CHƯƠNG 4: HỒI QUI ĐA BIẾN

Mô hình hồi quy đơn đã trình bày ở các chương 2 và 3 là khá hữu dụng cho rất nhiều trường hợp khác nhau. Mặc dù vậy, nó trở nên không còn phù hợp nữa khi có nhiều hơn một yếu tố tác động đến biến cần được giải thích. **Hồi quy đa biến** cho phép chúng ta nghiên cứu những trường hợp như vậy. Hãy xét các ví dụ sau:

4.1 Giới thiệu về hồi quy đa biến

Ví dụ 4.1: Rất nhiều các nghiên cứu trên thế giới quan tâm tới mối quan hệ giữa thu nhập và trình độ học vấn. Chúng ta kỳ vọng rằng, ít ra về trung bình mà nói, học vấn càng cao, thì thu nhập càng cao. Vì vậy, chúng ta có thể lập phương trình hồi quy sau:

$$\textit{Thu nhập} = \beta_1 + \beta_2 \textit{Học vấn} + \varepsilon$$

Tuy nhiên, mô hình này đã bỏ qua một yếu tố khá quan trọng là mọi người thường có mức thu nhập cao hơn khi họ làm việc lâu năm hơn, bất kể trình độ học vấn của họ thế nào. Vậy nên, mô hình tốt hơn cho mục đích nghiên cứu của chúng ta sẽ là:

$$\textit{Thu nhập} = \beta_1 + \beta_2 \textit{Học vấn} + \beta_3 \textit{Tuổi} + \varepsilon$$

Nhưng người ta cũng thường quan sát thấy, thu nhập có xu hướng tăng chậm dần khi người ta càng nhiều tuổi hơn so với thời trẻ. Để thể hiện điều đó, chúng ta mở rộng mô hình như sau:

$$\textit{Thu nhập} = \beta_1 + \beta_2 \textit{Học vấn} + \beta_3 \textit{Tuổi} + \beta_4 \textit{Tuổi}^2 + \varepsilon$$

Và chúng ta sẽ kỳ vọng rằng, β_3 mang dấu dương, và β_4 mang dấu âm.

Như vậy, chúng ta đã rời bỏ thế giới của hồi quy đơn và bước sang hồi quy đa biến.

Ví dụ 4.2: Nghiên cứu về nhu cầu đầu tư ở Mỹ trong khoảng thời gian từ năm 1968 – 1982.

Ở Mỹ, thời kỳ này mang dấu ấn lịch sử là cuộc chiến tranh Việt Nam kéo dài, dẫn đến bội chi ngân sách và lạm phát. Một năm sau khi chiến tranh kết thúc, lạm phát ở Mỹ đã đạt tới mức kỷ lục là 9.31% vào năm1976. Điều đó dẫn đến việc ngân hàng trung ương phải áp dụng mạnh mẽ chính sách tiền tệ chặt, vốn đã được áp dụng trong vài năm trước, và đưa mức lãi suất lên tới mức cao kỷ lục là 7.83%. Khi sự dính líu của Mỹ về quân sự tại Việt Nam đã hoàn toàn chấm dứt, nguồn nhân lực trước đây phục vụ cho chiến tranh nay chuyển ào ạt sang khu vực thương mại. Và điều này lại lại làm dấy lên một đợt lạm phát mới, đạt tới 9.44% vào năm 1981, sau đó được đưa về mức 5.99% vào năm 1982 nhờ vào việc nâng lãi suất lên tới 13.42%. Như vậy, lịch sử kinh tế Mỹ trong thời kỳ này được đặc trưng bởi chính sách tiền tệ chặt, kéo theo xu hướng cắt giảm liên tục về đầu tư qua các năm.

Chính vì vậy, các nhà nghiên cứu Mỹ đã đề xuất mô hình nghiên cứu sau về cầu đầu tư vào giai đoạn này:

$$INV = \beta_1 + \beta_2 T + \beta_3 G + \beta_4 INT + \varepsilon$$

Trong đó, *INV* và *G* lần lượt là cầu về đầu tư và GNP thực tế, đơn vị trillions dollars; *INT* là lãi suất; và *T* là biến xu thế, tính theo thời gian đã trôi qua, kể từ năm 1968. Từ lý luận kinh tế vĩ mô, chúng ta kỳ vọng rằng, β_3 mang dấu dương, và β_4 mang dấu âm. Và vì đây là thời kỳ đầu tư ở Mỹ có xu thế bị co hẹp, chúng ta cũng kỳ vọng rằng β_2 mang dấu âm.

Sử dụng dữ liệu thống kê vĩ mô của nền kinh tế Mỹ, từ năm 1968 - 1982 [xem bảng dữ liệu 4.2 phía dưới], kết quả ước lượng của mô hình hồi quy này như sau:

Bảng** Error! No text of specified style in document.**.1: Bảng kết xuất mô hình hồi qui các yếu tố ảnh hưởng đến cầu về đầu tư của Mỹ trong giai đoạn từ 1968 - 1982

Dependent Variable: INV
Method: Least Squares
Date: 04/09/07 Time: 16:14
Sample: 1 15
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.509237	0.052526	-9.694973	0.0000
T	-0.016583	0.001880	-8.819528	0.0000
G	0.670266	0.052426	12.78506	0.0000
INT	-0.002365	0.001034	-2.287282	0.0430

R-squared	0.972420	Mean dependent var	0.203333
Adjusted R-squared	0.964898	S.D. dependent var	0.034177
S.E. of regression	0.006403	Akaike info criterion	-7.040816
Sum squared resid	0.000451	Schwarz criterion	-6.852003
Log likelihood	56.80612	F-statistic	129.2784
Durbin-Watson stat	1.958353	Prob(F-statistic)	0.000000

Dưới dạng báo cáo, kết quả đó có thể được viết tóm tắt như dưới đây:

$$\underset{}{INV} = \underset{(0.0525)}{-0.5092} - \underset{(0.0018)}{0.0165T} + \underset{(0.052)}{0.67G} - \underset{(0.001)}{0.0023\, INT}$$

$$R^2 = 0.972,\ N = 15,\ ESS = 0.00045$$

Nếu viết dưới dạng sai phân, ta có:

$$\Delta INV = -\ 0.0165\,\Delta T + 0.67\,\Delta G - 0.0023\,\Delta INT$$

Nói khác đi, **nếu các yếu tố khác được giữ không đổi**, cứ sau mỗi một năm, kể từ năm 1968 (tức là $\Delta T = 1$), nhu cầu đầu tư sẽ bị giảm là -0.0165 trillions dollars. Cũng như vậy, nếu bỏ qua yếu tố xu thế và lãi suất, tác động riêng phần của việc tăng GNP lên 0.1 trillions dollars ($\Delta G = 0.1$), sẽ làm cầu về đầu tư tăng lên thêm 0.067 trillions; và nếu đẩy lãi suất lên thêm 1% ($\Delta INT = 1$), trong khi giữ nguyên các yếu tố còn lại, thì sẽ làm đầu tư giảm đi là -0.0023 trillions dollars.

Những tính toán trên đây cho thấy có sự tương đồng rõ rệt về cách diễn giải ý nghĩa của các hệ số ước lượng trong mô hình hồi quy đa biến so với trường hợp đơn biến. Điều đó gợi ý rằng, về mặt bản chất, mô hình hồi quy đa biến sẽ chỉ là sự mở rộng của hồi quy đơn biến. Ta sẽ thấy rõ hơn điều đó ở các phần sau.

4.2 Biểu diễn đại số của mô hình hồi quy đa biến

Chúng ta hãy đưa ra bảng so sánh về dạng hàm của mô hình hồi quy đa biến so với trường hợp đơn biến:

	Hồi quy đơn biến	**Hồi quy đa biến**
Ví dụ	$CONS = \beta_1 + \beta_2 INC + \varepsilon$	$INV = \beta_1 + \beta_2 T + \beta_3 G + \beta_4 INT + \varepsilon$
Dạng mô hình	$Y = \beta_1 + \beta_2 X + \varepsilon$	$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$
Với mỗi quan sát	$y_n = \beta_1 + \beta_2 x_n + \varepsilon_n$	$y_n = \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \beta_4 x_{n4} + \varepsilon_n$

Như vậy, hồi quy đa biến là một sự mở rộng tự nhiên của trường hợp đơn biến, khi số biến giải thích lớn hơn 2, kể cả hằng số. Để cho tiện lợi, chúng ta sẽ đưa vào các ký hiệu vector:

Gọi **vector hàng** $x_n^{'} = (1, x_{n2,} x_{n3,} x_{n4})$ là vector các quan sát thứ $n = 1,2..., N$ của các biến giải thích. [Lưu ý, **dấu phẩy** ở bên phải, phía trên vector x_n là dấu chuyển vị. Như vậy, theo mặc định, mọi vector (mà không có dấu chuyển vị) đều được coi là vector cột].

Từng "cặp" quan sát dữ liệu do vậy, sẽ là $\{y_n, x_n^{'}\}_{n=1}^{N}$.

Để minh họa, trong ví dụ 4.2 về cầu về đầu tư ở Mỹ (1968 – 82), những cặp $(y_5, x_5^{'})$ và $(y_{11}, x_{11}^{'})$ được tô màu:

Bảng** Error! No text of specified style in document.**.2: Dữ liệu vĩ mô về đầu tư và các biến giải thích của nền kinh tế Mỹ (1968 – 82).

Obs (n)	**INV (Y)**	**C (X1)**	**T (X2)**	**G (X3)**	**INT (X4)**	
1	0.161	1	1	1.058	5.16	
2	0.172	1	2	1.088	5.87	
3	0.158	1	3	1.086	5.95	
4	0.173	1	4	1.122	4.88	
5	0.195	1	5	1.186	4.5	$\boxed{(y_5, x_5^{'})}$
6	0.217	1	6	1.254	6.44	
7	0.199	1	7	1.246	7.83	
8	0.163	1	8	1.232	6.25	
9	0.195	1	9	1.298	5.5	
10	0.231	1	10	1.37	5.46	
11	0.257	1	11	1.439	7.46	$\boxed{(y_{11}, x_{11}^{'})}$
12	0.259	1	12	1.479	10.28	
13	0.225	1	13	1.474	11.77	
14	0.241	1	14	1.503	13.42	
15	0.204	1	15	1.475	11.02	

Nguồn: Economic Report of the President. Government, Printing Office, Washington D.C., 1983.

Tiếp theo, ta gọi **vector cột** $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix}$ là vector các tham số tổng thể, cần được ước lượng

Lưu ý rằng, **tích vô hướng** giữa hai vector $x_n^{'}$ và β sẽ tạo lại **phần xu thế** trong vế phải của phương trình hồi quy (4.2):

$$\underset{1\times 4}{x_n^{'}}\underset{4\times 1}{\beta} = (1, x_{n2,} x_{n3,} x_{n4}) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \beta_1 + \beta_2 x_{n2} + \beta_3 x_{n3} + \beta_4 x_{n4}, \qquad n = 1,2..N$$

Vì vậy, ứng với từng "cặp" quan sát $\{x_n^{'}, y_n\}_{n=1}^{N}$, ta có thể viết lại phương trình hồi quy đó như sau:

$$y_n = x_n^{'}\beta + \varepsilon_n \qquad n = 1,2,.., N \qquad (4.3)$$

Như vậy, mọi ký hiệu ta đã sử dụng trong ước lượng mô hình hồi quy đơn biến, nay có thể được sử dụng lại cho mô hình hồi quy đa biến. Cụ thể là:

$$\hat{y}_n = x_n^{'}\hat{\beta} \qquad n = 1,2,.., N \qquad (4.3)$$

Và sai số ước lượng hay số dư (residual) sẽ có dạng:

$$e_n = y_n - \hat{y}_n \qquad (4.4)$$

Việc tiến hành ước lượng các tham số của mô hình bằng phương pháp bình phương cực tiểu tương đương với việc giải bài toán sau:

$$S(\hat{\beta}) = \sum_n e_n^2 = \sum (y_n - x_n^{'}\hat{\beta})^2 \rightarrow \min_{\hat{\beta}} \qquad (4.5)$$

Tương tự như trong hồi quy đơn, ở đây, ta sử dụng điều kiện cực trị, (first order condition, FOC), để tìm các tham số ước lượng $\hat{\beta}_k, k = 1,2,3,4.$. Nói khác đi, ta đi giải hệ phương trình sau:

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}_1} = 0$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}_2} = 0$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}_3} = 0$$

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}_4} = 0 \qquad (4.6)$$

Đây là hệ gồm 4 phương trình với 4 ẩn số, mà việc giải nó cho chúng ta tham số ước lượng $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4)'$. Sử dụng phần mềm Eviews, kết quả tính toán các tham số này đã được nêu trong bảng báo cáo 4.1 ở trên.

Mặc dù dạng biểu diễn giải tích của vector $\underset{4\times 1}{\hat{\beta}}$ là khá phức tạp. Tuy nhiên, về bản chất chúng vẫn không khác gì trường hợp đơn biến. Cụ thể là, tương tự như $\hat{\alpha}$, phương trình đầu tiên của hệ (4.6) để ước lượng $\hat{\beta}_1$ dẫn đến cái điều là, đường hồi quy đi qua điểm trung bình $(\bar{y}_n, \bar{x}'_n)$. Và vì vậy, ta cũng có thể nói đến tiêu chuẩn đo lường độ phù hợp của đường hồi quy R^2. Cụ thể là từ mối quan hệ (4.4):

$$y_n = \hat{y}_n + e_n$$

Hay cũng hệt như thế:

$$(y_n - \bar{y}) = \hat{y}_n - \bar{y} + e_n$$

Người ta có thể viết lại nó như sau:

$$(y_n - \bar{y}) = (x'_n - \bar{x}')\hat{\beta} + e_n$$

Tức là, sự giao động so với trung bình của biến *Y* được giải thích một phần bởi mô hình, và phần còn lại là sai số e_n, chưa được giải thích bởi mô hình. Sử dụng các điều kiện tìm cực trị (4.6), ta cũng có thể viết lại quan hệ đó như sau:

$$\sum_n (y_n - \bar{y})^2 = \sum_n (\hat{y}_n - \bar{y})^2 + \sum_n e_n^2$$

Hay cũng vậy,

TSS = RSS + ESS

Vì thế, chúng ta có thể đưa ra định nghĩa:

$$R^2 = 1 - \frac{ESS}{TSS} \qquad (0 \le R^2 \le 1).$$

và sử dụng nó làm thước đo mức độ phù hợp của đường hồi quy với dữ liệu quan sát.

Phần tiếp sau sẽ đề cập đến bản chất thống kê của mô hình hồi quy đa biến.

4.3 Bản chất thống kê của mô hình hồi quy đa biến

Từ bây giờ, chúng ta sẽ sử dụng dạng tổng quát của mô hình hồi quy đa biến:

$$\begin{aligned} Y &= \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + ... + \beta_K X_K + \varepsilon \qquad (4.7) \\ y_n &= \beta_1 + \beta_2 x_{n2} + ... + \beta_k x_{nk} + ... + \beta_K x_{nK} + \varepsilon_n \\ &= x_n^{'} \beta + \varepsilon_n, \qquad n = 1,2,3,..., N \end{aligned}$$

Việc hồi quy mô hình (4.7) sẽ cho ta biểu diễn sau:

$$\begin{aligned} Y &= \hat{\beta}_1 + \hat{\beta}_2 X_2 + ... + \hat{\beta}_k X_k + ... + \hat{\beta}_K X_K + e \qquad (4.8) \\ y_n &= \hat{\beta}_1 + \hat{\beta}_2 x_{n2} + ... + \hat{\beta}_k x_{nk} + ... + \hat{\beta}_K x_{nK} + e_n \\ &= x_n^{'} \hat{\beta} + e_n, \qquad n = 1,2,3,..., N \end{aligned}$$

Trong đó, N là số quan sát, và K là số biến giải thích.

Ta phát biểu định lý sau[1]:

Định lý 4.1: Phương pháp bình phương cực tiểu, áp dụng cho mô hình hồi quy đa biến, sẽ cho ta các tham số ước lượng dưới dạng sau:

$$\hat{\beta}_k = \beta_k + \sum_n c_{kn} \varepsilon_n, k = 1,2,..,K \qquad (4.9)$$

Cũng như trường hợp đơn biến, phương trình (4.9) chỉ ra rằng: $\hat{\beta}_k$ bị tác động bởi các yếu tố ngẫu nhiên ε_n, làm giá trị của nó không trùng khít với β_k tổng thể. Và vì bị tác động

[1] Xem chứng minh chi tiết ở chương 8, phần Maximum likelihood.

bởi các yếu tố ngẫu nhiên, $\hat{\beta}_k$ cũng là một biến ngẫu nhiên. Do đó, độ tốt của ước lượng sẽ phụ thuộc trực tiếp vào bản chất của các quá trình ngẫu nhiên $\{\varepsilon_n\}_{n=1}^{N}$.

Điều này dẫn đến việc cần phải khắc họa bản chất thống kê của mô hình hồi quy, như chúng ta đã làm cho trường hợp đơn biến. Ta sẽ tiếp tục sử dụng các giả thuyết đã đưa ra về ε_n. Cụ thể là:

A1 $E\varepsilon_n = 0,$ với mọi *n*.

A2 $Var\varepsilon_n = \sigma^2$, với mọi *n*.

A3 $\varepsilon_n \overset{iid}{\sim} N(0,\sigma^2)$, với mọi *n*. Và:

A4 $E(y_n \mid x_n') = x_n'\beta$, với mọi *n*.

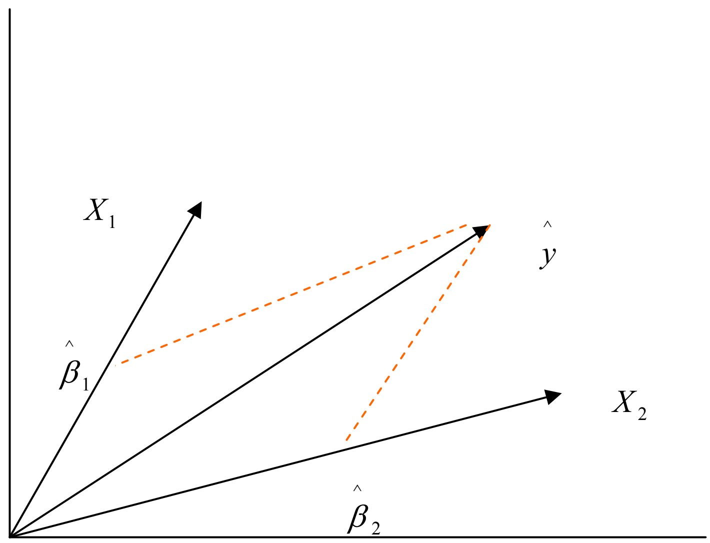
Đối với trường hợp đa biến, chúng ta đưa thêm đòi hỏi sau:

Gọi $\underset{N\times K}{X} = [X_1, X_2, ..., X_k, ..., X_K]$ là ma trận tạo bởi các vector cột của *K* biến giải thích [xem lại ví dụ minh họa về ma trận *X* ở bảng 4.2 về dữ liệu của mô hình đầu tư]. Khi đó, ta đòi hỏi rằng:

A5 Các cột $\{X_1, X_2, ..., X_k, ..., X_K\}$ là độc lập tuyến tính. Hay cũng vậy, rank *X* = *K*.

Về mặt hình học, giả thuyết này có ý nghĩa như sau. Hãy xét trường hợp *K = 2*, phương pháp bình phương cực tiểu có thể được biểu diễn bởi lược đồ dưới đây:

Đồ thị *Error! No text of specified style in document.***.1: Biểu diễn hình học của hồi quy***



Việc ước lượng tham số $\hat{\beta}$ cũng giống như là tìm các hệ số $\underset{2\times 1}{\hat{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)'$ sao cho $\hat{y} = \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$. Để làm được điều đó, điều kiện cần là các vector X_1, X_2 không được trùng khít với nhau. Hay cũng vậy, X_1, X_2 phải độc lập tuyến tính. Đây được gọi là **điều kiện xác định** (identification condition). Trong trường hợp tổng quát, khi $K \geq 2$, điều kiện đó được phát biểu dưới dạng giả thuyết A5. Chúng ta sẽ sử dụng giả thuyết này khi bàn tới vấn đề đa cộng tuyến (multicollinearity) trong chương 7.

4.4 Kiểm định các giả thuyết thống kê

Bây giờ hãy chỉ chú ý đến giả thuyết đầu tiên A1 – A3, và sử dụng chúng để đánh giá tính tốt của ước lượng theo các tiêu chuẩn thống kê.

Từ phương trình (4.9), ta đã có: $\hat{\beta}_k = \beta_k + \sum c_{kn}\varepsilon_n$. Bây giờ, hãy áp dụng toán tử kỳ vọng vào hai vế của (4.9). Ta có:

$$\begin{aligned} E\hat{\beta}_k &= E(\beta_k + \sum c_{kn}\varepsilon_n) \\ &= \beta_k + \sum c_{kn} E\varepsilon_n \\ &= \beta_k \end{aligned} \qquad (4.10)$$

[ở đây, ta sử dụng giả thiết A1: $E\varepsilon_n = 0$]. Do vậy, $\hat{\beta}_k$ là ước lượng không chệch của β_k.

Tiếp theo, sử dụng lại công thức: $Var(x) = Var(x - Ex)$ [xem chương 1, phần ôn tập], và lưu ý (4.9), (4.10), ta có:

$$
\begin{aligned}
Var\hat{\beta}_k &= Var(\hat{\beta}_k - \beta_k) \\
&= Var(\sum c_{kn}\varepsilon_n)
\end{aligned}
$$

Sử dụng giả thiết A3 về tính độc lập của các yếu tố ngẫu nhiên, cuối cùng ta nhận được:

$$
\begin{aligned}
Var\hat{\beta}_k &= \sum c_{kn}^2 Var\varepsilon_n \\
&= \sigma^2 \sum c_{kn}^2 \text{ , hay}
\end{aligned}
$$

$$Var\,\hat{\beta}_k = \frac{\sigma^2}{S_{kk}}, k = 1,2,..,K \tag{4.11}$$

(ở đây, mặc dù ta không đưa ra được tính toán trực tiếp; nhưng về cơ bản S_{kk} cũng là phương sai mẫu của biến X_k, tương tự như S_{XX} trong trường hợp đơn biến).

Định Lý 4.2 [Gauss – Markov]: Phương pháp bình phương cực tiểu có sai số ước lượng, đo lường bởi $Var\,\hat{\beta}_k, k = 1,2,..,K$, là nhỏ nhất trong lớp tất cả các ước lượng tuyến tính và không chệch.

Ta cũng nên nhấn mạnh lại rằng, chúng ta có được những tính chất rất tốt: **không chệch** và **hiệu quả** của ước lượng bình phương cực tiểu, mà chỉ đòi hỏi có trung bình bằng zero, tính độc lập, và phương sai giống nhau của các yếu tố ngẫu nhiên – tức là giả thiết A3.

Sử dụng (4.9) – (4.11), chúng ta đi đến kết luận rằng: $\hat{\beta}_k \sim N(\beta_k, \frac{\sigma^2}{S_{kk}})$. Điều đó có nghĩa là, sau khi chuẩn hóa, $Z_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2/S_{kk}}} \sim N(0,1)$. Thay thế σ^2 bởi ước lượng không chệch của nó là $s^2 = \frac{1}{N-K}\sum_n e_n^2$, ta có thống kê $t_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{s^2/S_{kk}}} = \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim t(N-K)$. Chúng ta bây giờ có thể xây dựng khoảng tin cậy cho $\beta_k, k = 1,2,..,K$, và tiến hành kiểm định các giả thuyết thống kê về tham số của tổng thể. Chẳng hạn như nếu muốn kiểm định tính có ý nghĩa của biến giải thích $X_k, k = 1,2,..,K$, chúng ta lập giả thuyết sau:

$$H_0 : \beta_k = 0 \ \text{ .vs. } \ H_1 : \beta_k \neq 0$$

Việc kiểm định bao gồm các bước sau:

Bước 1: Xác định thống kê – t (t-stat):

$$t_k = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \sim t(N-K)$$

Bước 2: Tra bảng thống kê t-student với *(N-K)* bậc tự do $t(N-K)$ để tìm giá trị t-tra bảng (t-critical) $t_{\lambda/2}(N-K)$, ứng với mỗi mức ý nghĩa (significance) λ [Chẳng hạn, 0.05 (5%); hay 0.1 (10%)]

Bước 3: Bác bỏ giả thuyết H_0 (viết tắt là RH_0), nếu $|t_k| = \left|\frac{\hat{\beta}_k}{se(\hat{\beta}_k)}\right| \geq t_{\lambda/2}(N-K)$, và không bác bỏ giả thuyết đó ($DNRH_0$), nếu $\left|\frac{\hat{\beta}_k}{se(\hat{\beta}_k)}\right| \leq t_{\lambda/2}(N-K)$.

Cũng như trong trường hợp đơn biến, người ta thường hay sử dụng *p-value*, hơn là phải tính toán và tra bảng qua các bước 1 đến 3 như trên.

Cụ thể, ứng với từng biến giải thích $X_k, k = 1,2,..,K$, ta cũng đặt:

$$p - value = \Pr ob\{|\, t(N-K)\,| \geq |\, t_k\,|)$$

Cũng hệt như ở đồ thị 3-9, chúng ta sẽ bác bỏ giả thuyết $H_0 : \beta_k = 0$, nếu: $p - value \le \lambda$, [trong trường hợp đó, ta nói X_k là có ý nghĩa ở mức $\lambda\%$]. Và chúng ta sẽ không bác bỏ giả thuyết đó, nếu $p - value \ge \lambda$.

Trong ví dụ 4.1 về đầu tư ở Mỹ (1968-82), *p-value* của cả 3 biến giải thích: *T, G, INT* , đều nhỏ hơn 5%. Vì vậy, ta nói rằng tất cả các biến này là có ý nghĩa ở mức $\lambda = 5\%$.