b.\beta_s \neq a^* \end{cases}$ Tiêu chuẩn kiểm định : $T = \dfrac{a.\hat{\beta}_j + b.\hat{\beta}_s - a^*}{Se(a.\hat{\beta}_j + b.\hat{\beta}_s)}$

Với kết quả ước lượng, ta có:

$$T_{qs} = \frac{a.\hat{\beta}_j + b.\hat{\beta}_s - a^*}{Se(a.\hat{\beta}_j + b.\hat{\beta}_s)}$$

Với α cho trước, miền bác bỏ H_0:

$$W_\alpha = \left\{T : |T| > t^{(n-k)}_{\alpha/2}\right\}$$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

b2. Cặp giả thuyết 2 $\begin{cases} H_0 : a.\beta_j + b.\beta_s = a^* \\ H_1 : a.\beta_j + b.\beta_s > a^* \end{cases}$

Với α cho trước, miền bác bỏ H_0:

$$W_\alpha = \left\{T : T > t^{(n-k)}_{\alpha}\right\}$$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

c2. Cặp giả thuyết 3 $\begin{cases} H_0 : a.\beta_j + b.\beta_s = a^* \\ H_1 : a.\beta_j + b.\beta_s < a^* \end{cases}$

Với α cho trước, miền bác bỏ H_0:

$$W_\alpha = \left\{T : T < -t^{(n-k)}_{\alpha}\right\}$$

Nếu $T_{qs} \in W_\alpha$ thì bác bỏ H_0

Nếu ngược lại : chấp nhận H_0.

4. Hệ số xác định của mô hình và kiểm định giả thuyết về sự phù hợp của hàm hồi quy

- Hệ số xác định $R^2 = \dfrac{ESS}{TSS} = 1 - \dfrac{RSS}{TSS}$ = **R – Squared** $\rightarrow$ ***Cho biết tỉ lệ sự biến động của biến phụ thuộc được giải thích bởi sự biến động của tất cả các biến độc lập (biến giải thích) có trong mô hình.***

RSS = Residual Sum of Squares

TSS = (n-1)*(S.D. Dependent Variable)^2

- Hệ số xác định đã hiệu chỉnh $\overline{R^2} = 1 - (1 - R^2)\dfrac{n-1}{n-k}$ = ***Adjusted -R - Squared*** Hệ $\rightarrow$ cách tính R^2 như sau:

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

Hệ số $\overline{R^2}$ còn được sử dụng để đánh giá việc đưa thêm 1 biến độc lập mới vào mô hình có cần thiết hay không. So sánh hệ số này của mô hình đã thêm biến và mô hình chưa thêm biến mới, nếu $\overline{R^2}$ tăng lên khi đưa thêm biến thì biến độc lập mới là cần thiết cho mô hình và ngược lại.

- Kiểm định sự phù hợp của hàm hồi quy

Cặp giả thuyết $\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 \neq 0 \end{cases} \Leftrightarrow \begin{cases} H_0 : \beta_2 = ... = \beta_k = 0 \\ H_1 : \exists \beta_j \neq 0 : (j \neq 1) \end{cases}$

H_0 : Hàm hồi quy không phù hợp (tất cả các biến độc lập cùng không tác động tới biến phụ thuộc)

H_1 : Hàm hồi quy phù hợp (có ít nhất một biến độc lập có giải thích cho biến phụ thuộc)

Kiểm định F: $$F_{qs} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{R^2}{1-R^2} \times \frac{n-k}{k-1} = \boldsymbol{F - Statistic}$$

- Nếu $F_{qs} > F_{\alpha}(k - 1; n - k)$ thì bác bỏ H_0 : hàm hồi qui là phù hợp.
- Ngược lại, hàm hồi qui không phù hợp.

Có thể sử dụng mức xác suất (p-value) đã được phần mềm tính ra để thực hiện kiểm định cặp giả thuyết trên theo quy tắc: Prob (F-Statistic) < α $\rightarrow$ Bác bỏ H_0

Prob > α $\rightarrow$ chấp nhận H_0

- Chú ý: Có thể từ công thức kiểm định trên $\rightarrow$ cách tính R^2

$$\boxed{R^2 = \frac{1}{1 + \frac{1}{F - statistic} \times \frac{n-k}{k-1}}}$$

5. Kiểm định thu hẹp hồi quy (kiểm định thêm biến hay bớt biến bằng kiểm định F) (Kiểm định nhiều điều kiện ràng buộc với các hệ số hồi quy)

$E(Y/X_2,..,X_{k-m},..,X_k) = \beta_1 + \beta_2 X_2 + \ldots + \beta_{k-m} X_{k-m} + \ldots + \beta_k X_k$ (UR)

$E(Y/X_2,\ldots, X_{k-m}) = \beta_1 + \beta_2 X_2 + \ldots + \beta_{k-m} X_{k-m}$ (R)

$$\begin{cases} H_0 : \beta_{k-m+1} = \beta_{k-m+2} \ldots = \beta_k = 0 & \textit{(Có thể bỏ m biến…ra khỏi mô hình (UR))} \\ H_1 : \exists \beta_j \neq 0 : (j = k-m+1 \div k) & \textit{(Không thể bỏ…………….)} \end{cases}$$

$\Leftrightarrow$ *Không cần đưa thêm m biến ….vào mô hình* (R)

*Nên đưa thêm m biến …… vào mô hình (*R*)*

$$F_{qs} = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k)} = \frac{R^2_{UR} - R^2_R}{1-R^2_{UR}} \times \frac{n-k}{m} = \frac{RSS_R - RSS_{UR}}{RSS_{UR}} \times \frac{n-k}{m}$$

Trong đó:

m – số điều kiện ràng buộc

k – số hệ số hồi quy của mô hình (UR)

n – số quan sát

Nếu $F_{qs} > F_{\alpha}(m, n - k)$ $\rightarrow$ bác bỏ H_0 và ngược lại.

6. Các mô hình có chứa biến giả:

Biến giả D1 = $\begin{cases} 1 & A_1 \\ 0 & A_2 \end{cases}$

(+) Mô hình có biến độc lập là biến giả

$$PRM: \quad Y_i = \beta_1 + \beta_2 X_i + \beta_3 D1_i + u_i$$

(A_1) hoặc $(D1_i = 1)$: $Y_i = (\beta_1 + \beta_3) + \beta_2 X_i + u_i$

(A_2) hoặc $(D1_i = 0)$: $Y_i = \beta_1 + \beta_2 X_i + u_i$

(+) Mô hình có biến tương tác giữa biến độc lập và biến giả

PRM: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 (X_i * D1_i) + u_i$

(A_1) hoặc $(D1_i = 1)$: $Y_i = \beta_1 + (\beta_2 + \beta_3).X_i + u_i$

(A_2) hoặc $(D1_i = 0)$: $Y_i = \beta_1 + \beta_2 X_i + u_i$

(+) Mô hình có cả biến giả và biến tương tác

PRM: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 D1_i + \beta_4 (X_i * D1_i) + u_i$

(A_1) hoặc $(D1_i = 1)$: $Y_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4).X_i + u_i$

(A_2) hoặc $(D1_i = 0)$: $Y_i = \beta_1 + \beta_2 X_i + u_i$

7. Phương sai sai số ngẫu nhiên thay đổi

- ***Giả thiết OLS***: Phương sai các yếu tố ngẫu nhiên là đồng nhất : Var(U_i) = σ^2 không đổi.
- Giả thiết không thỏa mãn: Var(U_i) = σ_i^2 không đồng nhất → PSSS thay đổi

Kiểm định WHITE: thường dùng cho hồi quy nhiều biến

Mô hình gốc: $Y = \beta_1 + \beta_2 X2 + \beta_3 X3 + U$

Bước 1: Hồi qui mô hình gốc thu được phần dư e_i

Bước 2: Tạo biến e_i^2, $X2_i^2$, $X3_i^2$, $(X2_i \times X3_i)$

Hồi qui mô hình hồi qui phụ:

(2) $e_i^2 = \alpha_1 + \alpha_2 X2_i + \alpha_3 X2_i^2 + \alpha_4 X3_i + \alpha_5 X3_i^2 + V_i$ (no cross terms)

(3) $e_i^2 = \alpha_1 + \alpha_2 X2_i + \alpha_3 X2_i^2 + \alpha_4 (X2_i \times X3_i) + \alpha_5 X3_i + \alpha_6 X3_i^2 + V_i$ (cross terms)

(**i**) được các hệ số xác định R_2^2 và R_3^2 (kí hiệu là R_i^2)

m là số hệ số của mô hình (**i**)

Bước 3: Kiểm định cặp giả thuyết

$$\begin{cases} H_0 : R_i^2 = 0 \\ H_1 : R_i^2 \neq 0 \end{cases}$$

H_0 : Mô hình ban đầu có phương sai của sai số đồng đều

H_1 : Mô hình ban đầu có phương sai của sai số thay đổi

Kiểm định F, χ^2

Kiểm định χ^2 : $\chi_{qs}^2 = nR_i^2$ = ***Obs*R-squared*** (White test) nếu $\chi_{qs}^2 > \chi_\alpha^2(m-1)$ thì bác bỏ H_0

Kiểm định F: $F_{qs} = \boxed{\dfrac{R_i^2/(m-1)}{(1-R_i^2)/(n-m)} = \dfrac{R_i^2}{1-R_i^2} \times \dfrac{n-m}{m-1}}$ = ***F-statistic*** (White test)

nếu $F_{qs} > F_\alpha(m-1, n-m)$ thì bác bỏ H_0.

Có thể sử dụng mức xác suất đã được máy tính tính ra trong kiểm định White để thực hiện kiểm định cặp giả thuyết theo quy tắc: Prob < α → Bác bỏ H_0

Prob > α → Chưa bác bỏ H_0

Chú ý: Nếu mô hình ban đầu chỉ có 1 biến độc lập thì không phân biệt kiểm định có hệ số chéo hay không và hồi quy phụ trong cả 2 trường hợp kiểm định đều là:

$$e_i^2 = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + V_i$$

7. Tự tương quan

- MH ban đầu: $Y_t = \beta_1 + \beta_2 X_t + U_t$

Giả thiết OLS : Các yếu tố ngẫu nhiên không tương quan

Nếu giả thiết bị vi phạm : hiện tượng TTQ bậc ρ

- Xét trường hợp $\rho = 1$ → lược đồ tự tương quan bậc 1 – AR(1)

$$u_t = \rho u_{t-1} + \varepsilon_t$$

với $-1 \le \rho \le 1$ và ε_t thỏa mãn các giả thiết của OLS

$-1 < \rho < 0$ tự tương quan âm

$\rho = 0$ không có tự tương quan

$0 < \rho < 1$ tự tương quan dương

$$\rho = \frac{\sum u_t u_{t-1}}{\sum u_t^2}$$

- **Kiểm định Durbin – Watson** *(Dùng để kiểm định tự tương quan bậc 1)*

Trong thực tế ta dùng ước lượng $\hat{\rho}$ để thay thế ρ khi quan sát hiện tượng tự tương quan

$$\hat{\rho} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

Thống kê Durbin Watson được tính theo công thức:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} \approx 2(1-\hat{\rho})$$

Với $-1 \le \hat{\rho} \le 1 \rightarrow 0 \le d \le 4$

Với n, $k' = k - 1$ cho trước, tra bảng phụ lục 5 → d_L (giá trị cận dưới thống kê d) và d_U (giá trị cận trên thống kê d)

Tự tương quan dương $\rho > 0$	Không có kết luận	Không có tự tương quan $\rho = 0$	Không có kết luận	Tự tương quan âm $\rho < 0$

0 d_L d_U $4 - d_U$ $4 - d_L$ 4

<u>Chú ý:</u> Kiểm định DW sẽ ***không dùng được*** trong các trường hợp sau:

* khi mô hình ***không có hệ số chặn***

$$Y_t = \beta_2 X_t + \beta_3 Z_t + U_t$$

* ***có biến trễ của biến phụ thuộc đóng vai trò biến độc lập giải thích trong mô hình gốc***

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + U_t$$

- **Kiểm định Breusch - Godfrey**

Bước 1: Hồi quy mô hình ban đầu được e_t và e_{t-1}

Bước 2: Hồi quy phụ

(2) $e_t = \beta_1 + \beta_2 X_t + V_t$

(3) $e_t = \beta_1 + \beta_2 X_t + \beta_3 e_{t-1} + V_t$

Bước 3: Kiểm định cặp giả thuyết

H_0 *: Mô hình không có tự tương quan*

H_1 *: Mô hình có tự tương quan*

Kiểm định χ^2 : $\chi^2_{qs} = (n-1) \times R_3^2$ = ***Obs*R-squared*** (Breusch – Godfrey test) nếu $\chi^2_{qs} > \chi^2_\alpha(1)$ thì bác bỏ H_0 và ngược lại (trong phần mềm EVIEWS số quan sát được lấy đủ là *n* quan sát vì quan sát bị thiếu do biến trễ của phần dư gây ra được gán trị bằng 0)

Kiểm định F: F_{qs} = $\boxed{\frac{R_3^2 - R_2^2}{1 - R_3^2} \times \frac{n-k-1}{1}}$ = ***F-statistic*** (Breusch – Godfrey test) nếu F_{qs} > $F_\alpha(1,n\text{-}k\text{-}1)$ thì bác bỏ H_0 và ngược lại

Chú ý: *k* là số hệ số hồi quy của mô hình ban đầu và mô hình ban đầu có bao nhiêu biến độc lập đều được đưa vào trong hồi quy phụ (2) và (3). Dạng ban đầu của các biến độc lập cũng được giữ nguyên trong các hồi quy phụ này (nếu trong mô hình gốc là dạng $ln(X_i)$ thì trong các hồi quy phụ cũng là $ln(X_i)$)

Có thể sử dụng mức xác suất đã được máy tính tính ra trong kiểm định Breusch – Godfrey để kết luận về cặp giả thuyết theo quy tắc: Prob < α → Bác bỏ H_0

Prob > α → Chưa bác bỏ H_0

8. Phát hiện mô hình thiếu biến giải thích

Bước 1: Hồi quy mô hình ban đầu thu được e_t và $\hat{Y}_t$

Bước 2: Hồi quy phụ

(2) $Y_t = \beta_1 + \beta_2 X_t + \beta_3 \hat{Y}_t^2 + u_t$

(3) $e_t = \beta_1 + \beta_2 X_t + \beta_3 \hat{Y}_t^2 + u_t$

Bước 3: Kiểm định cặp giả thuyết

H_0: *Mô hình ban đầu không thiếu biến (mô hình có dạng hàm đúng)*
H_1 : *Mô hình ban đầu thiếu biến (mô hình có dạng hàm sai)*

Kiểm định χ^2 : $\chi^2_{qs} = nR^2_3$ nếu $\chi^2_{qs} > \chi^2_{\alpha}(1)$ thì bác bỏ H_0

Kiểm định F: $F_{qs} = \boxed{\dfrac{R^2_2 - R^2_1}{1 - R^2_2} \times \dfrac{n-k-1}{1}} = $ ***F - statistic*** (Ramsey Reset test) nếu Fqs > $F_{\alpha}(1,n\text{-}k\text{-}1)$ thì bác bỏ H_0.

Có thể sử dụng p-value để thực hiện kiểm định cặp giả thuyết

Prob < α $\rightarrow$ Bác bỏ H_0

Prob > α $\rightarrow$ Chưa bác bỏ H_0

Chú ý: *k* là số hệ số hồi quy của mô hình ban đầu và mô hình ban đầu có bao nhiêu biến độc lập đều được đưa vào trong hồi quy phụ (2) và (3). Dạng ban đầu của các biến độc lập cũng được giữ nguyên trong các hồi quy phụ này (nếu trong mô hình gốc là dạng $ln(X_i)$ thì trong các hồi quy phụ cũng là $ln(X_i)$)

9. Kiểm định về quy luật phân phối xác suất của yếu tố ngẫu nhiên (Kiểm định Jarque Bera)

H_0 : *Yếu tố ngẫu nhiên phân phối chuẩn*

H_1 : *Yếu tố ngẫu nhiên không phân phối chuẩn*

Kiểm định χ^2 : $\chi^2_{qs} = n\left[\dfrac{S^2}{6} + \dfrac{(K-3)^2}{24}\right]$ = Jarque – Bera

Với *S* là hệ số bất đối xứng, *K* là hệ số nhọn của phần dư e trong mô hình ban đầu

$$S = \frac{\sum_{i=1}^{n} e_i^3 \Big/ n}{\left(\sum_{i=1}^{n} e_i^2 \Big/ n\right)^{3/2}} \qquad K = \frac{\sum_{i=1}^{n} e_i^4 \Big/ n}{\left(\sum_{i=1}^{n} e_i^2 \Big/ n\right)^{2}}$$

Nếu $\chi^2_{qs} > \chi^2_{\alpha}(2)$ thì bác bỏ H_0, ngược lại chấp nhận H_0

Có thể sử dụng p-value để thực hiện kiểm định cặp giả thuyết.

Prob < α $\rightarrow$ Bác bỏ H_0,

Prob > α $\rightarrow$ Chưa bác bỏ H_0

So sánh 2 mô hình với nhau: Mô hình nào tốt hơn mô hình nào. Thứ tự ưu tiên so sánh như sau

Thứ nhất nên so sánh kiểm định Ram sey (nếu mô hình nào vi phạm sẽ không tốt bằng mô hình kia)

Thư hai : so sánh kđ White, DW,BG(Nếu cùng tuân thủ Ramsey thì so sánh tiếp 3 kđ trên)

Thứ ba: So sánh R^2

Tuy nhiên bài thường hỏi tốt hơn những điêm nào nên ta phải kđ hết các khuyết tật ra

Các câu hỏi về không chệch tốt nhất và hiệu quả, có đáng tin cậy

- Các ước lượng là tốt nhất: Đối với mh sử dụng số liệu chéo các ul là tốt nhất khi tuân thủ kđ White và Ramsey, Đối với mô hình sử dụng số liệu thời gian thì cần tuân thủ thêm kđ DW và BG

- Các ước lượng có đáng tin cậy không: Chỉ cần nói vi phạm 1 trong các kđ trên thì sẽ không tin cậy

- Các ước lượng không chệch (Xét kđ ramsey nếu tuân thủ thì là không chệch còn vi phạm thì là ul chệch mà đã chệch thì k xét đến hiệu quả). Khi tuân thủ kđ Ramsey với mh sử dụng số liệu chéo nếu tuân thủ kđ White thì sẽ là ul hiệu quả còn với mh sử dụng số liệu thời gian thì phải tuân thủ thêm các kđ DW và BG

Các bạn nên thuộc nguyên nhân hậu quả, cách phát hiện, 1 số khuyết tật nên học cách khắc phục vì bài có thể thi nhé.

TỔNG HỢP LẠI NHỮNG VẤN ĐỀ THẮC MẮC CỦA SINH TIÊN

(Sau buổi hệ thống sáng nay)

Băn khoăn của sinh viên	Gợi ý trả lời
Bình luận về việc bớt biến khỏi MH (câu hỏi lý thuyết) Ví dụ: Giả sử có MH Q=b1+b2K+b3L+b4L^2+u Đề bài hỏi đánh giá về ý kiến "Khi mục đích chỉ là đánh giá tác động của K lên Q thì chỉ cần hồi quy Q theo K mà không cần đưa biến L, L^2 vào mô hình"?	Theo thay em nen trinh bay 2 y sau: 1) Khang dinh y kien tren la chua co co so. Boi vi: Ngoai K ra co rat nhieu yeu to anh huong den Q trong do L mot trong cac yeu to quan trong thuc su co anh huong den Q (ly thuyet ve ham san xuat trong kinh te vi mo) 2) De bo bien L va L^2 ra khoi MH thi co the lam cho MH thieu bien quan trong vi: L co the tac dong den Q va L co the co tuong quan voi K. Do do de co can cu danh gia y kien tren ta tien hanh kiem dinh gia thuyet thong ke "Bo 2 bien L, L^2 ra khoi MH". Cac buoc kiem dinh: 1) HQ MH tren --> R1^2 2) Thu hep MH bang cach bo 2 bien ra khoi MH va HQ --> R2^2 3) Thiet lap Fqs = ...; tra bang co gia tri toi han F 4) Ket luan: + Neu... + Neu...
Ý nghĩa của các hs trong các MH có dạng hàm khác nhau	Có 4 dạng hàm của MH thường gặp 1) MH tuyến tính: Y = a + bX + u → E(Y/X) = a + bX b = đạo hàm của E(Y/X) theo X và có ý nghĩa là X tăng lên 1 **<u>đơn vị</u>** thì E(Y/X) (hiểu theo nghĩa là trung bình của Y) thay đổi b **<u>đơn vị</u>** 2) MH loga tuyến tính: lnY = a + blnX + u → E(lnY/lnX) = a + blnX (X, Y>0) b = hệ số co giãn của E(Y/X) theo X và có ý nghĩa là X tăng lên **<u>1 %</u>** thì trung bình của Y thay đổi **<u>b %</u>** 3) MH bán loga: lnY = a + bX + u → E(lnY/X) = a + bX (Y>0) ý nghĩa là X tăng lên 1 **<u>đơn vị</u>** thì trung bình của Y thay đổi **<u>b*100(%)</u>** 4) MH bán loga: Y = a + blnX + u → E(Y/lnX) = a + blnX (X>0) ý nghĩa là X tăng lên **<u>1 %</u>** thì trung bình của Y thay đổi **<u>b/100</u>**
Giải thích ý nghĩa của MH khi có biến bậc 2 Y = beta1 + beta2X + beta3X^2 + beta4Z + U E(Y/X, Z) = beta1 + beta2X + beta3X^2 + beta4Z	- Mục đích đưa biến bậc 2 của X vào là muốn xem xét Quy luật "lợi suất cận biên giảm dần" có tác động lên mối quan hệ của biến X và biến Y hay không - Giải thích ý nghĩa như sau: + Đạo hàm bậc 1 của E(Y) theo X = beta2 + 2beta3*X: cho biết khi X tăng lên 1 đơn vị thì trung bình của Y thay đổi (beta2 + 2beta3*X) đơn vị (giá trị thay đổi này tùy thuộc vào giá trị của biến X) + Đạo hàm bậc 2 của E(Y) theo X = 2beta3: Nó cho biết khi X tăng lên thì "giá trị cận biên của E(Y) theo X" thay đổi như thế nào Nếu mqh giữa y và X chịu sự chi phối của quy luật lợi suất cận biện giảm dần thì 2beta3<0 → beta 3 <0 - Ngoài ra nếu cho mô hình lnY = a + blnX + u → E(lnY/lnX) = a + blnX (X, Y>0) Nếu mqh giữa y và X chịu sự chi phối của quy luật lợi suất cận biện giảm dần thì 0<b3<1 → nếu phải KĐ thì tiến hành kiểm định 2 cặp giả thuyết: Cặp 1: H0: b = 0; H1: b>0 và cặp 2: H0:b=1; H1: b<1 (chú ý nếu phải tính toán cụ thể thì các beta và b được thay bằng các giá trị ước lượng điểm của nó lấy từ kết quả của mô hình)
Vấn đề về kỳ vọng của các hs hồi quy Y = beta1 + beta2X + beta3Z + U E(Y/X, Z) = beta1 + beta2X+beta3Z	- Sinh viên cần dựa vào lý thuyết kinh tế hoặc hiểu biết thực tế để đưa ra kỳ vọng về các hs hồi quy của MH. Ở đây là kỳ vọng dấu của các hs trong MH tổng thể (để sau này đối chiếu với các giá trị ước lượng được xem có phù hợp với kỳ vọng hay không). Một số trường hợp có thể kỳ vọng nhiều hơn là dấu. ví dụ mô hình: Chi tiêu = a + bThu nhập + U (rõ ràng bạn sẽ kỳ vọng 0 < b<1) - Ý nghĩa của beta1 = E(Y/X=Z=0): đây là ý nghĩa về mặt lý thuyết. Ý nghĩa thực tế tùy thuộc vào tình huống kinh tế (nội dung của các biến Y, X, Z) TC = beta1 + beta2Q + beta3Q^2+beta4Q^3 +U → Beta1 = E(TC/Q=0) = chi phí cố định trung bình Chi tiêu = beta1+beta2TN + U → beta1 = E(chi tiêu/TN = 0) = chi tiêu tối thiểu ……………….. Tuy nhiên: nói chung khi được hỏi ý nghĩa của các hs hồi quy thì nếu thấy beta1 không có ý nghĩa thực tế thì cũng không cần kỳ vọng và cũng không cần giải thích ý nghĩa của beta1^

Vấn đề về các khuyết tật Không rõ khi nào thì phải viết MH hồi quy phụ, cách làm thế nào?	1) Nếu câu hỏi: Kiểm định các khuyết tật của MH + VIết cặp giả thuyết + Đề xuất phương pháp Kiểm định (tương ứng với các khuyết tật) + Kết luận dựa vào kết quả 2) Nếu câu hỏi: Giá trị thống kê F-statistic hoặc Khi bình phương qs được tính từ như thế nào, có kết luận gì về MH B1: Nêu các bước để tính Fqs/Khi bình phương qs + Viết MH ban đầu (là MH cần kiểm định khuyết tật) + HQ MH này thu được giá trị phần dư/Y^ (tùy theo khuyết tật) + HQ mô hình phụ (tùy theo khuyết tật) + Viết công thức để tính Fqs/Khi bình phương qs B2: Dựa vào kết quả đã có → KĐ + VIết cặp giả thuyết + Đề xuất phương pháp Kiểm định (tương ứng với các khuyết tật) + Kết luận dựa vào kết quả
Câu hỏi "Các hs hồi quy ước lượng được có đáng tin cậy hay không?" Y = beta1 + beta2X + beta3Z + U E(Y/X, Z) = beta1 + beta2X+beta3Z	- Khái niệm "Đáng tin cậy" là một khái niệm cần phải xem xét kỹ dựa vào thông tin cho trong bảng kết quả (tránh hiện tượng suy diễn kiểu như "biết đâu vẫn có ĐCT, dạng hàm sai…." - Theo Định lý Gauss-Markov +MH với số liệu chéo: thỏa mãn 4 giả thiết (GT1 về mẫu ngẫu nhiên – đương nhiên thỏa mãn – không cần kiểm định; GT2: E(U)=0 – Kiểm định Ramsey; GT3: PSSS đồng đều → KĐ white; MH không có ĐCT hoàn hảo – Phần này thường được thỏa mãn vì bản thân MH đã ước lượng được thì tức là không có ĐCT hoàn hảo) + MH với số liệu theo thời gian thì có thêm GT về TTQ (KĐ DW với TTQ bậc 1 hoặc KĐ BG với TTQ bậc bất kỳ) NẾU CÓ ÍT NHẤT 1 GIẢ THIẾT TRONG CÁC GIẢ THIẾT TRÊN BỊ VI PHẠM VÌ CÓ THỂ COI CÁC KẾT QUẢ ƯỚC LƯỢNG ĐƯỢC LÀ KHÔNG ĐÁNG TIN CẬY - Nếu giả thiết về + PSSS không phân phối chuẩn + ĐCT không hoàn hảo ở mức độ cao ➔ Mặc dù không vi phạm giả thiết theo ĐL Gauss – Markov nhưng lại dẫn đến các hậu quả về việc các SUY DIỄN THỐNG KÊ (kIểm định và ƯL khoảng tin cậy không ĐÁNG TIN CẬY) KẾT LUẬN: Việc trả lời "ĐÁNG TIN CẬY" cần được phân biệt như 2 ý trên
Vấn đề với việc hiệu chỉnh MH bằng cách đưa biến giả vào MH - MH ban đầu Y = beta1 + beta2X + beta3Z + U E(Y/X, Z) = beta1 + beta2X+beta3Z - MH sau khi hiệu chỉnh Y = beta1 + beta2X + beta3Z + beta4D+beta5D*X+beta6D*Z+ U D = 0 với các quan sát có thuộc tình nào đó D=1 với các quan sát còn lại	- Khi được yêu cầu so sánh hoặc đánh giá tác động của Z và/hoặc Z đến Y là "khác nhau" khi D nhận giá trị khác nhau (quan sát có thuộc tính khác nhau) chúng thực hiện theo các bước: + Xác định chính xác nội dung và đặt biến giả + Đọc yêu cầu để biết đưa biến giả vào MH như thế nào MH1: nếu chỉ muốn xem Trung bình của Y có khác nhau khi D nhận giá trị khác nhau hay không: E(Y/X,Z,D) = beta1 + beta2X + beta3Z + beta4D MH2: Khi muốn đánh giá tác động của X → Y có khác nhau khi D nhận giá trị khác nhau hay không: E(Y/X,Z,D) = beta1 + beta2X + beta3Z + beta4D*X MH3: Khi muốn đánh giá tác động của Z → Y có khác nhau khi D nhận giá trị khác nhau hay không: E(Y/X,Z,D) = beta1 + beta2X + beta3Z + beta4D*Z ……….. Câu hỏi sẽ xoay quan: + Ý nghĩa của beta4 + Kiểm định liên quan đến hs beta4 trong MH
Mô hình với số liệu theo thời gian Y_t = beta1 + beta2X_t +beta3Z_t+ U_t	- MH có thể có biến trễ của biến độc lập Y_t = beta1 + beta2X_t +beta3Z_t+ beta3X_{t-1} +U_t - MH có thể có biến trễ của biến phụ thuộc Y_t = beta1 + beta2X_t +beta3Z_t+ beta3Y_{t-1} +U_t (MH này không thể sử dụng kiểm định TTQ bậc 1 bằng phương pháp DW vì có biến trễ của biến phụ thuộc với vai trò là biến độc lập) - Ý nghĩa hs gắn với biến trễ được giải thích một cách bình thường như các biến độc

	lập khác (ví dụ: giá ở năm trước; thu nhập ở tháng trước,….) - MH có thể có biến xu thế Y_t = beta1 + beta2X_t +beta3Z_t+ beta4T+ U_t Biến T nhận các giá trị = 1, 2, ….n tương ứng với các quan sát theo thứ tự Ý nghĩa của beta4 = khi T tăng lên 1 đơn vị (có thể là sau 1 thời kỳ - sau 1 năm, 1 quý, 1 tháng…) thì TB của Y tăng lên/giảm xuống beta4 đơn vị trong điều kiện các yếu tố khác không thay đổi - MH có thể có yếu tố mùa vụ: hãy xem xét sự thay dổi của Y khi chuyển từ mùa nóng → mùa lạnh (nhu cầu về bia, nước mía,….) + Đưa biến giả vào để đánh giá sự thay đổi này + Phân tích hs của biến giả
Vấn đề dự báo trong MH hồi quy	- Dự báo dựa trên ước lượng điểm + Cho giá trị của các biến độc lập + Dựa vào SRF: Y^ =beta^+…..--> tìm được Y^ (đây là ước lượng cho E(Y/X)- một giá trị dự báo) - Dự báo dựa trên ước lượng khoảng TC Y^-Se(Y^)T(n-k),a/2<= E(Y)<= Y^+Se(Y^)T(n-k),a/2 Vấn đề là phải có Se(Y^) (cái này bài cho)