

Chương 13

Chuỗi dừng và không dừng

(Gujarati: Econometrics by example, 2011)¹.

Người dịch và diễn giải: Phùng Thanh Bình

<http://vnp.edu.vn/>



Trong phân tích hồi quy với dữ liệu chuỗi thời gian, một giả định rất quan trọng là chuỗi thời gian đang xem xét là chuỗi dừng (stationary). Nói chung, một chuỗi thời gian dừng nếu trung bình (mean) và phương sai (variance) của nó không đổi qua thời gian và giá trị hiệp phương sai (covariance) giữa hai giai đoạn chỉ phụ thuộc vào khoảng cách giữa hai giai đoạn ấy chứ không phụ thuộc vào thời gian thực sự tại đó hiệp phương sai được tính². [Diễn giải: Các hệ số tự tương quan giữa Y_t và Y_{t-1} , Y_{t-5} và Y_{t-6} , hoặc Y_{t-12} và Y_{t-13} sẽ là như nhau; nhưng khác với các hệ số tự tương quan giữa Y_t và Y_{t-3} , Y_{t-5} và Y_{t-8} , ... Nói cách khác, nếu ta ngắt dữ liệu thành các giai đoạn khác nhau, thì giản đồ tự tương quan của các giai đoạn đó là như nhau].

Một chuỗi thời gian là một ví dụ về điều mà ta gọi là một quy trình ngẫu nhiên (stochastic process), đó là một chuỗi của các biến ngẫu nhiên được sắp xếp thứ tự theo thời gian³.

13.1 Tỷ giá hối đoái có phải là chuỗi dừng?

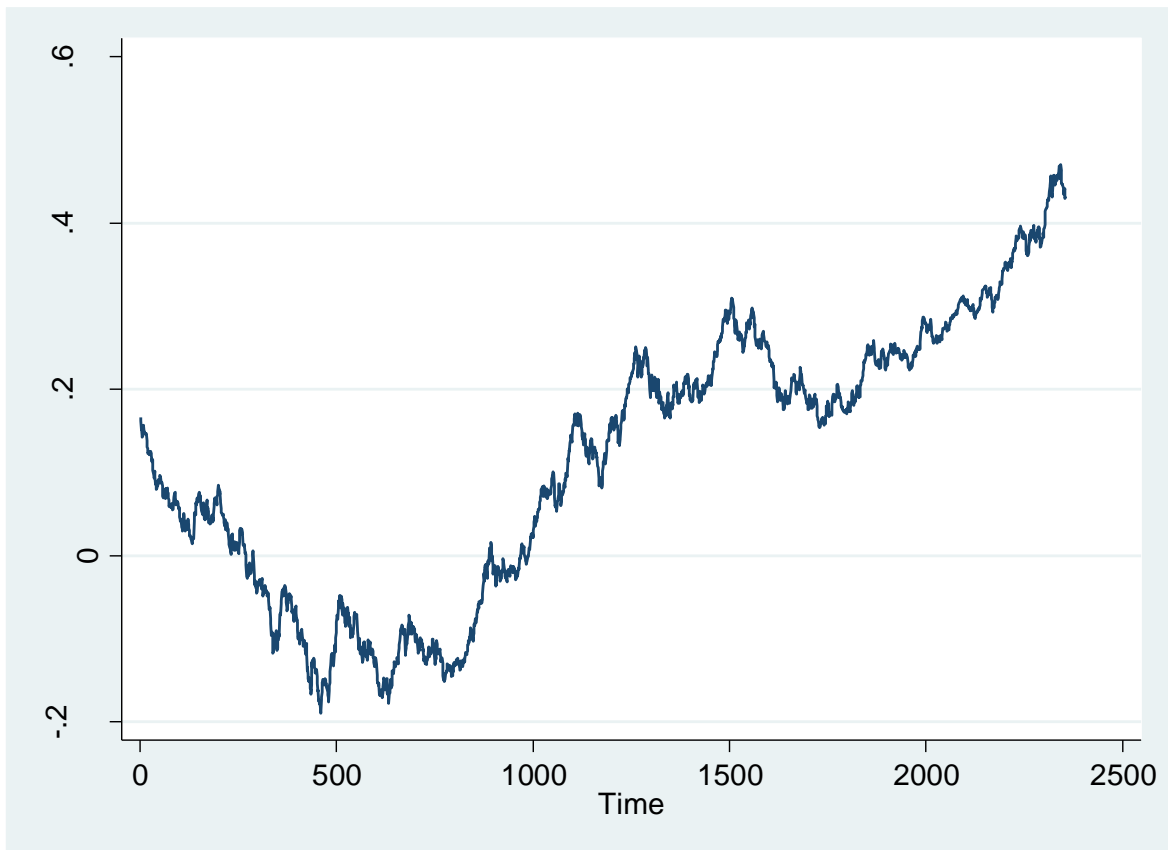
Để giải thích chuỗi dừng có nghĩa là gì, chúng ta xem xét một chuỗi thời gian kinh tế cụ thể: tỷ giá hối đoái giữa đôla Mỹ và đồng Euro (đặt tên biến này là EX), được định nghĩa là số đồng đôla trên một đơn vị đồng Euro. Dữ liệu tỷ giá hối đoái được thu thập theo ngày, từ ngày 4 tháng 1 năm 2000 đến ngày 8 tháng 5 năm 2008, với tổng số là 2.355 quan sát. Dữ liệu này không liên tục, vì các thị trường ngoại hối không mở cửa mỗi ngày và phải trừ những ngày nghỉ lễ. Dữ liệu này được cho sẵn trong tập tin **Table 13.1**, bạn có thể tìm thấy trên trang web của cuốn sách này.

¹ Hiện nay đã có ấn bản mới (lần 2, năm 2015). Dữ liệu của phiên bản 2011:

<https://www.macmillanihe.com/companion/Gujarati-Econometrics-By-Example/student-zone/>

² Một chuỗi thời gian với các đặc điểm này được biết như chuỗi dừng yếu (weakly stationary) hoặc chuỗi dừng hiệp phương sai (covariance stationary). Một chuỗi thời gian là dừng 'ng nghiêm ngặt' (strictly stationary) nếu tất cả các mô men (moments) của phân phối xác suất của nó chứ không chỉ là hai mô men đầu tiên (tức trung bình và phương sai) là không đổi qua thời gian. Tuy nhiên, nếu quy trình dừng (stationary process) có phân phối chuẩn, thì quy trình ngẫu nhiên dừng yếu (weakly stationary stochastic process) cũng là quy trình dừng nghiêm ngặt, vì quy trình chuẩn được xác định một cách đầy đủ bởi hai mô men: trung bình và phương sai.

³ Thuật ngữ 'ngẫu nhiên' (stochastic) có từ tiếng Hy Lạp là *stokhos*, có nghĩa là một mục tiêu (target) hoặc điểm đen (bull's-eye). Bất kỳ ai ném phi tiêu vào một tấm bia đều biết rằng quá trình ném trúng điểm đen là quá trình ngẫu nhiên; một số mũi phi tiêu văng ra ngoài, một số sẽ trúng mục tiêu, nhưng hầu hết chúng sẽ phân tán xung quanh tâm điểm theo một cách hoàn toàn ngẫu nhiên.



Hình 13.1: LEX của tỷ giá hối đoái theo ngày tỷ đồng đôla/euro.

[Lưu ý: Lệnh trên Stata: **twoway (tsline lnex)**]

Trong Hình 13.1, chúng ta trình bày log của tỷ giá hàng ngày giữa đôla/euro (LEX). Ý tưởng cơ bản của việc vẽ đồ thị về log của tỷ giá thay vì tỷ giá là thay đổi trong log của một biến thể hiện một sự thay đổi tương đối (hoặc suất sinh lợi), trong khi sự thay đổi trong một biến không ở dạng logarithm thể hiện một sự thay đổi tuyệt đối. Vì các mục đích so sánh, cách thể hiện tương đối nói chung sẽ thú vị hơn.

Nhìn sơ qua, đồ thị này cho thấy chuỗi LEX là không dừng, vì nhìn chung nó có xu hướng di chuyển đi lên, mặc dù có sự biến thiên đáng kể. Điều này cho thấy cả giá trị trung bình và phương sai của chuỗi tỷ giá đều không dừng. *Nói một cách chính thức hơn, một chuỗi thời gian được cho là dừng nếu trung bình và phương sai của nó là không đổi qua thời gian và giá trị hiệp phương sai giữa hai giai đoạn chỉ phụ thuộc vào khoảng cách giữa hai giai đoạn ấy chứ không phụ thuộc vào thời gian thực sự tại đó hiệp phương sai được tính.* Chuỗi thời gian như thế được gọi là chuỗi dừng yếu hoặc chuỗi dừng hiệp phương sai⁴.

13.2 Tầm quan trọng của chuỗi dừng

Tại sao chúng ta lo lắng chuyện một chuỗi thời gian có dừng hay không? Có một số lý do. Thứ nhất, nếu một chuỗi không dừng, chúng ta có thể nghiên cứu hành vi của nó

⁴ Như đã lưu ý ở trên, một chuỗi được cho là dừng nghiêm ngặt nếu tất cả các mô men của phân phối xác suất của nó và không chỉ có trung bình và phương sai không đổi qua thời gian.

chỉ cho riêng giai đoạn đang xem xét, chẳng hạn tỷ giá hối đoái trong giai đoạn 4/1/2000 đến 8/5/2008. Vì thế, mỗi chuỗi thời gian là một giai đoạn riêng biệt. Cho nên, chúng ta không thể khái quát hóa kết quả phân tích cho các giai đoạn khác. Đối với các mục đích dự báo, chuỗi không dừng sẽ không có giá trị ứng dụng thực tiễn. [Diễn giải: Một điều kiện ngầm ẩn trong dự báo là xu thế vận động của dữ liệu trong quá khứ và hiện tại sẽ được duy trì cho tương lai. Như thế, chỉ có chuỗi dừng thì điều kiện này mới thỏa mãn].

Thứ hai, nếu chúng ta có hai hoặc nhiều chuỗi không dừng, phân tích hồi quy với các chuỗi như thế có thể dẫn đến hiện tượng hồi quy giả mạo (spurious regression) hoặc hồi quy vô nghĩa (nonsense regression). Đó là, nếu bạn hồi quy một chuỗi không dừng với một hoặc nhiều chuỗi không dừng khác, bạn có thể có một giá trị R^2 cao và một số hoặc tất cả các hệ số hồi quy có thể có ý nghĩa thống kê dựa trên các kiểm định t hoặc F thông thường. Nhưng không may, trong trường hợp chuỗi không dừng, thì các kiểm định này là không thể tin cậy, vì chúng giả định rằng các chuỗi thời gian về cơ bản là các chuỗi dừng. Chúng ta sẽ thảo luận chủ đề hồi quy giả mạo chi tiết hơn ở chương tiếp theo. [Diễn giải: Ví dụ bạn thực hiện hồi quy giữa GDP Việt Nam theo cung tiền M2 của Nepal thì bạn sẽ có giá trị R^2 cao và hệ số hồi quy dương và có ý nghĩa thống kê. Nhưng, kết quả hồi quy như thế có ý nghĩa gì về mặt chính sách? Không hề có. Việc R^2 cao và hệ số có ý nghĩa chẳng qua là do mối quan hệ giữa hai xu thế chứa trong hai chuỗi dữ liệu tạo ra, chứ bản chất kinh tế là vô nghĩa. Bạn cũng có thể hồi quy giữa chỉ số giá chứng khoán VN Index với tiền công thuê lao động ở Gò Công, hay ở Cẩm Mỹ, bạn vẫn có kết quả y như thế. Tôi dám chắc điều này].

13.3 Các kiểm định dừng

Vì các lý do vừa được nói ở trên, nên điều quan trọng là chúng ta cần xem xét một chuỗi có dừng hay không dừng. Cơ bản ta có ba cách để khảo sát tính dừng của một chuỗi thời gian: (1) phân tích đồ thị (graphical analysis), (2) giản đồ tự tương quan (correlogram), và (3) phân tích nghiệm đơn vị (unit root analysis). Chúng ta sẽ thảo luận hai cách đầu tiên trong mục này và tiếp tục bàn luận cách thứ ba ở mục tiếp theo.

Phân tích đồ thị

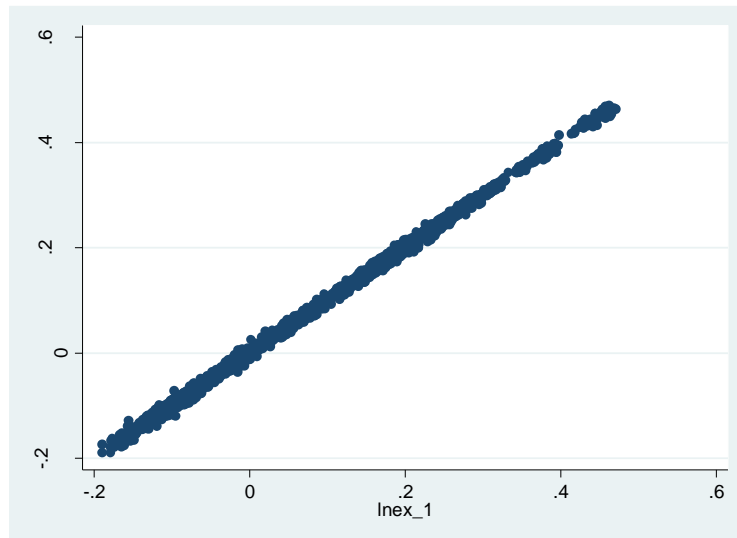
Một phương pháp đơn giản để kiểm định tính dừng là vẽ đồ thị chuỗi thời gian, như chúng ta thực hiện ở Hình 13.1. Thường thì một phân tích không chính thức như thế sẽ cho ta một dấu hiệu sơ bộ nào đó để biết một chuỗi có dừng hay không. Một cảm nhận trực quan như thế là điểm khởi đầu cho các kiểm định chính thức hơn về tính dừng. Và điều đáng ghi nhớ là “bất kỳ ai cố gắng phân tích một chuỗi thời gian mà không vẽ đồ thị tức xem trước tức là đang tự chuốc lấy phiền phức”⁵.

[Diễn giải: Trên Stata, đầu tiên bạn tạo một biến trễ một giai đoạn của chuỗi log tỷ giá hối đoái như sau:

```
gen lnex_1=lnex[_n-1]
```

```
twoway (scatter lnex lnex_1)]
```

⁵ Chris Chatfield. The analysis of time series: An introduction, 6th edn, Chapman & Hall/CRC Press.



Hình 13.2: Giá trị hiện hành và độ trễ một giai đoạn của LEX.

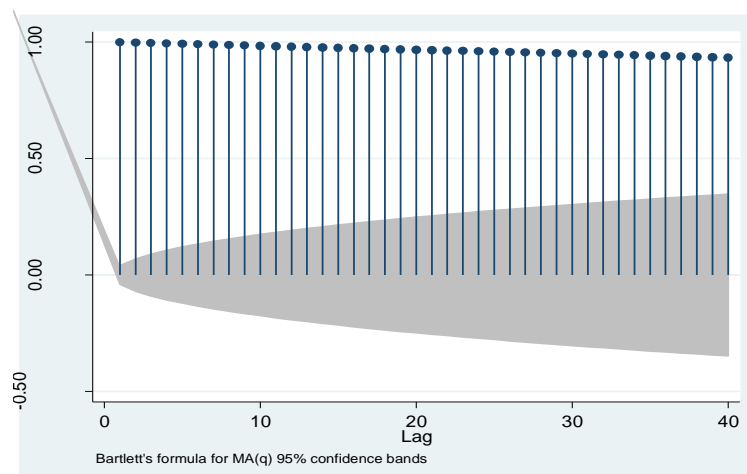
Hàm tự tương quan và giản đồ tự tương quan

Trong Hình 13.2, chúng ta vẽ LEX tại thời điểm t (LEX_t) theo biến trễ một giai đoạn của nó (LEX_{t-1}). Hình này cho thấy sự tương quan rất cao giữa LEX hiện hành và LEX trễ một ngày. Nhưng rất có thể sự tương quan này kéo dài qua một số ngày. Nghĩa là, giá trị hiện hành của LEX có thể tương quan với LEX trễ một số ngày. Để biết sự tương quan kéo dài trong bao lâu, chúng ta có thể sử dụng hàm tự tương quan (autocorrelation function, AFC). AFC tại độ trễ k được định nghĩa như sau:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{hiệp phương sai tại độ trễ } k}{\text{phương sai}} \quad (13.1)$$

Trong thực tế, chúng ta tính AFC từ một mẫu cho trước, và được ký hiệu là $\hat{\rho}_k$, dựa trên hiệp phương sai mẫu tại độ trễ k và phương sai mẫu. Chúng ta không nên bị công thức thực tế này làm mình nản lòng, vì các phần mềm hiện đại đều đã tính sẵn.

[Diễn giải: Ví dụ, trên cửa sổ lệnh của Stata, bạn chỉ cần sử dụng lệnh `ac lnex` thì bạn sẽ có đồ thị sau đây:]



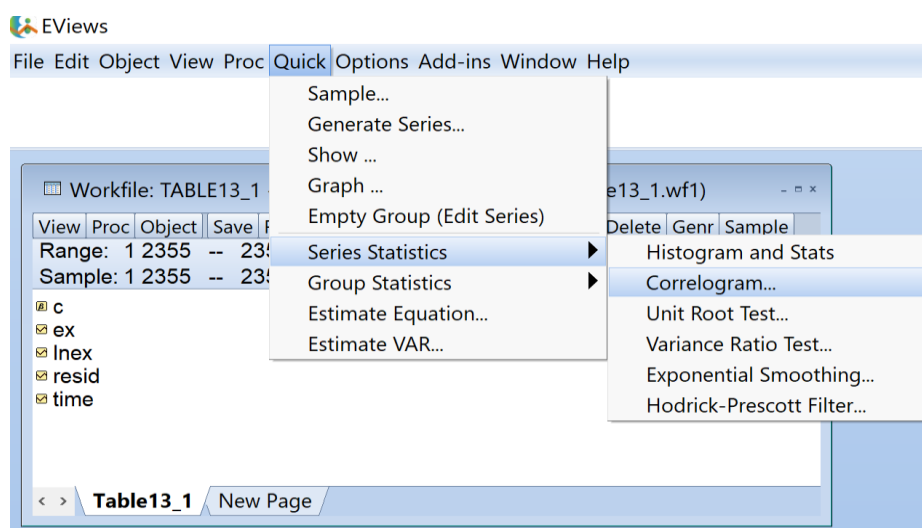
Một câu hỏi thực tiễn quan trọng là về độ dài của độ trễ k . Chúng ta có thể sử dụng tiêu chí thông tin Akaike (AIC) hoặc Schwarz (SIC) để xác định độ dài của độ trễ⁶. Nhưng một quy tắc thực nghiệm là tính AFC trong khoảng từ $\frac{1}{4}$ đến $\frac{1}{3}$ độ dài của chuỗi thời gian. Chúng ta có 2.355 quan sát. Như vậy, $\frac{1}{4}$ con số này là khoảng 589 độ trễ. Chúng ta sẽ không trình bày AFC ở tất cả các độ trễ này, nhưng chỉ cần xem 30 độ trễ đầu tiên để bạn có một ý tưởng gì đó về bản chất của AFC. Đồ thị về \hat{p}_k theo k , chiều dài độ trễ, được gọi là giản đồ tự tương quan mẫu. Lúc này, chúng ta tạm lờ đi cột về hệ số tự tương quan riêng (PAC), hệ số này sẽ cần thiết ở chương 16 khi chúng ta bàn luận về dự báo chuỗi thời gian. [Diễn giải: PAC có ý nghĩa tương tự như hệ số hồi quy riêng giữa Y_t với từng độ trễ của nó. Và PAC dường như được phát minh ra chỉ cho duy nhất một chuyện là xác định độ trễ trong các mô hình ARIMA].

Đối với chuỗi tỷ giá đôla/euro, giản đồ tự tương quan được cho ở Bảng 13.2. [Diễn giải: Trên Stata, ta sử dụng lệnh `corrgram lnx, lags(20)`].

Bảng 13.2: Giản đồ tự tương quan mẫu của tỷ giá đôla/euro.

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial Autocor]		
1	0.9985	1.0001	2350.9	0.0000						
2	0.9970	-0.0021	4695.7	0.0000						
3	0.9954	0.0009	7034.2	0.0000						
4	0.9939	0.0169	9366.6	0.0000						
5	0.9924	-0.0516	11693	0.0000						
6	0.9909	0.0369	14013	0.0000						
7	0.9893	-0.0163	16326	0.0000						
8	0.9877	-0.0214	18633	0.0000						
9	0.9861	0.0285	20934	0.0000						
10	0.9845	-0.0079	23228	0.0000						
11	0.9829	0.0154	25516	0.0000						
12	0.9812	0.0046	27796	0.0000						
13	0.9795	-0.0161	30070	0.0000						
14	0.9778	-0.0306	32337	0.0000						
15	0.9761	0.0011	34597	0.0000						
16	0.9744	0.0316	36850	0.0000						
17	0.9727	-0.0103	39097	0.0000						
18	0.9710	-0.0005	41336	0.0000						
19	0.9693	-0.0218	43569	0.0000						
20	0.9676	-0.0210	45795	0.0000						

Với Eviews, ta thực hiện như sau:



⁶ Chúng ta đã thảo luận các tiêu chí này ở Chương 2.

Chọn biến *Inex*.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.998	0.998	2350.9	0.000
		2	0.997	0.004	4695.7	0.000
		3	0.995	-0.017	7034.2	0.000
		4	0.994	0.012	9366.6	0.000
		5	0.992	-0.014	11693.	0.000
		6	0.991	0.012	14013.	0.000
		7	0.989	-0.020	16326.	0.000
		8	0.988	-0.018	18633.	0.000
		9	0.986	0.006	20934.	0.000
		10	0.984	0.001	23228.	0.000
		11	0.983	0.001	25516.	0.000
		12	0.981	-0.024	27796.	0.000
		13	0.979	-0.019	30070.	0.000
		14	0.978	-0.001	32337.	0.000
		15	0.976	0.016	34597.	0.000
		16	0.974	-0.007	36850.	0.000
		17	0.973	-0.010	39097.	0.000
		18	0.971	0.020	41336.	0.000
		19	0.969	-0.011	43569.	0.000
		20	0.968	-0.005	45795.	0.000

Trước khi đi tiếp, chúng ta nên đề cập đến một loại đặc biệt của chuỗi thời gian, đó là chuỗi nhiễu trắng (white noise), hoặc chuỗi ngẫu nhiên thuần túy (purely random). Chuỗi thời gian như thế có trung bình không đổi, phương sai không đổi (tức là phương sai đồng nhất), và không có tương quan chuỗi (serially uncorrelated); giá trị trung bình của nó thường giả định bằng không. Nhớ lại hạng nhiễu u_t trong mô hình hồi quy tuyến tính cổ điển được giả định là một quá trình ngẫu nhiên nhiễu trắng, được ký hiệu như sau: $u_t \sim IID(0, \sigma^2)$, nghĩa là phân phối độc lập và giống nhau (independently and identically distributed) với trung bình bằng 0 và phương sai không đổi. Hơn nữa, nếu u_t có phân phối chuẩn, thì nó được gọi là một quy trình nhiễu trắng Gauss. Đối với một chuỗi thời gian như thế, thì AFC tại các độ trễ khác nhau sẽ xoay quanh 0 và giản đồ tự tương quan không có biên dạng rõ ràng.

Trở lại với ví dụ của chúng ta, chúng ta hãy tập trung vào cột ACF và thể hiện bằng đồ thị của nó (tức giản đồ tự tương quan) được cho ở cột thứ 6 (Bảng 13.2) [Lưu ý: Cột thứ nhất, đối với giản đồ tự tương quan trong Eviews]. Như bạn có thể thấy, thậm chí thứ 30 (trong Bảng trên chỉ dừng lại ở 20 độ trễ thôi, nếu cho 30 độ trễ bảng này sẽ nhảy sang trang khác 😊)), hệ số tự tương quan cũng còn rất cao, khoảng 0.95. Không chỉ có thế, các hệ số tự hồi quy ước lượng, \hat{p}_k , giảm rất chậm về không. Điều này hoàn toàn trái ngược với giản đồ tự tương quan của một chuỗi thời gian nhiễu trắng (xem Bảng 13.5). [Diễn giải: Một chuỗi có các hệ số tự tương quan đầu tiên rất cao, gần bằng 1, và giảm về 0 rất chậm sau rất nhiều độ trễ (như Bảng trên) là một chuỗi không dừng; một chuỗi có một vài độ trễ đầu tiên khác không, nhưng không cao quá (ví dụ chừng 0.7, hoặc 0.6), và sau đó giảm rất nhanh xuống 0 chỉ sau vài độ trễ ban đầu, đó là một chuỗi

dừng; và một chuỗi có tất cả các hệ số tự tương quan bằng 0 (kể cả độ trễ đầu tiên), đó là một chuỗi ngẫu nhiên hay nhiễu trắng. Theo quy tắc kinh nghiệm, trong Eviews, nếu hệ số tự tương quan nào nằm ngoài đường viền (khoảng tin cậy 95%, như ở cột 2 ở Bảng trên), được xem là khác không. Trong Stata, thì hệ số nào nằm ngoài khoảng màu sậm, như Hình ở trang 6, được xem là khác không].

Chúng ta có thể kiểm định ý nghĩa của mỗi hệ số tự tương quan bằng cách tính sai số chuẩn (standard error) của nó. Nhà thống kê Barlett chỉ ra rằng nếu một chuỗi thời gian là ngẫu nhiên thuần túy, thì hệ số tự tương quan mẫu, \hat{p}_k , (đối với các mẫu lớn) theo phân phối xấp xỉ chuẩn như sau:

$$\hat{p}_k \sim N(0, 1/n) \quad (13.2)$$

Nghĩa là, trong các mẫu lớn, \hat{p}_k theo phân phối xấp xỉ chuẩn với trung bình bằng 0 và phương sai bằng 1 chia cho cỡ mẫu. Cỡ mẫu của chúng ta là 2.355. Vì thế, phương sai là $1/2355$ hoặc 0.00042, và sai số chuẩn là căn bậc hai của con số $0.00042 = 0.0206$. Vì thế, theo tính chất của phân phối chuẩn, khoảng tin cậy 95% của p_k là $[0 \pm 1.96(0.0206)]$ hay $(-0.0404$ đến $0.0404)$.

Không có hệ số tự tương quan nào nằm trong khoảng này. Vì thế, chúng ta có thể kết luận rằng tất cả các hệ số tự tương quan trong bảng trên là có ý nghĩa thống kê [Diễn giải: Nói đúng hơn là khác 0 một cách có ý nghĩa thống kê]. Kết luận này không thay đổi thậm chí nếu chúng ta tính ACF cho đến 150 độ trễ! Đây là một dấu hiệu rất mạnh để kết luận rằng chuỗi LEX là không dừng.

Thay vì đánh giá ý nghĩa thống kê của từng hệ số tự tương quan riêng lẻ, chúng ta cũng có thể xem xét liệu rằng tổng các hệ số tự tương quan bình phương có ý nghĩa thống kê hay không [Diễn giải: Nói cách khác, ta xem xét liệu rằng ít nhất một hệ số tự tương quan nào đó trong tổng số m độ trễ có khác không một cách có ý nghĩa thống kê hay không; ở đây, giả thuyết H_0 là tất cả các hệ số tự tương quan trong tổng số m độ trễ đồng thời bằng không]. Điều này có thể được thực hiện với sự hỗ trợ của thống kê Q (**Q statistic**) do Box và Pierce đưa ra, công thức tính như sau:

$$Q = n \sum_{k=1}^m \hat{p}_k^2 \quad (13.3)$$

Ở đây, n là cỡ mẫu (2.355 trong ví dụ của chúng ta), và m là tổng số độ trễ được dùng để tính ACF, 30 trong ví dụ của chúng ta. [Diễn giải: Ví dụ, với $m = 10$, thống kê $Q = 23228$, và xác suất là 0.000; như vậy ta có thể bác bỏ giả thuyết cho rằng tất cả các hệ số trong tổng số 10 hệ số tự tương quan đầu tiên bằng không; nghĩa là có ít nhất một hệ số khác không. Nhưng ở đây, chắc chắn là tất cả 10 hệ số đều khác không]. Thống kê Q thường được sử dụng để kiểm định xem liệu rằng một chuỗi thời gian có phải là nhiễu trắng, hoặc ngẫu nhiên thuần túy hay không. [Diễn giải: Nghĩa là chúng ta không dùng để kiểm định xem một hệ số nào đó có khác không hay không, vì để kiểm định một hệ số nào đó có khác không hay không, ta sử dụng thống kê t , như được trình bày theo ý tưởng của Barlett].

Trong các mẫu lớn, Q có phân phối xấp xỉ phân phối Chi bình phương (chi-square) với m bậc tự do. Nếu trong một ứng dụng cụ thể, giá trị Q tính toán (như trong kết quả Eviews

và Stata ở trên) lớn hơn giá trị Q phê phán (tra bảng) từ phân phối Chi bình phương tại một mức ý nghĩa được chọn (ví dụ 5%), thì chúng ta bác bỏ giả thuyết H_0 cho rằng tất cả các hệ số tự tương quan tổng thể, p_k , đều bằng 0; tức là, ít nhất một vài hệ số trong số đó phải khác 0.

Cột thứ 5 của Bảng 13.2 (hoặc cột cuối trong kết quả Eviews ở trang 8) cho ta biết giá trị xác suất p của thống kê Q. Như bảng này cho thấy, giá trị Q cho tới 30 độ trễ là 67.666 và xác suất để giá trị đó bằng 0. Nghĩa là, chuỗi thời gian của chúng ta (LEX) là chuỗi không dừng.

Tóm lại, có một bằng chứng rất mạnh cho rằng chuỗi tỷ giá đôla/euro là một chuỗi không dừng.

13.4 Kiểm định nghiệm đơn vị về tính dừng

Không cần đi chi tiết vào khía cạnh kỹ thuật, chúng ta có thể diễn đạt kiểm định nghiệm đơn vị cho ví dụ về tỷ giá đôla/euro như sau⁷:

$$\Delta \text{LEX}_t = B_1 + B_2 t + B_3 \text{LEX}_{t-1} + u_t \quad (13.4)$$

Trong đó, $\Delta \text{LEX}_t = \text{LEX}_t - \text{LEX}_{t-1}$, nghĩa là, sai phân bậc 1 của chuỗi log của tỷ giá hối đoái, t là thời gian hay biến xu thế có giá trị 1, 2, ... cho đến hết mẫu dữ liệu, và u_t là hạng nhiễu. [Diễn giải: Tùy vào bản chất của ΔLEX_t (bằng cách xem đồ thị dạng line graph), có thể có ba dạng phương trình kiểm định nghiệm đơn vị DF: (1) không có hằng số B_1 , (2) có hằng số B_1 , nhưng không có xu thế, và dạng thứ (3) như ở phương trình (13.4). Phương trình trên được ước lượng bằng OLS vì chúng ta kỳ vọng rằng LEX_t không dừng, nhưng sau khi lấy sai phân bậc thì chuỗi sai phân là một chuỗi dừng. Vấn đề này sẽ được bàn luận ở phần sau].

Diễn đạt bằng lời, chúng ta hồi quy sai phân bậc một của chuỗi log của tỷ giá theo biến xu thế và giá trị trễ một giai đoạn của chuỗi tỷ giá.

Giả thuyết không (H_0) là B_3 , hệ số của LEX_{t-1} , bằng 0. Đây được gọi là giả thuyết nghiệm đơn vị (unit root hypothesis)⁸. Giả thuyết khác (H_1) là $B_3 < 0$ ⁹. Không bác bỏ giả thuyết không sẽ gợi ra rằng chuỗi thời gian đang xem xét là một chuỗi không dừng.

Có vẻ như chúng ta có thể kiểm định giả thuyết không $B_3 = 0$ bằng kiểm định t thông thường. Không may, chúng ta không thể làm như thế bởi vì kiểm định t chỉ có hiệu lực nếu chuỗi thời gian cơ bản phải là chuỗi dừng. Tuy nhiên, chúng ta có thể sử dụng một kiểm định được phát triển bởi hai nhà thống kê Dickey và Fuller, gọi là **thống kê τ** (tau test) với các giá trị phê phán của nó được tính toán bằng cách chạy mô phỏng và các

⁷ Thảo luận chi tiết hơn, xem Gujarati/Porter (2009), Chương 21.

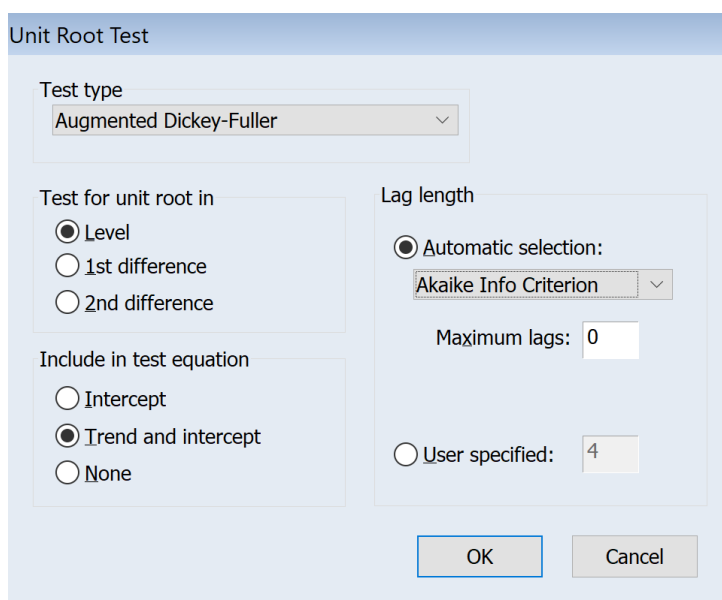
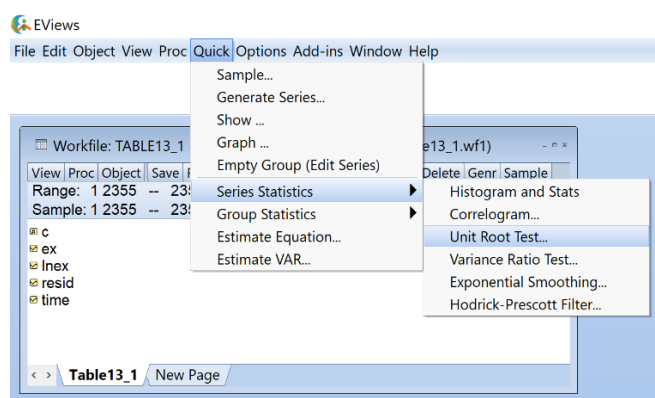
⁸ Để biết về mặt cảm quan tại sao thuật ngữ *nghiệm đơn vị* được sử dụng, chúng ta có thể tiến hành như sau: $\text{LEX}_t = B_1 + B_2 t + B_3 \text{LEX}_{t-1} + u_t$. Bây giờ trừ hai vế của phương trình này cho LEX_{t-1} , ta có: $(\text{LEX}_t - \text{LEX}_{t-1}) = B_1 + B_2 t + C \text{LEX}_{t-1} - \text{LEX}_{t-1} + u_t$, sau đó ta viết lại như sau $\Delta \text{LEX}_t = B_1 + B_2 t + (C-1) \text{LEX}_{t-1} + u_t = B_1 + B_2 t + B_3 \text{LEX}_{t-1} + u_t$, trong đó $B_3 = C - 1$. Như vậy, nếu C thực sự bằng 1, thì B_3 trong phương trình (13.4) sẽ bằng 0, và vì thế ta có tên gọi là *nghiệm đơn vị*.

⁹ Chúng ta loại bỏ khả năng $B_3 > 0$, vì trong trường hợp đó $C > 1$, như vật chuỗi thời gian cơ bản là chuỗi bùng phát (explosive time series). [Diễn giải, chuỗi bùng phát nghĩa là một chuỗi tăng rất nhanh qua thời gian, điều này chỉ có thể xảy ra trong trường hợp ta khảo sát một giai đoạn rất ngắn của một chuỗi thời gian].

phần mềm thống kê hiện đại, chẳng hạn như Eviews và Stata, đều cung cấp sẵn trong kết quả kiểm định. Trong lý thuyết, kiểm định τ được biết như **kiểm định Dickey-Fuller**, viết tắt là DF.

Trong thực tế, chúng ta ước lượng phương trình (13.4) bằng OLS, thấy giá trị t tính toán thông thường của hệ số LEX_{t-1} ($=B_3$) [Diễn giải: Đúng ra là b_3 mới chính xác, vì B_3 là tham số tổng thể, không quan sát được. Ở dưới đây, tôi xin mạo muội đổi B_3 thành b_3 nếu tôi nghĩ Gurajati đang hàm ý về hệ số ước lượng từ dữ liệu mẫu], nhưng sử dụng *các giá trị phê phán DF* để xem giá trị t tính toán có lớn hơn giá trị phê phán DF hay không. Nếu trong một ứng dụng cụ thể, giá trị t tính toán ($=\tau$) của hệ số ước lượng b_3 lớn hơn (về mặt giá trị tuyệt đối) giá trị phê phán DF, chúng ta bác bỏ giả thuyết nghiệm đơn vị - nghĩa là, chúng ta kết luận rằng chuỗi thời gian đang xem xét là một chuỗi dừng. Trong trường hợp đó, kiểm định t thông thường là có hiệu lực. Trái lại, nếu nó nhỏ thua giá trị τ phê phán, chúng ta không bác bỏ giả thuyết nghiệm đơn vị và kết luận rằng chuỗi thời gian là một chuỗi không dừng. Lý do phải xem xét giá trị tuyệt đối là vì nói chung các hệ số b_3 được kỳ vọng là mang dấu âm¹⁰.

Chúng ta hãy quay trở lại ví dụ minh họa. Các kết quả ước lượng phương trình (13.4) được cho trong Bảng 13.3. [Diễn giải: Bảng kết quả này từ Eviews].



¹⁰ Lưu ý $B_3 = (C - 1)$. Nên nếu $C < 1$, $B_3 < 0$.

Bảng 13.3: Kiểm định nghiệm đơn vị của tỷ giá đôla/euro.

Dependent Variable: D(LNEX)				
Method: Least Squares				
Date: 10/08/17 Time: 21:01				
Sample (adjusted): 2 2355				
Included observations: 2354 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNEX(-1)	-0.004088	0.001351	-3.026490	0.0025
C	-0.000846	0.000292	-2.897774	0.0038
@TREND("1")	1.21E-06	3.22E-07	3.761596	0.0002
R-squared	0.005995	Mean dependent var		0.000113
Adjusted R-squared	0.005149	S.D. dependent var		0.005926
S.E. of regression	0.005911	Akaike info criterion		-7.422695
Sum squared resid	0.082147	Schwarz criterion		-7.415349
Log likelihood	8739.512	Hannan-Quinn criter.		-7.420020
F-statistic	7.089627	Durbin-Watson stat		1.999138
Prob(F-statistic)	0.000852			

Nhìn vào hệ số của LEX trễ một giai đoạn [LNEX(-1)]. Giá trị t (=tau) của nó là -3.0265. Nếu bạn nhìn vào giá trị xác suất p tính toán thông thường của hệ số này, 0.0025, giá trị này rất thấp. Vì thế, bạn có thể sẽ dễ bị cám dỗ mà vội đi đến kết luận rằng hệ số ước lượng -0.004 là khác không một cách có ý nghĩa thống kê và vì thế chuỗi tỷ giá US/EU là chuỗi dừng¹¹.

[Diễn giải: Eviews hoặc Stata đều cung cấp sẵn các giá trị phê phán tại các mức ý nghĩa lần lượt 10%, 5%, và 1% tương ứng với phương trình kiểm định được chọn và một cỡ mẫu cho trước].

Null Hypothesis: LNEX has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on AIC, maxlag=0)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.026490	0.1251
Test critical values: 1% level	-3.961944	
5% level	-3.411717	
10% level	-3.127739	

*MacKinnon (1996) one-sided p-values.

Tuy nhiên, các giá trị phê phán DF lần lượt là: -3.9619 (mức ý nghĩa 1%), -3.4117 (mức ý nghĩa 5%), và -3.1277 (mức ý nghĩa 10%). Giá trị t tính toán là -3.0265. Theo giá trị tuyệt đối, 3.0265 nhỏ hơn bất cứ các giá trị phê phán DF nào (như trong bảng trên). Vì thế, chúng ta kết luận rằng chuỗi tỷ giá US/EU là một chuỗi không dừng.

Nói một cách khác, để bác bỏ giả thuyết không của nghiệm đơn vị, giá trị t tính toán của LEX_{t-1} phải âm nhiều hơn bất kỳ giá trị phê phán DF nào. Trên cơ sở của giá trị phê phán DF, thì xác suất để có một giá trị tau (=t) là -3.0265 là khoảng 12% (xem bảng trên). Như bạn có thể thấy từ bảng kết quả hồi quy ở trên, thống kê t thông thường cho thấy giá

¹¹ Trong trường hợp này, $(C-1) = -0.004$, tức $C = 0.996$, nó thực sự không chính xác bằng 1. Điều này sẽ cho rằng chuỗi LEX là chuỗi dừng.

trị -3.0264 có ý nghĩa tại mức ý nghĩa 0.0025. Điều này là bằng chứng cho ta thấy rằng mức ý nghĩa theo tính toán thông thường của giá trị t ước lượng có thể rất sai lầm khi áp dụng cho một chuỗi thời gian không dừng.

[Diễn giải: Kiểm định nghiệm đơn vị trên Stata như sau:]

```
. dfuller lnex, trend regress lags(0)
```

Dickey-Fuller test for unit root		Number of obs =		2354	
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value		
Z(t)	-3.026	-3.960	-3.410	-3.120	
MacKinnon approximate p-value for Z(t) = 0.1248					

D.lnex	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnex						
L1.	-.004088	.0013508	-3.03	0.003	-.0067368	-.0014392
_trend	1.21e-06	3.22e-07	3.76	0.000	5.79e-07	1.84e-06
_cons	-.0008456	.0002918	-2.90	0.004	-.0014178	-.0002734

Một vài khía cạnh thực tế của kiểm định DF

Kiểm định DF có thể được thể hiện trong ba hình thức khác:

Bước ngẫu nhiên (random walk):

$$\Delta \text{LEX}_t = B_3 \text{LEX}_{t-1} + u_t \quad (13.5)$$

Bước ngẫu nhiên có hệ số cắt (random walk with drift):

$$\Delta \text{LEX}_t = B_1 + B_3 \text{LEX}_{t-1} + u_t \quad (13.6)$$

Bước ngẫu nhiên có hệ số cắt quanh một xu thế tất định (random walk with drift around a deterministic trend):

$$\Delta \text{LEX}_t = B_1 + B_2 t + B_3 \text{LEX}_{t-1} + u_t \quad (13.7)$$

Trong mỗi trường hợp, giả thuyết không đều là $B_3 = 0$ (tức có nghiệm đơn vị) và giả thuyết khác là $B_3 < 0$ (tức không có nghiệm đơn vị). Tuy nhiên, các giá trị phê phán DF là khác nhau cho mỗi mô hình kiểm định này. Mô hình nào thích hợp trong một ứng dụng là một câu hỏi thực nghiệm. [Diễn giải: Tùy vào dữ liệu sai phân xoay quanh số không, hay xoay quanh một giá trị khác không; tùy vào dữ liệu sai phân có xu thế hay không. Cho nên, nếu cẩn thận bạn nên vẽ đồ thị trước khi tiến hành kiểm định nghiệm đơn vị]. Nhưng hãy lưu ý các lỗi do xác định sai dạng mô hình. Nếu mô hình (13.7) là mô hình đúng, thì việc chọn hoặc mô hình (13.5) hoặc mô hình (13.6) sẽ dẫn đến lỗi do xác định sai dạng mô hình: ở đây, tức là lỗi do bỏ sót biến (các biến quan trọng). [Diễn giải: Như vậy sẽ có khả năng hệ số ước lượng của b_3 bị chệch].

Chúng ta nên sử dụng phương trình nào trong số các phương trình (13.5), (13.6), và (13.7) trong thực tế? Đây là một số hướng dẫn¹²:

¹² Xem R. Carter Hill, William F. Griffiths và Guay C. Lim, *Principles of Econometrics*, 3rd edn, John Wiley & Sons, New York, 2008, p.336.

1. Sử dụng phương trình (13.5) nếu chuỗi thời gian dao động xung quanh trung bình mẫu là không. [Diễn giải: Theo tôi sai phân bậc một của chuỗi thời gian thì chính xác hơn].
2. Sử dụng phương trình (13.5) nếu chuỗi thời gian dao động xung quanh trung bình mẫu là khác không. [Diễn giải: Theo tôi sai phân bậc một của chuỗi thời gian thì chính xác hơn].
3. Sử dụng phương trình (13.5) nếu chuỗi thời gian dao động xung quanh một xu thế tuyến tính. Đôi khi xu thế có thể là dạng hàm bậc hai. [Diễn giải: Theo tôi sai phân bậc một của chuỗi thời gian thì chính xác hơn].

Trong lý thuyết, mô hình (13.5) được gọi là mô hình bước ngẫu nhiên không có hệ số cắt [random walk model without drift (no intercept)], mô hình (13.6) được gọi là mô hình bước ngẫu nhiên có hệ số cắt, B_1 là hệ số dịch chuyển (drift hoặc shift parameter), và mô hình (13.7) là mô hình bước ngẫu nhiên có hệ số cắt và có xu thế tất định, được gọi như vậy bởi vì một giá trị xu thế tất định B_2 được cộng thêm vào cho mỗi giai đoạn thời gian. Chúng ta sẽ nói nhiều hơn về xu thế tất định ở phần tiếp theo.

Chúng ta hãy xem liệu rằng hồi quy (13.7) có thể hiện được đặc điểm của chuỗi LEX. Các kết quả được cho trong Bảng 13.4.

Bảng 13.4: Kiểm định nghiệm đơn vị tỷ giá đôla/euro với hệ số cắt và xu thế.

Null Hypothesis: LNEX has a unit root				
Exogenous: Constant, Linear Trend				
Lag Length: 0 (Automatic - based on AIC, maxlag=0)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-3.026490	0.1251
Test critical values:	1% level		-3.961944	
	5% level		-3.411717	
	10% level		-3.127739	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LNEX)				
Method: Least Squares				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNEX(-1)	-0.004088	0.001351	-3.026490	0.0025
C	-0.000846	0.000292	-2.897774	0.0038
@TREND("1")	1.21E-06	3.22E-07	3.761596	0.0002
R-squared	0.005995	Mean dependent var		0.000113
Adjusted R-squared	0.005149	S.D. dependent var		0.005926
S.E. of regression	0.005911	Akaike info criterion		-7.422695
Sum squared resid	0.082147	Schwarz criterion		-7.415349
Log likelihood	8739.512	Hannan-Quinn criter.		-7.420020
F-statistic	7.089627	Durbin-Watson stat		1.999138
Prob(F-statistic)	0.000852			

Lưu ý: @Trend là lệnh của Eviews để tạo ra biến xu thế. D là ký hiệu của Eviews cho lấy sai phân bậc một.

Kết quả Eviews được cho trong bảng này được chia thành hai phần. Phần dưới là kết quả hồi quy OLS thông thường của phương trình (13.7). Kết quả này cho thấy tất cả các hệ số ước lượng đều có ý nghĩa thống kê cao trên cơ sở kiểm định t và giá trị thống kê F cũng có ý nghĩa cao, điều này chỉ ra rằng tất cả các biến giải thích đồng thời ảnh hưởng có ý nghĩa lên LEX¹³.

Đối với các mục đích hiện tại, hệ số quan trọng là hệ số của biến trễ một giai đoạn của chuỗi LEX [tức LEX_{t-1}]. Giá trị t của hệ số này có ý nghĩa ở mức 0.0025, trong khi nếu bạn nhìn vào giá trị *tau* của hệ số này được cho trong nửa trên của bảng này, thì bạn thấy nó có ý nghĩa ở mức 0.125, mức này cao hơn nhiều so với các giá trị *tau* phê phán ở 1%, 5%, và 10%. Nói cách khác, trên cơ sở kiểm định *tau*, hệ số của biến trễ của LEX là không khác không, như vậy cho ta thấy rằng chuỗi LEX là chuỗi không dừng. Điều này khẳng định lại kết luận dựa trên phân tích đồ thị đơn giản cũng như giản đồ tự tương quan.

Bài tập này cho thấy các kiểm định t và F thông thường có thể mắc sai lầm nếu chúng ta dùng để xử lý dữ liệu chuỗi thời gian.

Kiểm định Dickey-Fuller mở rộng (ADF)

Trong các mô hình (13.5), (13.6), và (13.7), chúng ta giả định rằng hạng nhiễu u_t không có tương quan. [Diễn giải: Không có hiện tượng tương quan chuỗi]. Nhưng nếu hạng nhiễu tương quan, nhất là trong trường hợp của mô hình (13.7). Dickey và Fuller đã phát triển một kiểm định khác, được gọi là **kiểm định Dickey-Fuller mở rộng** (augmented Dickey-Fuller test). Kiểm định này được thực hiện bằng cách mở rộng ba phương trình trên bằng việc đưa thêm các biến trễ của biến phụ thuộc ΔLEX_t như sau:

$$\Delta LEX_t = B_1 + B_2t + B_3LEX_{t-1} + \sum_{i=1}^m \alpha_i \Delta LEX_{t-i} + \varepsilon_t \quad (13.8)$$

Trong đó, ε_t là nhiễu trắng và m là độ trễ tối đa của biến phụ thuộc, độ trễ này được xác định một cách thực nghiệm¹⁴. [Diễn giải: Cách xác định độ trễ tối ưu phổ biến là chọn độ trễ sao cho tối thiểu hóa một tiêu chí nào đó, ví dụ AIC, SIC, ... Nghĩa là nếu có hiện tượng tương quan chuỗi, khi tăng số độ trễ của biến phụ thuộc vào mô hình thì AIC hoặc SIC sẽ giảm, và đến một độ trễ nào đó thì AIC hoặc SIC sẽ bắt đầu tăng. Trong Eviews, việc xác định độ trễ tối ưu được thực hiện một cách tự động theo cách thức này]. Mục tiêu là làm cho phần dư từ phương trình (13.7) là một chuỗi ngẫu nhiên thuần túy.

Vẫn như kiểm định DF, giả thuyết không vẫn là B_3 trong phương trình (13.8) bằng không.

Đối với ví dụ minh họa của chúng ta, chúng ta sử dụng $m = 26$. Và kết luận chuỗi tỷ giá đôla/euro là chuỗi không dừng vẫn không thay đổi.

¹³ Chúng tôi cũng ước lượng mô hình với cả biến xu thế tuyến tính và xu thế bậc hai, nhưng biến xu xu thế bậc hai không có ý nghĩa thống kê, giá trị xác suất p của hệ số xu thế bậc hai là 26%.

¹⁴ Nhưng lưu ý rằng nếu chúng ta đưa quá nhiều độ trễ, chúng sẽ tiêu tốn rất nhiều bậc tự do, đây có thể là một vấn đề trong các mẫu nhỏ. Đối với dữ liệu theo năm, chúng ta có thể đưa một hoặc hai độ trễ, trong khi dữ liệu theo tháng chúng ta có thể đưa 12 độ trễ. Dĩ nhiên, mục đích đưa độ trễ của biến phụ thuộc ΔLEX là để làm cho hạng nhiễu không còn bị hiện tượng tương quan chuỗi (serial correlation).

Tóm lại, chúng ta có bằng chứng vững chắc để kết luận rằng chuỗi tỷ giá hối đoái đôla/euro là một chuỗi không dừng.

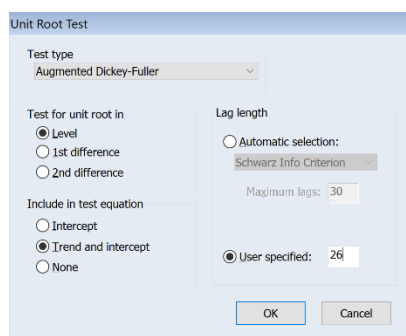
[Diễn giải: Bây giờ mình tập thực hành với Stata và Eviews thôi].

```
. dfuller lnex, trend regress lags(26)
```

Augmented Dickey-Fuller test for unit root Number of obs = 2328

Test Statistic	Interpolated Dickey-Fuller		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-2.846	-3.960	-3.410

MacKinnon approximate p-value for Z(t) = 0.1805



Null Hypothesis: LNEX has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 26 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.846115	0.1808
Test critical values:		
1% level	-3.961987	
5% level	-3.411739	
10% level	-3.127751	

*MacKinnon (1996) one-sided p-values.

Có một cách nào chúng ta có thể làm cho chuỗi tỷ giá đôla/euro trở thành chuỗi dừng không? (Có). Câu trả lời ở dưới đây.

13.5 Chuỗi dừng sai phân và chuỗi dừng xu thế

Như Hình 13.1 cho thấy, chuỗi tỷ giá đôla/euro nhìn chung là có xu thế đi lên. Một cách thực tế phổ biến để làm một chuỗi có xu thế như thế dừng là loại bỏ xu thế trong chuỗi dữ liệu. Điều này có thể được thực hiện bằng cách ước lượng hồi quy sau đây:

$$LEX_t = A_1 + A_2t + v_t \quad (13.9)$$

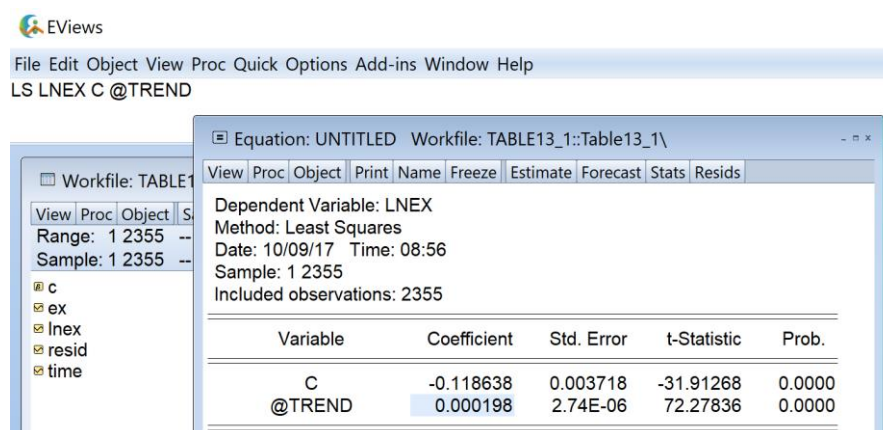
Trong đó, t (time) là một biến xu thế có giá trị theo thứ tự 1, 2, ..., 2355, và v_t là hạng nhiễu với các tính chất thông thường¹⁵. Sau khi chạy hồi quy này, chúng ta lưu phần dư:

$$\hat{v}_t = LEX_t - a_1 - a_2t \quad (13.10)$$

¹⁵ Một xu thế bậc hai cũng có thể được đưa vào mô hình.

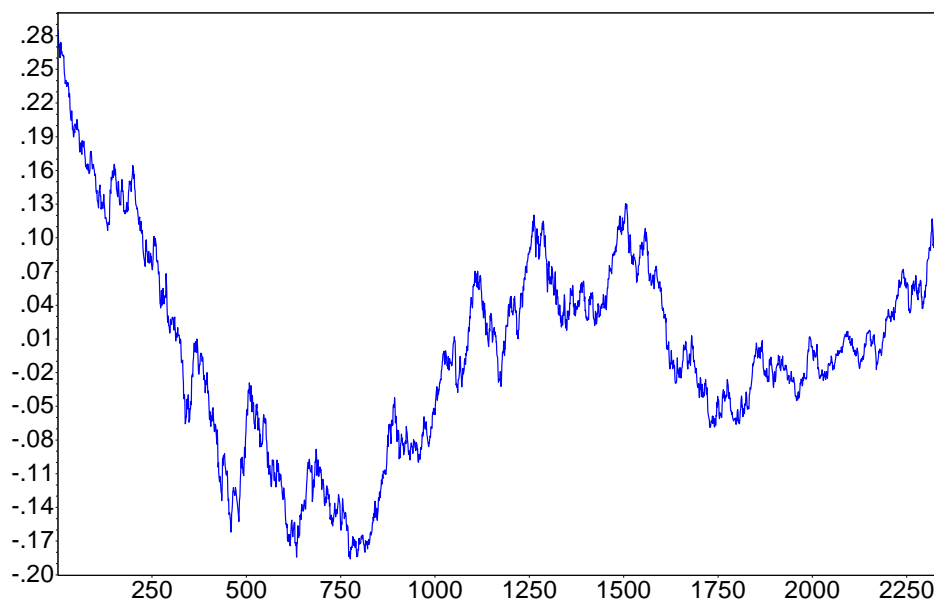
Hạng nhiễu ước lượng trong phương trình (13.10), \hat{u}_t , bây giờ thể hiện chuỗi LEX đã được khử xu thế (detrended), nghĩa là LEX với yếu tố xu thế đã được loại bỏ.

Thủ tục vừa được mô tả có hiệu lực nếu chuỗi LEX gốc có một xu thế tất định (deterministic trend). Phần dư thu được từ hồi quy phương trình (13.10) được trình bày ở Hình 13.3.



[Lưu ý: Biến 'resid' trong bảng trên chính là phần dư ngay sau ta khi thực hiện hồi quy phương trình (13.10). Sau đó, vẽ đồ thị của 'resid', chúng ta có được đồ thị như trong Hình 13.3].

Hình này rất giống với Hình 13.1. Nếu bạn thực hiện kiểm định nghiệm đơn vị chuỗi phần dư trong Hình 13.3, bạn sẽ nhận ra rằng chuỗi LEX khử xu thế (tất định) vẫn là chuỗi không dừng¹⁶. Vì thế, thủ tục khử xu thế như vừa mới thực hiện sẽ không làm cho một chuỗi không dừng thành một chuỗi dừng, bởi vì thủ tục như thế chỉ có hiệu lực nếu chuỗi thời gian có chứa một xu thế tất định. Vậy chúng ta sẽ làm gì đây?



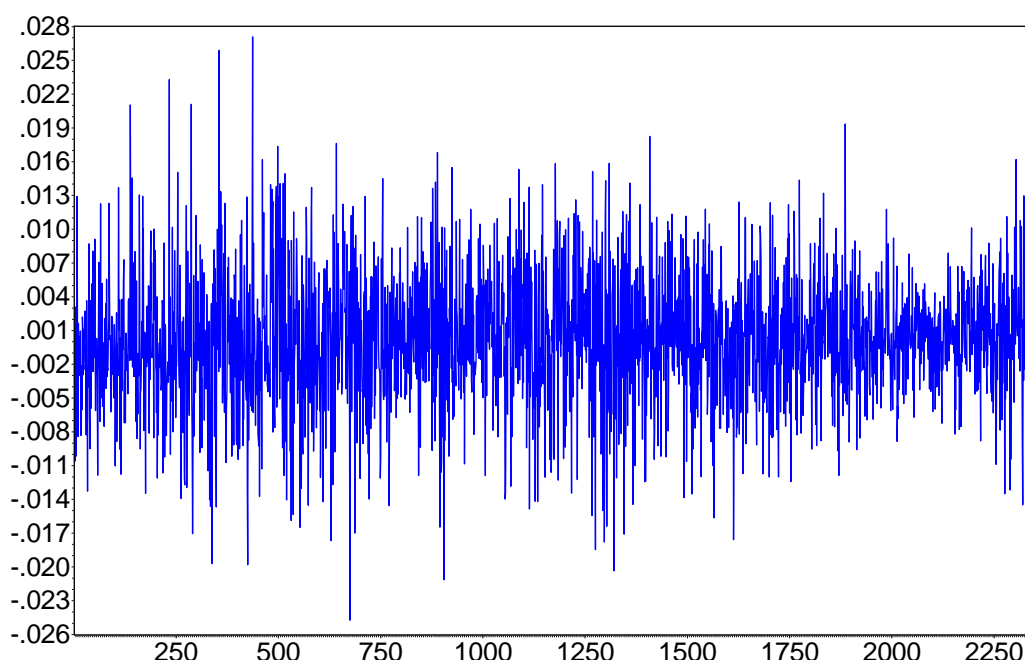
Hình 13.3: Phần dư từ hồi quy LEX theo thời gian.

¹⁶ Thậm chí nếu bạn đưa thêm xu thế bậc hai, t^2 , vào phương trình (13.9) thì phần dư từ hồi quy này vẫn cho thấy chúng là những chuỗi không dừng.

Nếu một chuỗi sẽ dừng nếu chúng ta khử xu thế của nó theo cách vừa mới được đề xuất, thì chuỗi đó được gọi là một quy trình ngẫu nhiên dừng xu thế [trend stationary (stochastic) process, (TSP)]. Chúng ta có thể chỉ ra ở đây rằng một quy trình có xu thế tất định là không dừng nhưng không phải là một quy trình có nghiệm đơn vị (unit root process).

Thay vì khử xu thế một chuỗi thời gian theo như cách vừa được đề xuất, giả sử chúng ta lấy sai phân bậc một (first differences) của LEX (nghĩa là $LEX_t - LEX_{t-1}$). Kết quả này cho ta đồ thị như trong Hình 13.4.

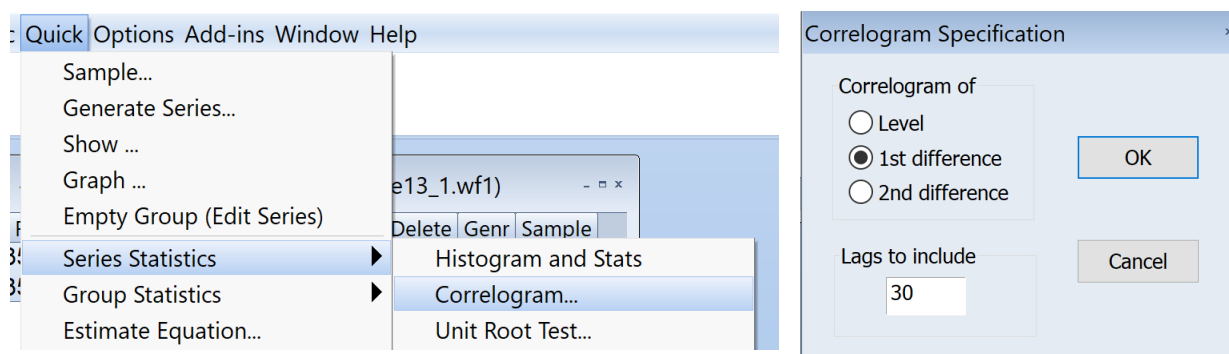
[Trong Eviews: PLOT D(LNEX)]



Hình 13.4: Sai phân bậc một của LEX.

Không giống Hình 13.1, chúng ta không thấy một xu thế rõ ràng trong chuỗi sai phân của LEX. Nếu chúng ta vẽ giản đồ tự tương quan cho chuỗi sai phân bậc một của LEX, chúng ta có kết quả như Bảng 13.5.

[Trong Eviews, ta không cần tạo biến sai phân. Thay vào đó ta chọn biến LNEX sau khi chọn Correlogram, và chọn '1st difference'].



Bảng 13.5: Giải đồ tự tương quan của sai phân bậc một của LEX.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.002	0.002	0.0113	0.915
		2	-0.001	-0.001	0.0125	0.994
		3	-0.017	-0.017	0.6673	0.881
		4	0.051	0.052	6.9213	0.140
		5	-0.036	-0.037	10.017	0.075
		6	0.016	0.016	10.643	0.100
		7	0.020	0.022	11.582	0.115
		8	-0.024	-0.028	12.970	0.113
		9	0.003	0.008	12.997	0.163
		10	-0.013	-0.015	13.379	0.203
		11	-0.003	-0.004	13.396	0.268
		12	0.012	0.016	13.735	0.318
		13	0.034	0.030	16.482	0.224
		14	-0.003	-0.001	16.501	0.284
		15	-0.032	-0.031	18.857	0.220
		16	0.011	0.010	19.140	0.261
		17	0.002	0.000	19.148	0.320
		18	0.021	0.022	20.222	0.320
		19	0.019	0.021	21.085	0.332
		20	0.022	0.017	22.193	0.330
		21	-0.035	-0.032	25.141	0.241
		22	0.041	0.041	29.088	0.142
		23	0.033	0.032	31.619	0.108
		24	0.038	0.037	35.079	0.067
		25	-0.007	-0.004	35.189	0.085
		26	0.008	0.001	35.341	0.104
		27	-0.015	-0.013	35.903	0.117
		28	-0.028	-0.027	37.786	0.103
		29	-0.014	-0.015	38.230	0.117
		30	0.012	0.010	38.570	0.136

Như bạn có thể thấy, cho đến 30 độ trễ, không có hệ số tự tương quan nào có ý nghĩa thống kê tại mức ý nghĩa 5% [Diễn giải: Tất cả hệ số nằm trong hai đường --- thể hiện khoảng tin cậy 95%]; thống kê Q cũng cho thấy tất cả 30 hệ số tự tương quan đồng thời không có ý nghĩa thống kê.

Một ứng dụng các kiểm định nghiệm đơn vị cũng cho thấy không có nghiệm đơn vị đối với chuỗi sai phân bậc một của LEX. Nghĩa là, sai phân bậc một của chuỗi LEX là một chuỗi dừng.

Null Hypothesis: D(LNEX) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 26 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-8.527415	0.0000
Test critical values: 1% level	-3.961989	
5% level	-3.411739	
10% level	-3.127752	

*MacKinnon (1996) one-sided p-values.

Nếu một chuỗi sẽ dừng sau khi chúng ta lấy sai phân bậc một, chúng ta gọi chuỗi như thế là một quy trình (ngẫu nhiên) dừng sai phân [difference stationary (stochastic) process, (DSP)]¹⁷.

Điều quan trọng cần lưu ý là nếu một chuỗi thời gian là DSP nhưng chúng ta xem nó là TSP, điều nhầm lẫn này được gọi là **lấy sai phân chưa tới** (under-differencing). Trái lại, nếu một chuỗi thời gian là TSP và chúng ta xử lý nó như DSP, điều nhầm lẫn này được gọi là **lấy sai phân quá mức** (over-differencing). Trong Hình 13.3, thực sự chúng ta lấy sai phân chưa tới cho chuỗi LEX.

Kết luận chính mà chúng ta rút ra ở đây là chuỗi tỷ giá đôla/euro LEX là một chuỗi dừng sai phân.

Chuỗi thời gian tích hợp (integrated time series)

Trong lý thuyết chuỗi thời gian, bạn sẽ thường gặp thuật ngữ “chuỗi tích hợp” [Diễn giải: Thông thường ta quen gọi là chuỗi dừng ở sai phân bậc mấy]. Nếu một chuỗi như thế trở thành chuỗi dừng sau khi lấy sai phân một lần, ta gọi đó là **tích hợp bậc một** (integrated of order one), ký hiệu là $I(1)$. Nếu phải lấy sai phân hai lần (tức là sai phân của sai phân) để trở thành chuỗi dừng, ta gọi đó là **tích hợp bậc hai** (integrated of order two), ký hiệu là $I(2)$. Nếu phải lấy sai phân d lần để thành chuỗi dừng, ta gọi đó là **tích hợp bậc d** , ký hiệu là $I(d)$. Một chuỗi dừng là $I(0)$, nghĩa là, tích hợp bậc 0. Vì thế, các thuật ngữ ‘chuỗi dừng’ và ‘chuỗi tích hợp bậc 0’ có nghĩa như sau. Vì vậy, nếu một chuỗi là tích hợp, có nghĩa đó là chuỗi không dừng.

Có thể nói thêm rằng một chuỗi $I(0)$ dao động xung quanh giá trị trung bình của nó với phương sai không đổi, trong khi một chuỗi $I(1)$ thì di chuyển rất ngoằn ngoèo. Một cách khác để diễn tả điều này là một chuỗi $I(0)$ có xu hướng **quay về giá trị trung bình** (mean reverting), trong khi một chuỗi $I(1)$ không cho thấy một xu hướng này. Nó có thể trôi dạt (drift away) xa dần giá trị trung bình ban đầu. Đó là lý do tại sao một chuỗi $I(1)$ được cho là có **xu thế ngẫu nhiên** (stochastic trend). Vì vậy, các hệ số tự tương quan trong giản đồ tự tương quan của một chuỗi $I(0)$ giảm về không rất nhanh khi ta gia tăng độ trễ, trong khi đối với một chuỗi $I(1)$ thì các hệ số tự tương quan giảm về không rất chậm, như giản đồ tự tương quan của chuỗi LEX trong Bảng 13.2 cho thấy rất rõ điều này.

¹⁷ Thỉnh thoảng chúng ta có thể phải lấy sai phân một chuỗi thời gian hơn một lần (tức sai phân bậc hai, bậc ba, ...) để làm cho chuỗi trở thành chuỗi dừng.

Hầu hết các chuỗi kinh tế không dừng nói chung là không cần phải lấy sai phân nhiều hơn một hoặc hai lần.

Tóm lại, một chuỗi thời gian không dừng được biết một cách khác như một chuỗi tích hợp hoặc một chuỗi với xu thế ngẫu nhiên.

Trước khi chúng ta kết luận chương này, chúng ta sẽ thảo luận ngắn gọn một loại đặc biệt của chuỗi không dừng được miêu tả rất nhiều trong lý thuyết tài chính, chuỗi này có tên là **bước ngẫu nhiên** (random walk time series).

13.6 Mô hình bước ngẫu nhiên (RWM)

Người ta thường nói rằng các giá tài sản, như giá chứng khoán và tỷ giá hối đoái, theo một bước ngẫu nhiên, nghĩa là, chúng là những chuỗi không dừng¹⁸. Chúng ta phân biệt hai loại bước ngẫu nhiên: (1) bước ngẫu nhiên không trôi dạt (random walk without drift, tức không có hằng số hay hệ số cắt) và (2) bước ngẫu nhiên trôi dạt (random walk with drift, tức có hằng số).

Bước ngẫu nhiên không có hằng số

Xem mô hình sau đây:

$$Y_t = Y_{t-1} + u_t \quad (13.11)$$

Trong đó, Y_t ví dụ là giá chứng khoán hôm nay và Y_{t-1} là giá chứng khoán hôm qua, và u_t là hạng nhiễu trắng với trung bình bằng 0 và phương sai là σ^2 .

Chúng ta có thể xem phương trình (13.11) như một hồi quy của Y tại thời điểm t theo giá trị trễ một giai đoạn của nó. Những người có niềm tin vào **giả thuyết thị trường hiệu quả** (efficient market hypothesis) khư khư bảo vệ quan điểm của mình rằng giá chứng khoán là ngẫu nhiên và vì thế không có chỗ cho đầu cơ sinh lợi trong thị trường chứng khoán¹⁹.

Bằng cách thay thế liên tục trong phương trình (13.11), phương trình này có thể được viết lại như sau:

$$Y_t = Y_0 + \sum u_t \quad (13.12)$$

Trong đó, Y_0 là giá chứng khoán ban đầu.

[Diễn giải: Ta thực hiện như sau:

$$\begin{aligned} Y_1 &= Y_0 + u_1 \\ Y_2 &= Y_1 + u_2 = Y_0 + u_1 + u_2 \\ Y_3 &= Y_2 + u_3 = Y_0 + u_1 + u_2 + u_3 \\ Y_4 &= Y_3 + u_4 = Y_0 + u_1 + \dots + u_4 \end{aligned}$$

¹⁸ Thuật ngữ bước ngẫu nhiên thường được so sánh với bước đi của anh chàng nát rượu. Sau khi rời quầy bar, gã say bước đi ngất ngưỡng một khoảng ngẫu nhiên u_t tại thời điểm t và, tiếp tục bước đi trong vô định, cuối cùng sẽ lang thang xa dần quầy bar. Giá chứng khoán cũng giống như thế. Giá chứng khoán hôm nay bằng giá chứng khoán hôm qua cộng một bước ngẫu nhiên.

¹⁹ Các nhà phân tích kỹ thuật (technical analysts), hoặc những nhà phân tích đầu tư bằng đồ thị (chartists) như cách họ được gọi, không tin và giả thuyết như vậy và tin rằng họ có thể dự đoán mô tiếp vận động của giá chứng khoán dựa vào giá chứng khoán được quan sát trong quá khứ.

$$\dots$$

$$Y_t = Y_{t-1} + u_t = Y_0 + u_1 + \dots + u_t]$$

Vì thế,

$$E(Y_t) = E(Y_0) + E(\sum u_t) = Y_0 \quad (13.13)$$

vì giá trị kỳ vọng của mỗi u_t là bằng không.

Bằng cách thay thế liên tục, ta có công thức phương sai như sau:

$$\text{Var}(Y_t) = t\sigma^2 \quad (13.14)$$

Từ phần thảo luận trước, chúng ta thấy rằng trung bình của Y bằng giá trị ban đầu của nó, hoặc giá trị khởi điểm, cố định, như khi t tăng lên vô cùng, thì phương sai của Y cũng tăng lên vô cùng, vì thế phá vỡ một trong số những điều kiện của tính dừng là phương sai không đổi qua thời gian.

Tóm lại, mô hình bước ngẫu nhiên không có hằng số là một trường hợp quan trọng và đặc biệt của một quy trình ngẫu nhiên không dừng.

Thật thú vị, nếu chúng ta viết phương trình (13.11) như sau:

$$Y_t - Y_{t-1} = \Delta Y_t = u_t \quad (13.15)$$

Trong đó Δ là toán tử sai phân bậc một.

Vì thế, mặc dù Y_t là chuỗi không dừng, nhưng sai phân bậc một của nó là một chuỗi dừng. Nói cách khác, mô hình bước ngẫu nhiên không có hằng số là một quy trình dừng sai phân.

Bước ngẫu nhiên có hằng số

Bây giờ chúng ta hãy sửa lại phương trình (13.11) và viết như sau:

$$Y_t = \delta + Y_{t-1} + u_t \quad (13.16)$$

Trong đó, δ được biết như **tham số dịch chuyển** (drift parameter), về cơ bản nó là hệ số cắt trong mô hình bước ngẫu nhiên.

Đối với mô hình bước ngẫu nhiên có hằng số, ta có thể viết:

$$E(Y_t) = Y_0 + \delta t \quad (13.17)$$

$$\text{Var}(Y_t) = t\sigma^2 \quad (13.18)$$

Như bạn có thể thấy, đối với mô hình bước ngẫu nhiên có hằng số thì cả trung bình và phương sai đều tăng qua thời gian, điều này lại phá vỡ các điều kiện của chuỗi dừng.

Chúng ta hãy viết lại phương trình (13.16) như sau:

$$Y_t - Y_{t-1} = \Delta Y_t = \delta + u_t \quad (13.19)$$

Đo là sai phân bậc một của mô hình bước ngẫu nhiên có hằng số. Ta dễ dàng kiểm tra các thống kê sau đây:

$$E(\Delta Y_t) = \delta \quad (13.20)$$

$$\text{Var}(\Delta Y_t) = \sigma^2 \quad (13.21)$$

$$\text{Cov}(\Delta Y_t, \Delta Y_{t-s}) = E(u_t, u_{t-s}) = 0 \quad (13.22)$$

bởi vì u_t là hạng nhiễu trắng.

Tất cả điều này có nghĩa rằng mặc dù mô hình bước ngẫu nhiên có hằng số là một chuỗi không dừng, nhưng sai phân bậc một của nó là một quy trình (ngẫu nhiên) dừng. Nói cách khác, mô hình bước ngẫu nhiên có hằng số là một quy trình $I(1)$, trong khi đó sai phân bậc một của nó là một quy trình $I(0)$. Ở đây, hằng số δ đóng vai trò như một xu thế tuyến tính bởi vì trong mỗi giai đoạn Y_t dịch chuyển, trung bình, bằng một lượng δ .

Một ví dụ minh họa: Giá đóng cửa hàng ngày của cổ phiếu IBM, giai đoạn 4/1/2000 – 20/8/2002

Để xem giá cổ phiếu IBM qua giai đoạn này có theo một bước ngẫu nhiên hay không, trước hết chúng ta vẽ đồ thị log của giá đóng cửa hàng ngày của cổ phiếu này, như trong Hình 13.5 (xem dữ liệu Table 13.6 trên trang web của cuốn sách).

Về mặt trực quan, dường như log của giá cổ phiếu IBM là chuỗi không dừng.

Chúng ta có thể kiểm tra điều này về mặt thống kê hay không? Bạn có thể sẽ bị thôi thúc chạy hồi quy sau đây (cho Y là log của giá đóng cửa hàng ngày của cổ phiếu IBM).

$$Y_t = B_1 + B_2 Y_{t-1} + u_t \quad (13.23)$$

Và kiểm định giả thuyết rằng $B_2 = 1$ bằng kiểm định t thông thường. Tuy nhiên, trong các trường hợp chuỗi thời gian không dừng, kiểm định t có xu hướng bị lệch về không. Để tránh điều này, chúng ta xử lý phương trình (13.23) như sau. Trừ hai vế của phương trình này cho Y_{t-1} , ta có:

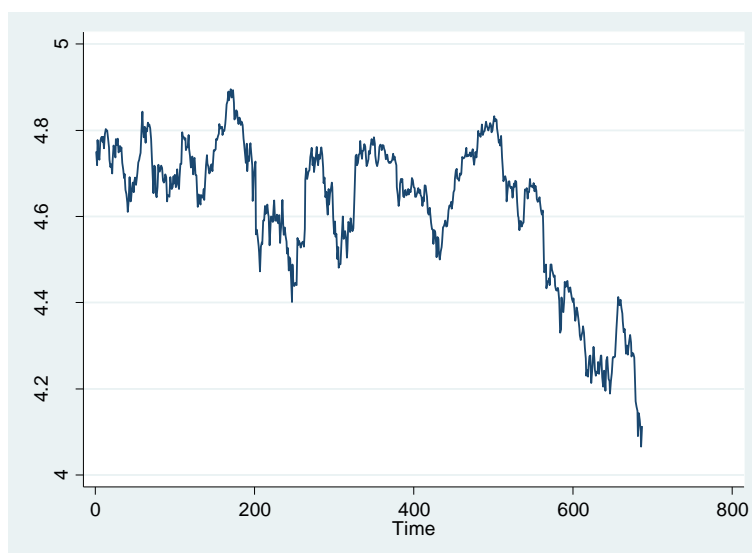
$$Y_t - Y_{t-1} = B_1 + B_2 Y_{t-1} - Y_{t-1} + u_t$$

Đó là

$$\Delta Y_t = B_1 + \lambda Y_{t-1} + u_t \quad (13.24)$$

Trong đó $\lambda = B_2 - 1$.

twoway (tsline lnclose)



Hình 13.5: Log của giá đóng cửa của cổ phiếu IBM.

Như thế thay vì ước lượng phương trình (13.23), chúng ta ước lượng phương trình (13.24) và kiểm định giả thuyết $\lambda = 0$ đối lại với giả thuyết khác $\lambda < 0$ ²⁰. Nếu $\lambda = 0$, thì $B_2 = 1$ và Y là một bước ngẫu nhiên (có hằng số), nghĩa là, đó là một chuỗi không dừng. Về mặt kỹ thuật, chuỗi Y có nghiệm đơn vị. Trái lại, nếu $\lambda < 0$, chúng ta có thể kết luận rằng Y_t là chuỗi dừng²¹.

Bảng 13.7: Kiểm định nghiệm đơn vị giá đóng cửa hàng ngày của cổ phiếu IBM

Null Hypothesis: LNCLOSE has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic - based on SIC, maxlag=0)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-1.026066	0.7455
Test critical values:	1% level		-3.439654	
	5% level		-2.865536	
	10% level		-2.568955	
*MacKinnon (1996) one-sided p-values.				
Dependent Variable: D(LNCLOSE)				
Method: Least Squares				
Date: 10/10/17 Time: 14:17				
Sample (adjusted): 1/04/2000 8/20/2002				
Included observations: 686 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNCLOSE(-1)	-0.006209	0.006051	-1.026066	0.3052
C	0.027766	0.027984	0.992236	0.3214
R-squared	0.001537	Mean dependent var		-0.000928
Adjusted R-squared	0.000077	S.D. dependent var		0.026385
S.E. of regression	0.026384	Akaike info criterion		-4.429201
Sum squared resid	0.476146	Schwarz criterion		-4.415991
Log likelihood	1521.216	Hannan-Quinn criter.		-4.424090
F-statistic	1.052811	Durbin-Watson stat		2.099601
Prob(F-statistic)	0.305223			

Sau khi bạn ước lượng phương trình (13.24), bạn không thể kiểm định giả thuyết không $\lambda = 0$ bằng kiểm định t thông thường bởi vì giá trị t của hệ số ước lượng của Y_{t-1} không theo phân phối t thậm chí trong các mẫu lớn.

Như đã lưu ý trước đây, trong các trường hợp giống như trường hợp này, chúng ta sử dụng thống kê τ của Dickey-Fuller, họ đưa ra công thức tính toán các giá trị phê phán, và từ đó được mở rộng bởi MacKinnon. Các giá trị phê phán MacKinnon bây giờ đã được đưa vào nhiều phần mềm kinh tế lượng.

Sử dụng Eviews 8, chúng ta có được các kết quả như trong Bảng 13.7. Phần thứ hai của Bảng này là kết quả hồi quy OLS thông thường. Giá trị t của hệ số hồi quy của biến trễ một giai đoạn của giá đóng cửa cổ phiếu IBM là -1.0026 với giá trị xác suất p khoảng 0.3, cho biết rằng hệ số này không khác không, và vì thế ủng hộ giả thuyết cho rằng giá đóng cửa của cổ phiếu IBM là một bước ngẫu nhiên hoặc chuỗi giá cổ phiếu IBM là một chuỗi không dừng.

²⁰ Về cơ bản chúng ta đang thực hiện một phân tích nghiệm đơn vị.

²¹ Nếu $\lambda = (B - 1)$, đối với chuỗi dừng thì B_2 phải nhỏ hơn 1. Để điều này xảy ra, λ phải âm.

Nếu bạn nhìn vào phần đầu của bảng kết quả này, bạn sẽ nhận thấy giá trị p của giá trị τ của Dickey-Fuller của hệ số của biến trễ giá cổ phiếu IBM là khoảng 0.75, một lần nữa ủng hộ giả thuyết bước ngẫu nhiên. Nhưng lưu ý rằng mức ý nghĩa của thống kê t thông thường và thống kê τ có thể khác nhau một cách đáng kể.

Sai phân bậc một của giá đóng cửa cổ phiếu IBM có dừng không?

Vì chúng ta biết rằng sai phân bậc một của log của giá đóng cửa cổ phiếu IBM là chuỗi dừng bởi vì sai phân bậc một của một mô hình bước ngẫu nhiên là chuỗi dừng, điều này có thể không làm chúng ta ngạc nhiên. Nếu bạn ước lượng giản đồ tự tương quan của sai phân bậc một, bạn sẽ nhận ra rằng các hệ số tương quan xoay quanh số không, đó điển hình là trường hợp của một chuỗi nhiễu trắng.

Nếu chúng ta thực hiện một phân tích nghiệm đơn vị chính thức, chúng ta có được kết quả như trong Bảng 13.8. Các kết quả này gợi ra rằng chúng ta có thể bác bỏ giả thuyết nghiệm đơn vị của chuỗi sai phân của log của chuỗi giá cổ phiếu IBM. Giá trị τ ước lượng ($=t$) mang dấu âm có ý nghĩa thống kê cao hơn giá trị τ phê phán ở mức ý nghĩa 1%. Trong trường hợp này, các thống kê τ và t là như nhau.

Bảng 13.8: Kiểm định nghiệm đơn vị của sai phân bậc một của giá đóng cửa hàng ngày cổ phiếu IBM

Null Hypothesis: D(LNCLOSE) has a unit root				
Exogenous: None				
Lag Length: 0 (Automatic - based on SIC, maxlag=0)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			-27.65371	0.0000
Test critical values:	1% level		-2.568342	
	5% level		-1.941286	
	10% level		-1.616388	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(LNCLOSE,2)				
Method: Least Squares				
Date: 10/10/17 Time: 14:23				
Sample (adjusted): 1/05/2000 8/20/2002				
Included observations: 685 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LNCLOSE(-1))	-1.057102	0.038226	-27.65371	0.0000
R-squared	0.527857	Mean dependent var		0.000116
Adjusted R-squared	0.527857	S.D. dependent var		0.038349
S.E. of regression	0.026351	Akaike info criterion		-4.433187
Sum squared resid	0.474941	Schwarz criterion		-4.426575
Log likelihood	1519.367	Hannan-Quinn criter.		-4.430629
Durbin-Watson stat	1.989376			

Chúng ta đã lưu ý trước đây rằng chúng ta không thể sử dụng một chuỗi thời gian không dừng cho các mục đích dự báo. Chúng ta có thể sử dụng các chuỗi sai phân của LEX hoặc giá cổ phiếu IBM để dự báo được không? Nếu được thì chúng ta làm thế nào để liên hệ việc dự báo chuỗi sai phân với chuỗi dữ liệu gốc (không lấy sai phân)? Chúng ta sẽ làm điều này ở chương sau (chương 16 về các mô hình ARIMA).

13.7 Tóm tắt và kết luận

Mặc dù chúng ta chỉ nghiên cứu hai chuỗi kinh tế tài chính, nhưng ý tưởng và kỹ thuật được thảo luận trong chương này có thể áp dụng cho nhiều chuỗi kinh tế và tài chính khác, đối với hầu hết các chuỗi dữ liệu kinh tế dưới dạng gốc (in level form) là không dừng. Các chuỗi như thế thường thể hiện các xu thế đi lên hoặc đi xuống qua một giai đoạn liên tục. Nhưng xu thế như thế thường mang tính ngẫu nhiên hơn là tất định. Điều này có các hàm ý quan trọng cho phân tích hồi quy, vì hồi quy một chuỗi không dừng theo một hay nhiều chuỗi không dừng thường dẫn đến hiện tượng hồi quy giả mạo hoặc hồi quy vô nghĩa. Như chúng ta sẽ chỉ ra ở chương sau, chỉ trong trường hợp các chuỗi đồng liên kết (đồng tích hợp) [cointegrated time series] thì mới tránh hiện tượng tương quan giả mạo, thậm chí nếu các chuỗi về cơ bản là không dừng.

Chúng ta đã xem ba công cụ phân tích chẩn đoán để tìm hiểu xem một chuỗi thời gian dừng hay không dừng. Công cụ đơn giản nhất trong số này là vẽ chuỗi thời gian theo thời gian. Một đồ thị như thế về chuỗi thời gian là một công cụ rất có giá trị để có cảm nhận sơ bộ về bản chất của chuỗi thời gian. Một cách chính thức hơn, chúng ta có thể phân tích giản đồ tự tương quan của chuỗi thời gian qua một số độ trễ. Giản đồ tự tương quan sẽ cho thấy các hệ số tự tương quan của chuỗi thời gian qua một số độ trễ sẽ giảm nhanh chóng hay giảm từ từ về không. Nếu nó giảm rất chậm về không, thì có lẽ chuỗi thời gian là không dừng.

Một kiểm định trở nên phổ biến là kiểm định nghiệm đơn vị. Nếu trên cơ sở kiểm định Dickey-Fuller hoặc kiểm định Dickey-Fuller mở rộng, chúng ta thấy có một hoặc nhiều nghiệm đơn vị trong một chuỗi thời gian, thì có lẽ nó cung cấp thêm bằng chứng về tính không dừng.

Vì mô hình hóa bằng phương pháp hồi quy truyền thống dựa trên giả định rằng chuỗi thời gian được sử dụng trong phân tích là các chuỗi dừng, nên điều quan trọng là chúng ta phải thực hiện các kiểm định tính dừng của chuỗi thời gian như vừa được thảo luận ở trên.

Nếu một chuỗi thời gian có xu thế tất định, nó có thể được chuyển sang chuỗi dừng bằng cách hồi quy chuỗi đó theo thời gian hoặc biến xu thế. Phần dư thu được từ hồi quy này sẽ thể hiện cho chuỗi thời gian không còn yếu tố xu thế.

Tuy nhiên, nếu một chuỗi thời gian có xu thế ngẫu nhiên, nó có thể được chuyển sang chuỗi dừng bằng cách lấy sai phân một hoặc nhiều lần./.