

Các khái niệm cơ bản về chuỗi thời gian (2)

Nội dung

- Chuỗi dừng và chuỗi không dừng (stationary and non-stationary).
- Hồi quy giả (spurious regression)
- Kiểm định chuỗi dừng.
- Chuyển đổi chuỗi dừng thành chuỗi không dừng.

Chuỗi Dừng và Chuỗi Không Dừng (Stationary and Non-stationary)

Chuỗi Dừng và Chuỗi Không Dừng

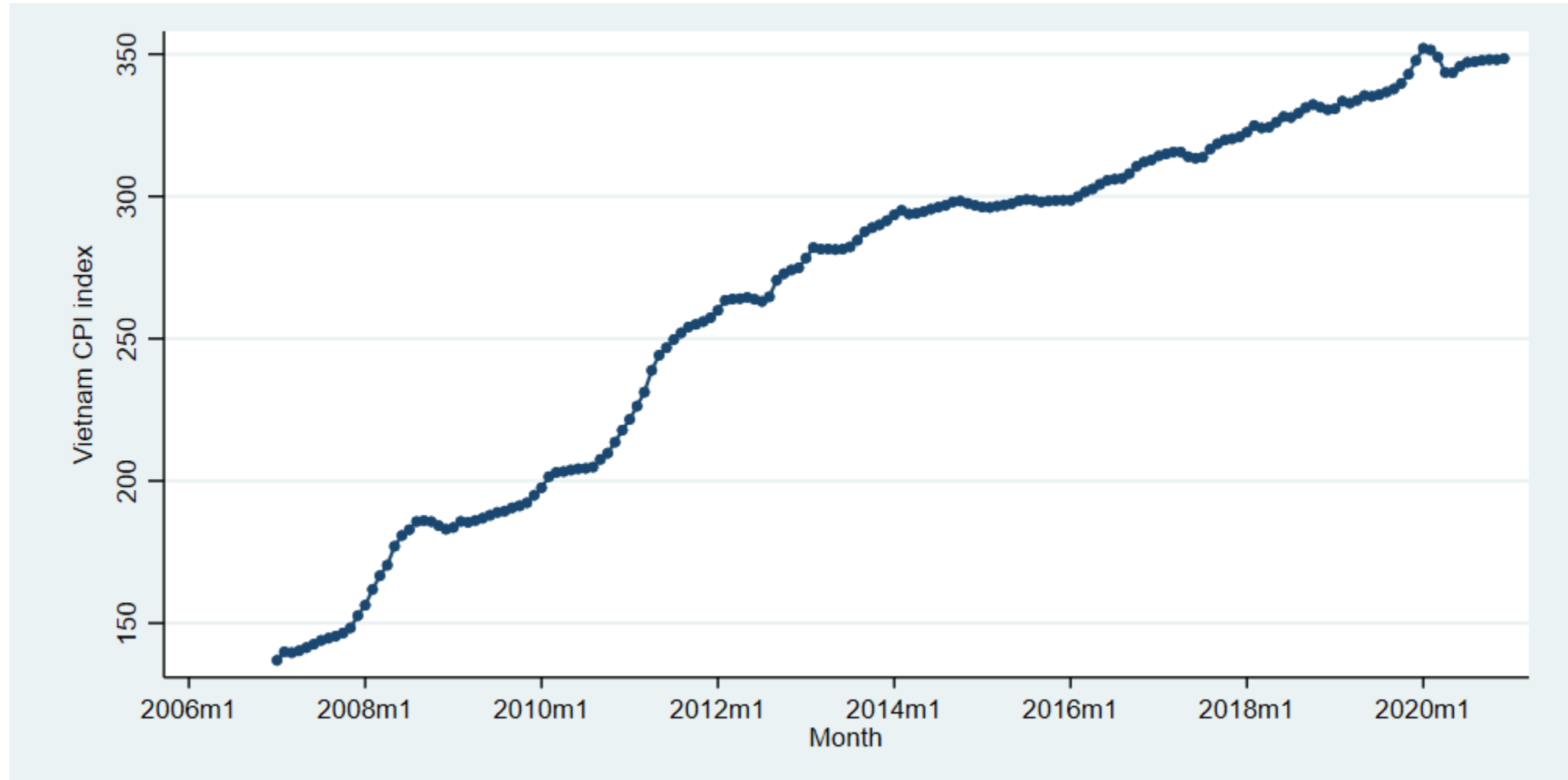
- Chuỗi dừng (stationary stochastic process) là một quá trình ngẫu nhiên mà giá trị trung bình và phương sai không thay đổi theo thời gian, cũng như hiệp phương sai (covariance) chỉ phụ thuộc thời khoảng cách thời gian.
- Về mặt toán học, ta có thể viết như sau:

$$E(y_t) = \mu, \quad t = 1, 2, 3 \dots, T$$

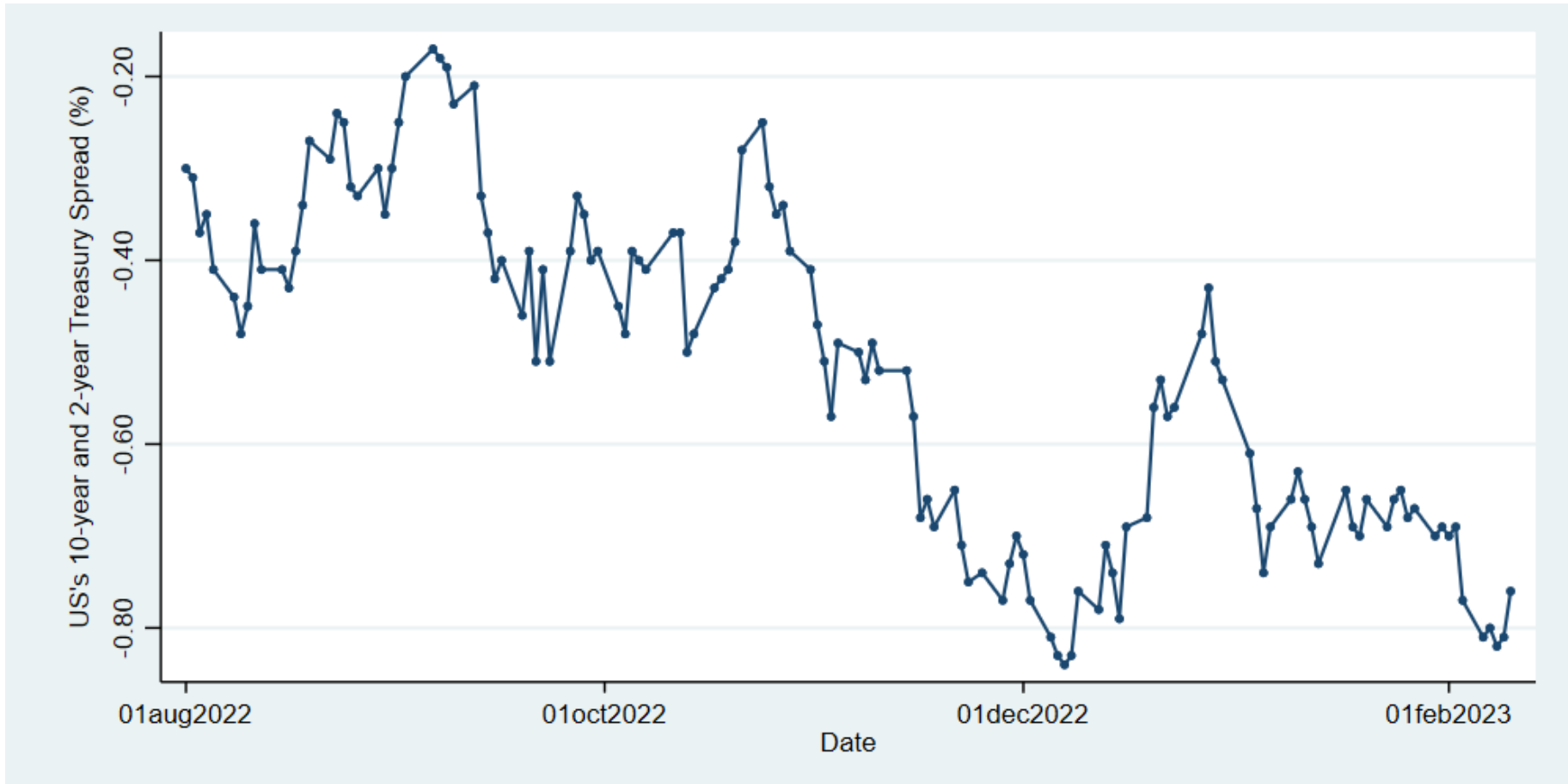
$$var(y_t) = \sigma^2, \quad t = 1, 2, 3 \dots, T$$

$$covariance(y_t, y_{t-k}) \text{ only depends on } k$$

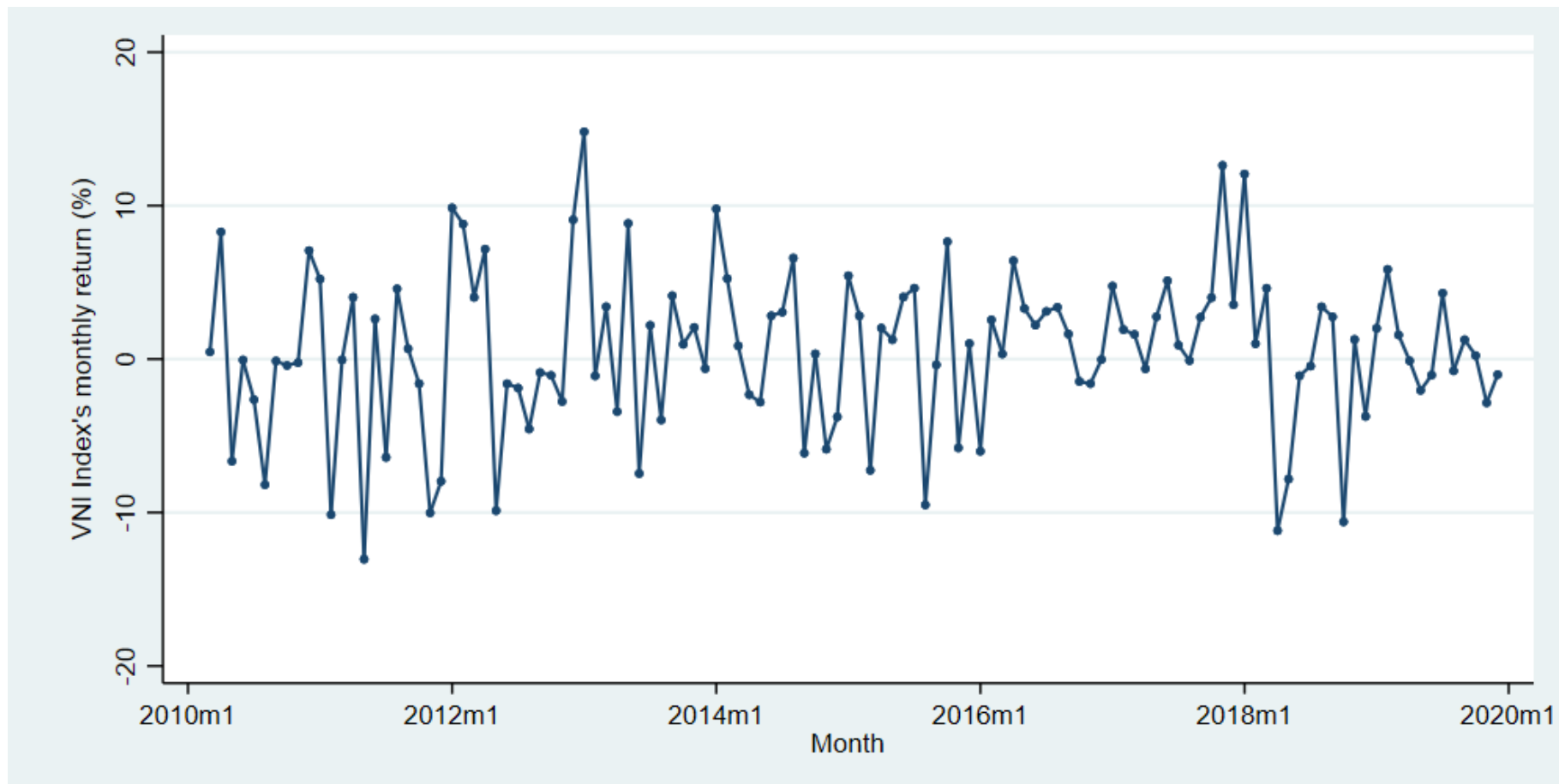
Minh họa (1)



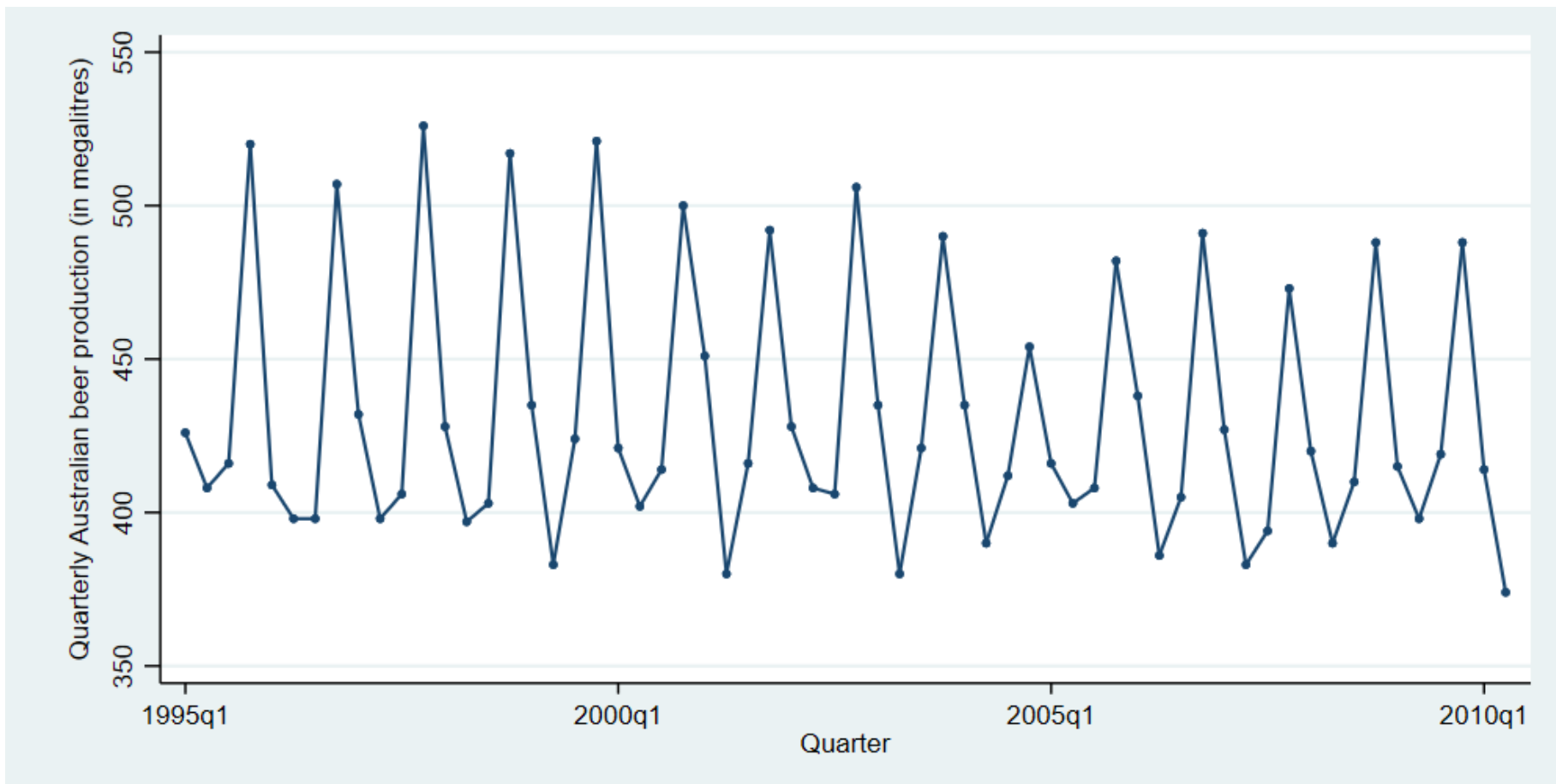
Minh họa (2)



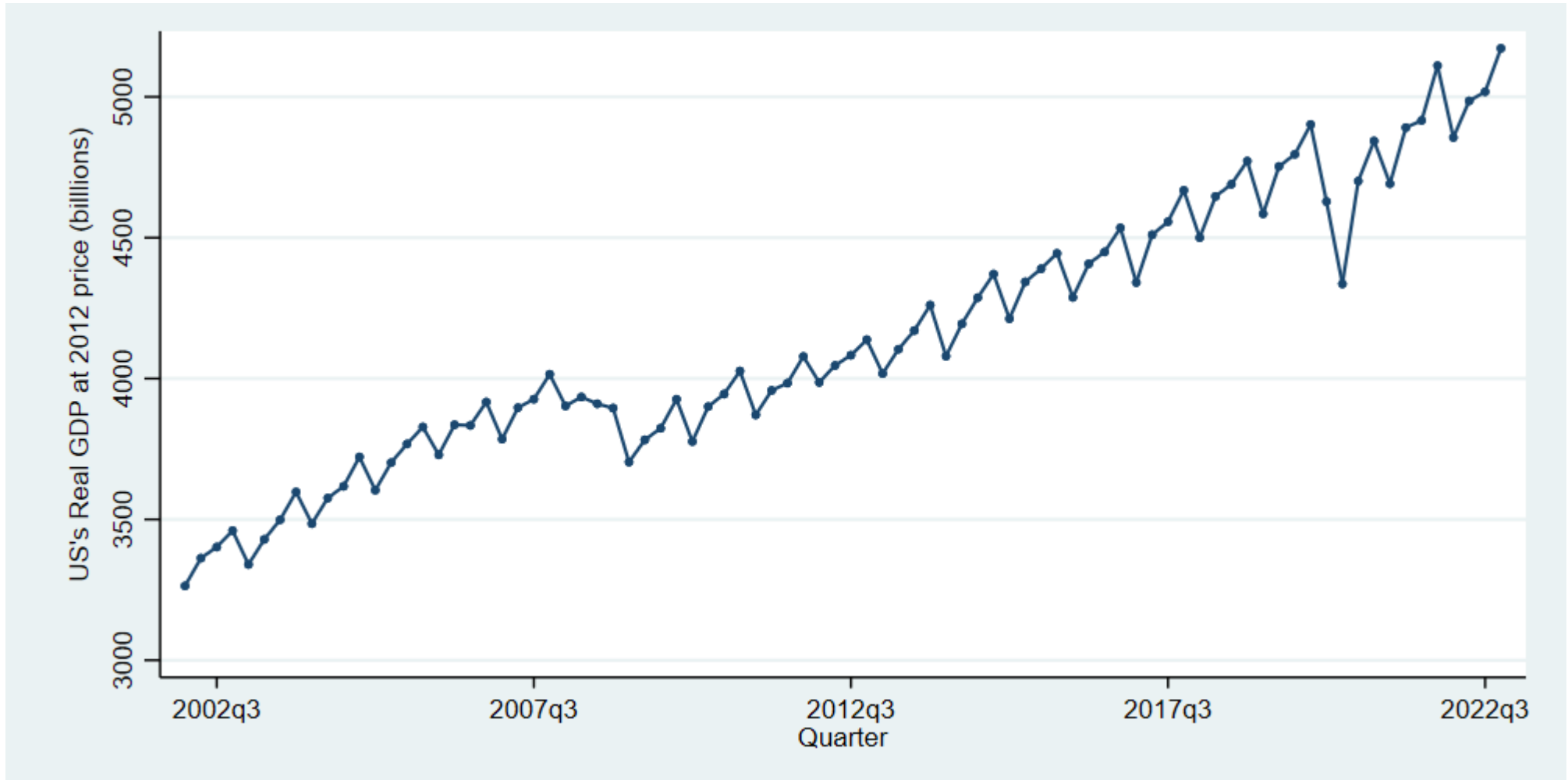
Minh họa (3)



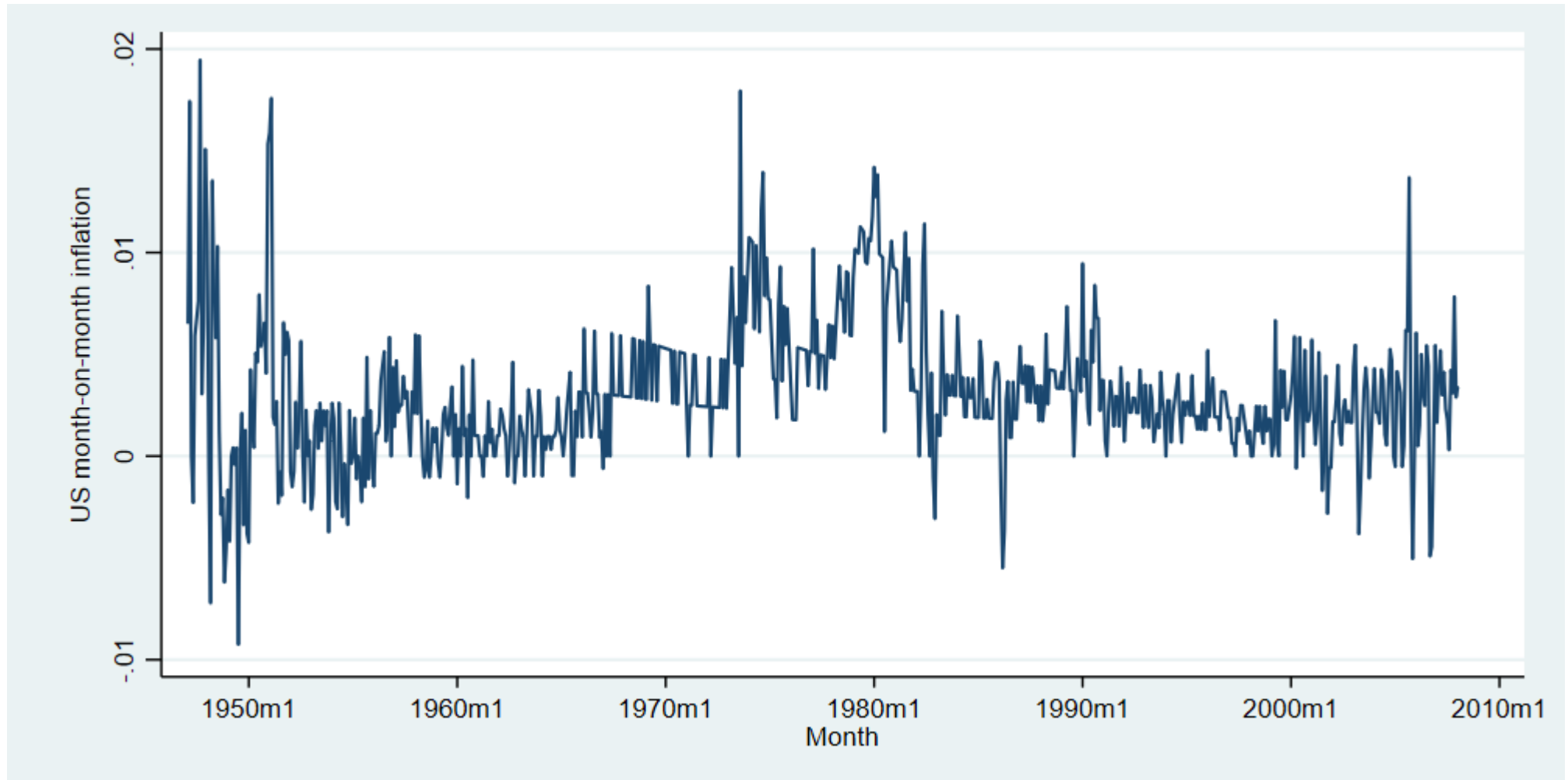
Minh họa (4)



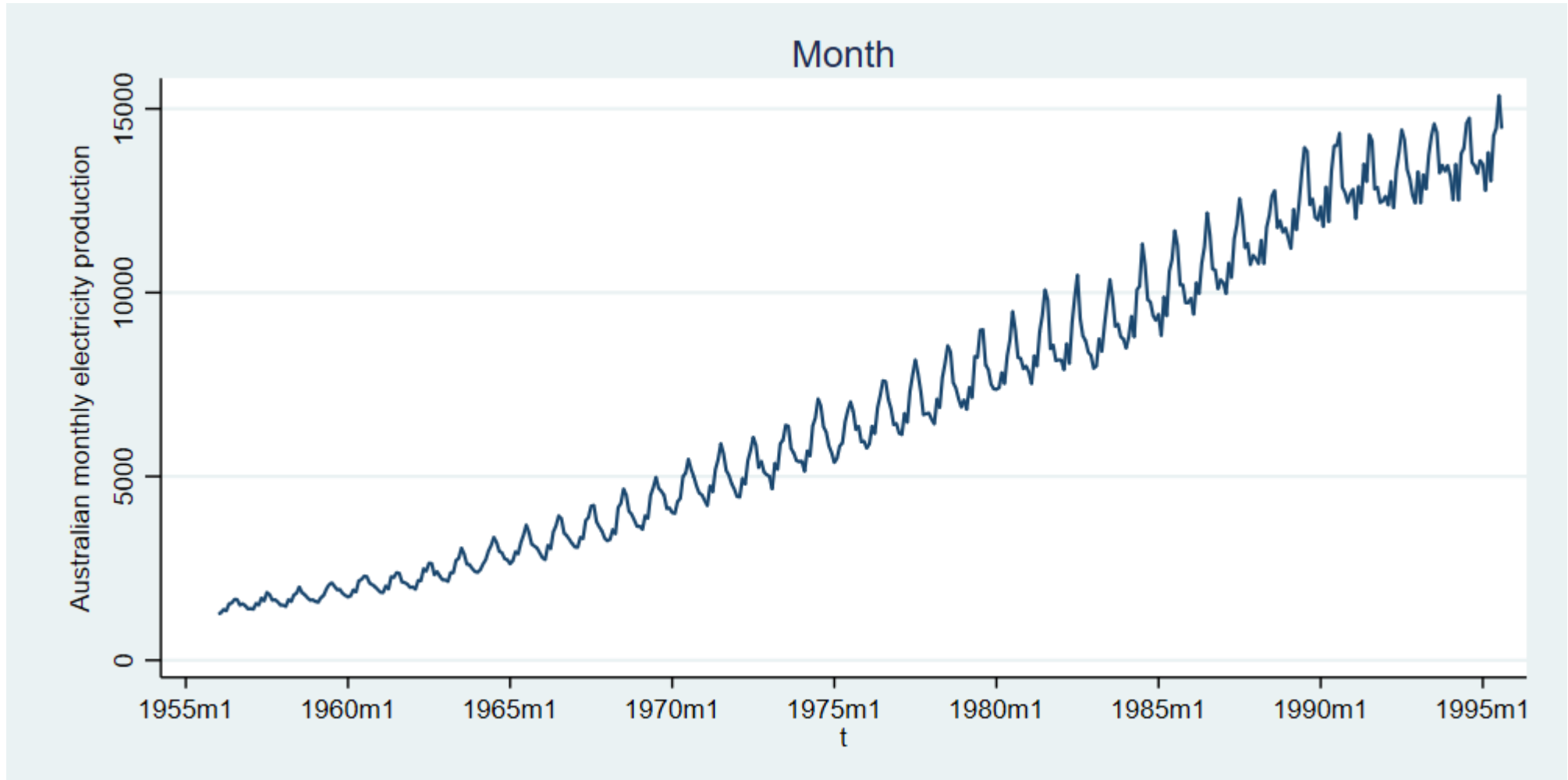
Minh họa (5)



Minh họa (6)



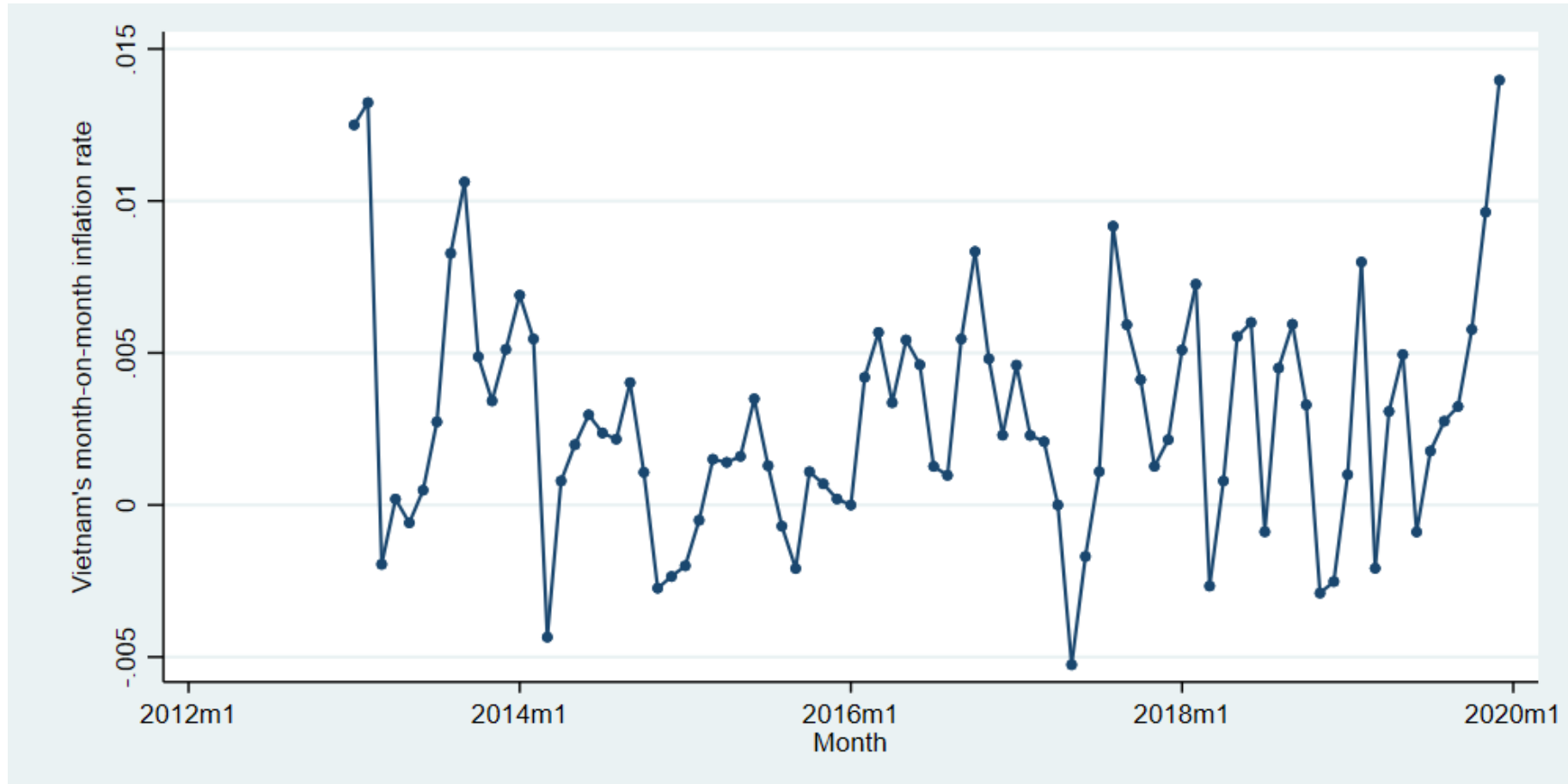
Minh họa (7)



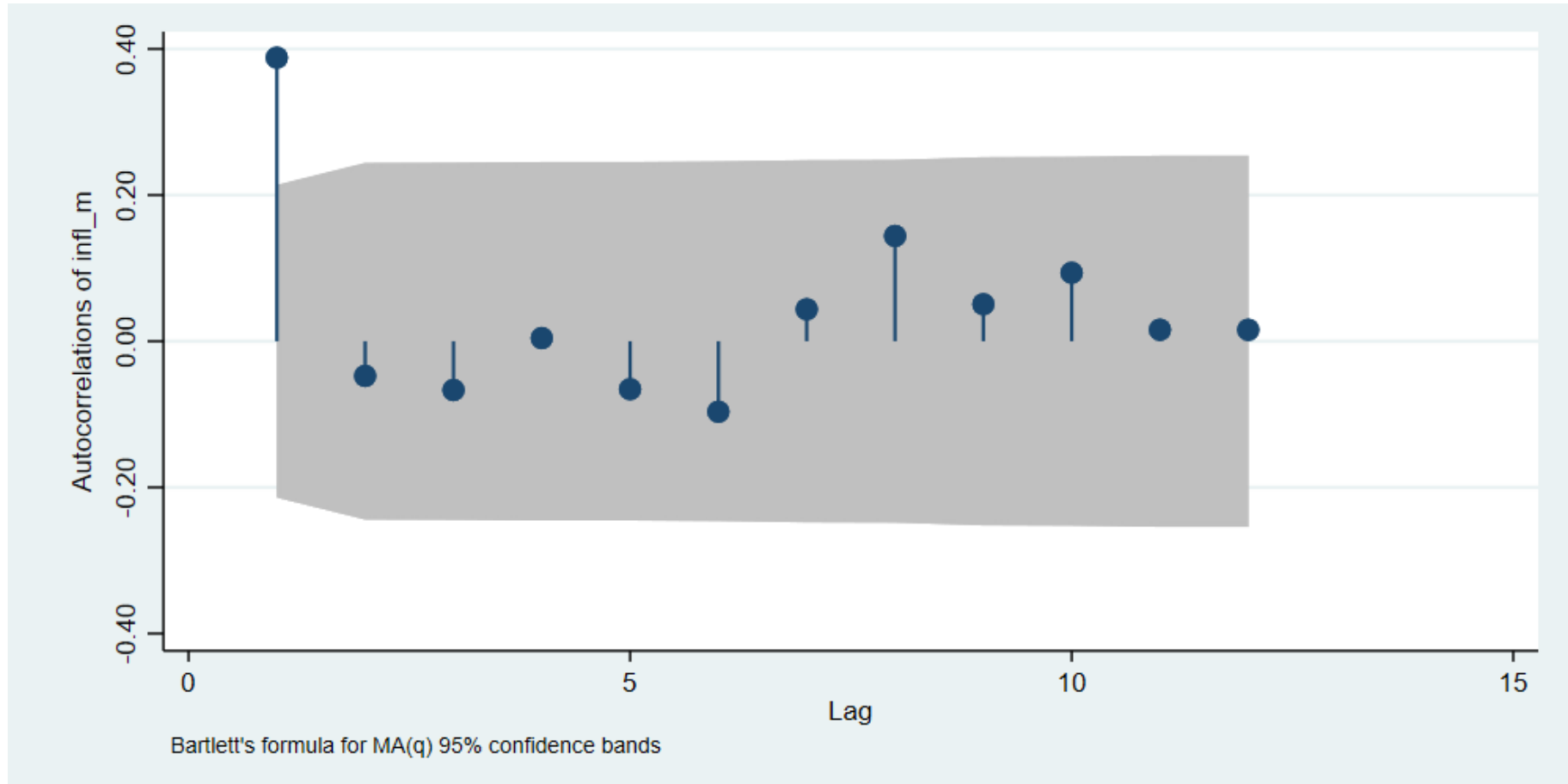
Nhiều Trắng và Chuỗi Dừng

- Nhiều trắng có phải là chuỗi dừng hay không?
- Tất cả chuỗi dừng có phải là nhiều trắng hay không?

Minh Họa: Tỷ Lệ Lạm Phát Giữa Các Tháng Với Nhau của Việt Nam (2013-2019)



Minh Họa: Biểu Đồ Correlogram



Chuỗi Không Dừng (Non-stationary)

- Một chuỗi không dừng (Non-stationary) là chuỗi không thỏa mãn các yêu cầu của một chuỗi dừng.
- Một trong các điều kiện sau thỏa mãn:

$$E(y_t) \neq \mu, \quad t = 1, 2, 3 \dots, T$$

$$var(y_t) \neq \sigma^2, \quad t = 1, 2, 3 \dots, T$$

$covariance(y_t, y_{t-k})$ không chỉ phụ thuộc vào k

Mô Hình Bước Đi Ngẫu Nhiên (Random Walk Model)

- Một trong chuỗi cơ bản nhất của chuỗi không dừng là mô hình bước đi ngẫu nhiên (random walk model).
- Mô hình này có thể chia làm hai dạng chủ yếu:
 - Bước ngẫu nhiên không trôi dạt (random walk without drift)
 - Bước ngẫu nhiên trôi dạt (random walk with drift)

Bước Ngẫu Nhiên Không Trôi Dạt (Random Walk Without Drift)

- Giả sử chuỗi thời gian y_t được hình thành như sau

$$y_t = y_{t-1} + u_t \quad (t = 1, 2, 3, \dots),$$

với u_t là nhiễu trắng với trung bình 0 và phương sai σ^2

- Bắt đầu từ giá trị đầu tiên, ta có:

$$y_1 = y_0 + u_0,$$

$$y_2 = y_1 + u_1 = y_0 + u_0 + u_1,$$

$$y_3 = y_2 + u_2 = y_0 + u_0 + u_1 + u_2$$

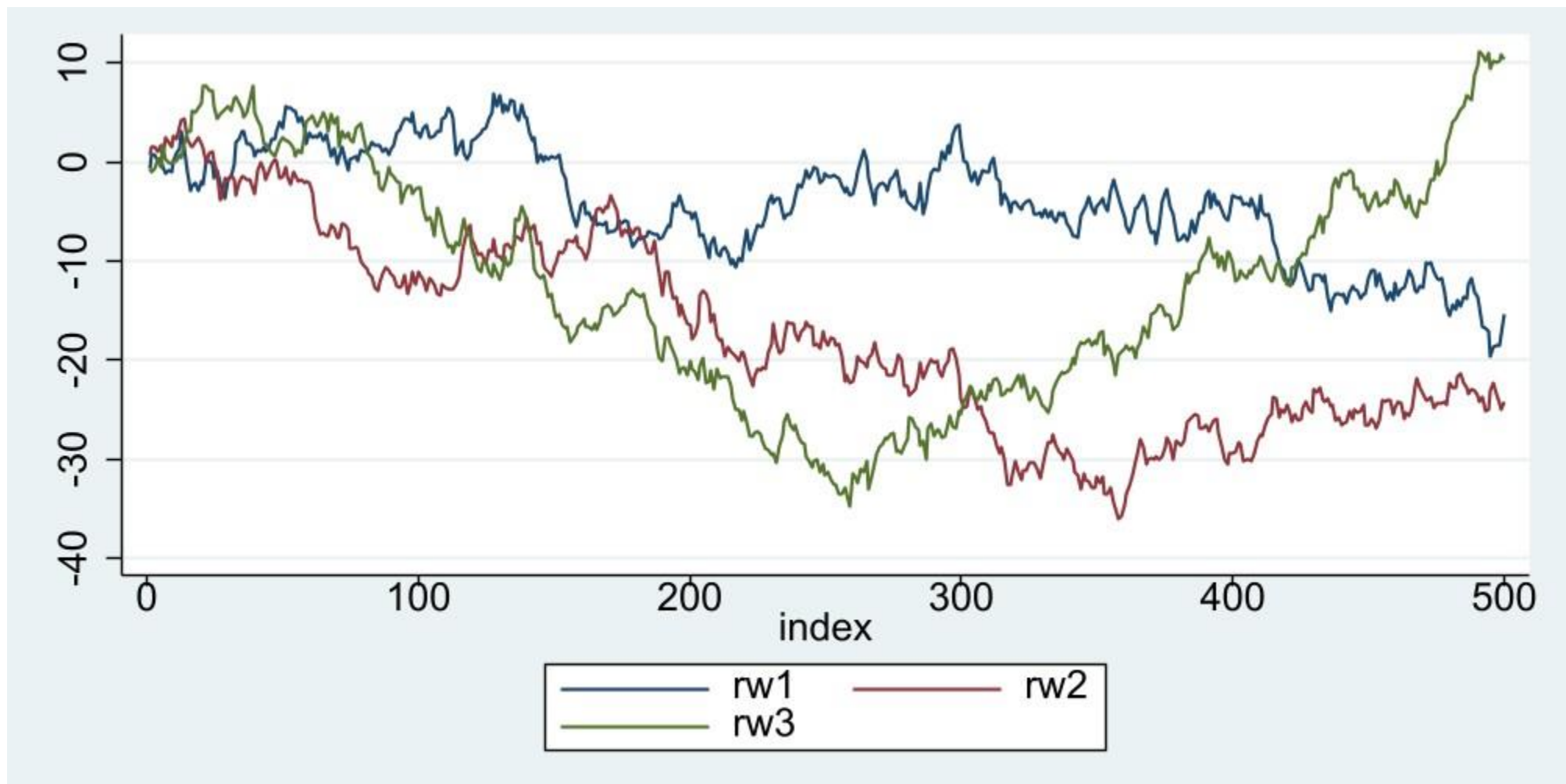
- Dần dần, ta có:

$$y_t = y_0 + \sum u_t$$

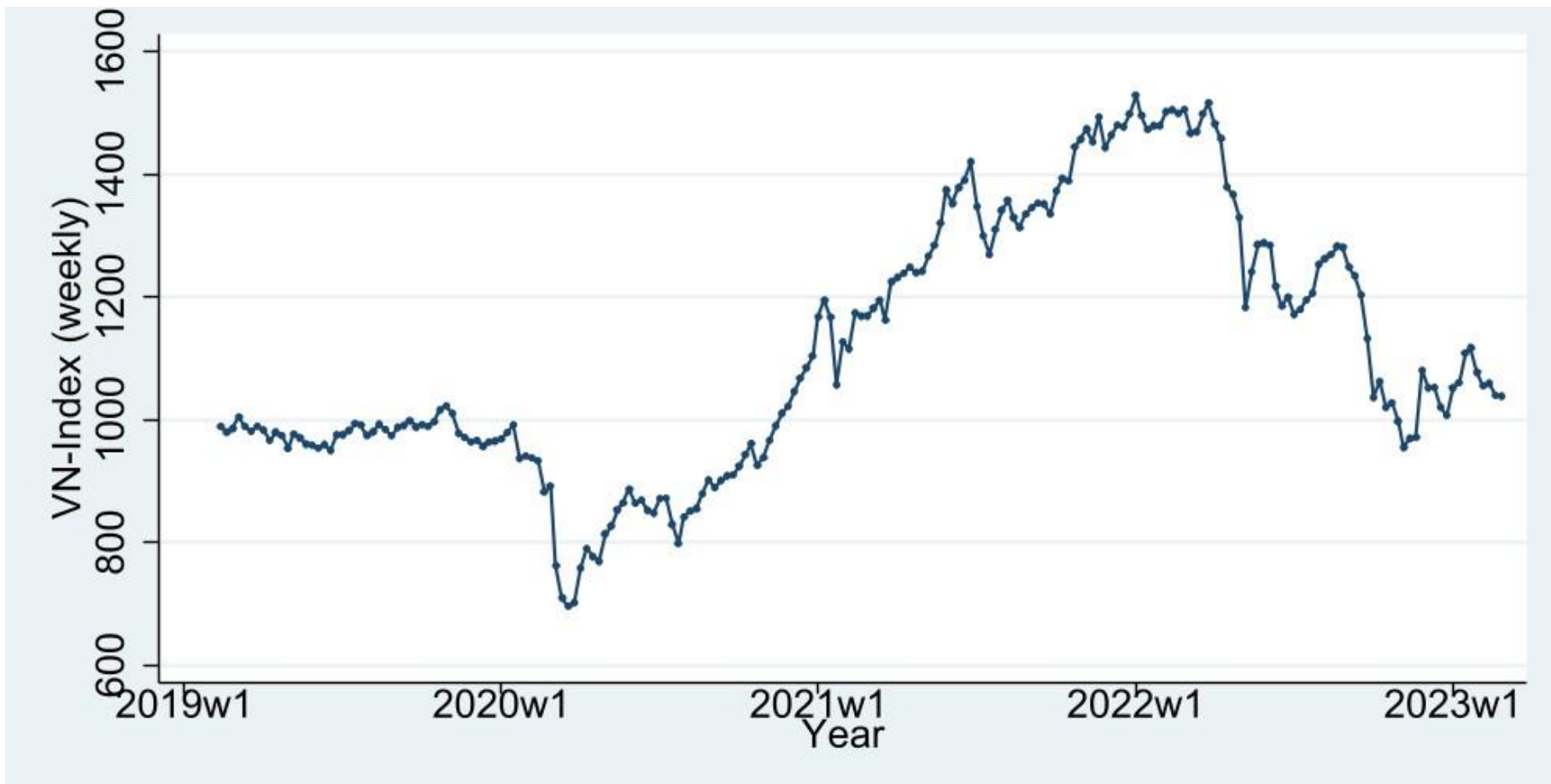
$$\rightarrow E(y_t) = E\left(y_0 + \sum u_t\right) = y_0$$

$$\rightarrow \text{var}(y_t) = t\sigma^2$$

Minh Họa: Bước Ngẫu Nhiên Không Trôi Dạt (1)



Minh Họa: Bước Ngẫu Nhiên Không Trôi Dạt (2)



Bước Đi Ngẫu Nhiên Trôi Dạt (Random Walk With Drift)

- Chúng ta có một chuỗi khác:

$$y_t = \delta + \beta_1 y_{t-1} + u_t \quad (t = 1, 2, 3, \dots)$$

- Chúng ta có thể chứng minh:

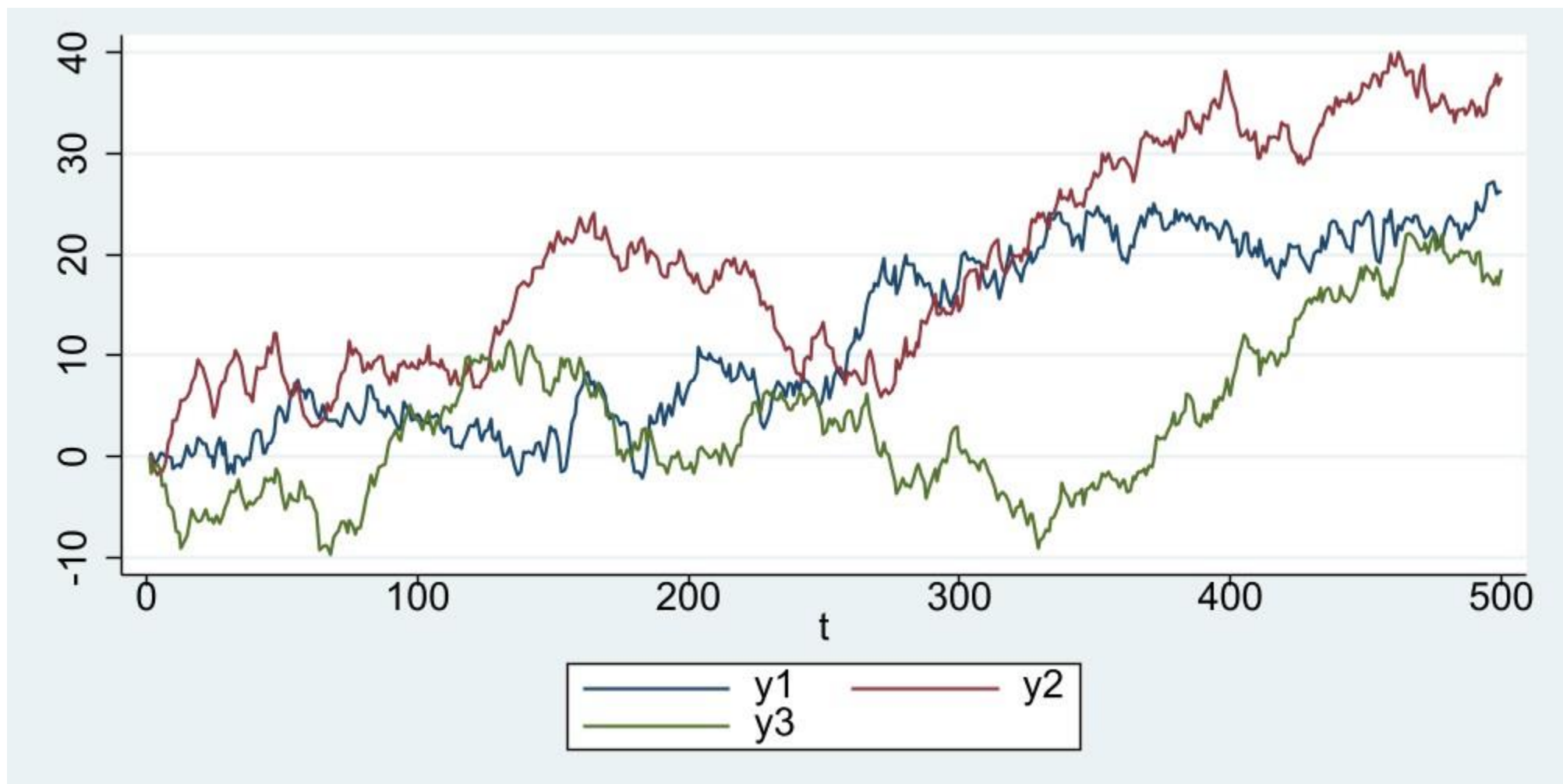
$$E(y_t) = y_0 + t\delta$$

$$\text{var}(y_t) = t\sigma^2$$

δ được biết như tham số dịch chuyển (drift parameter).

- Xu hướng tạo ra từ δ gọi là xu hướng ngẫu nhiên (stochastic trend).

Minh Hoạ: Bước Ngẫu Nhiên Trôi Dạt



Xu Hướng Xác Định (Deterministic Trend)

- Quá trình này được diễn tả như sau:

$$y_t = \delta + \beta_1 t + u_t \quad (t = 1, 2, 3, \dots)$$

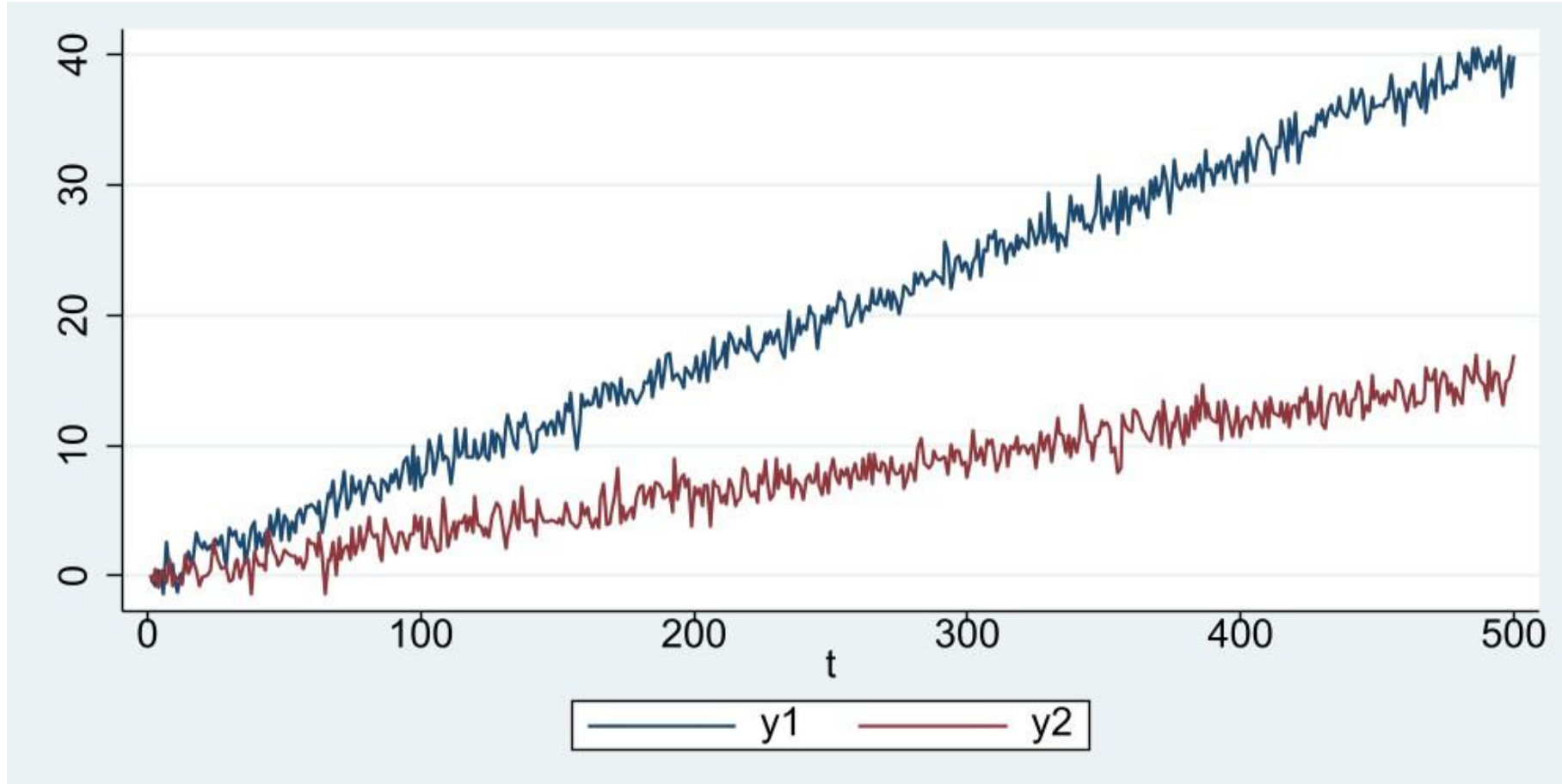
- Chúng ta có thể chứng minh:

$$E(y_t) = y_0 + \beta_1 t$$

$$\text{var}(y_t) = \sigma^2$$

- Quá trình còn được gọi là dừng có xu hướng (trend stationary). Loại bỏ xu hướng của quá trình (detrending), chúng ta sẽ được một chuỗi dừng.

Minh Họa: Xu Hướng Xác Định



Bước Đi Ngẫu Nhiên Theo Xu Hướng và Có Trượt (Random Walk With Drift and Deterministic Trend)

- Là quá trình kết hợp giữa hai quá trình không dừng chúng ta đã thảo luận ở trên:

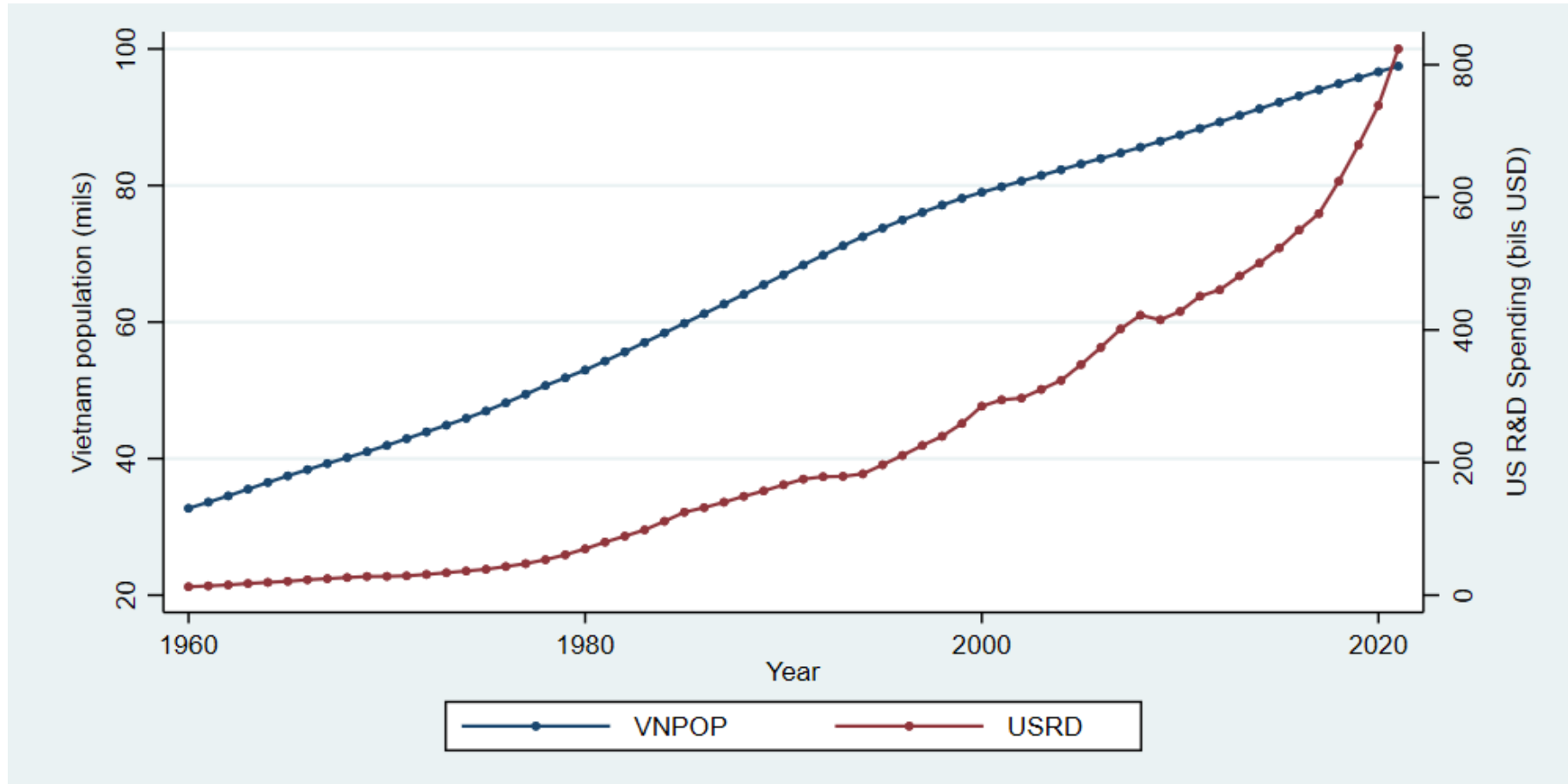
$$y_t = \delta + \beta_1 t + y_t + u_t$$

Hồi Quy Giả (Spurious Regression)

Mối Liên Hệ Thực Tế Giữa Hai Hiện Tượng Sau Đây?

- Dân số Việt Nam (y_t),
- Chỉ tiêu cho nghiên cứu và phát triển của Mỹ (x_t),
- Dữ liệu năm từ 1990 đến 2020.

Đồ Thị Thời Gian Của Hai Biến



Kết Quả Hồi Quy

reg VNPOP USRD

Source	SS	df	MS	Number of obs	=	62
Model	21454.2748	1	21454.2748	F(1, 60)	=	359.24
Residual	3583.32543	60	59.7220906	Prob > F	=	0.0000
Total	25037.6002	61	410.452462	R-squared	=	0.8569
				Adj R-squared	=	0.8545
				Root MSE	=	7.728

VNPOP	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
USRD	.0895685	.0047257	18.95	0.000	.0801157	.0990213
_cons	45.69107	1.454297	31.42	0.000	42.78205	48.6001

corr VNPOP USRD

obs=62)

	VNPOP	USRD
VNPOP	1.0000	
USRD	0.9257	1.0000

Các Ví Dụ Minh Họa Khác (1)

- Tỷ lệ tử vong vị thành niên của Ai Cập (y_t), 1971-1990, dữ liệu năm, theo tổng thu nhập của nông dân Mỹ (I_t) và cung tiền của Honduras (m_t).
- Chỉ số xuất khẩu của Mỹ (y_t), 1960-1990, dữ liệu năm, theo tuổi thọ đàn ông người Úc (x_t).

Các Ví Dụ Minh Họa Khác (2)

- Tỷ lệ tử vong vì thành niên của Ai Cập (y_t), 1971-1990, dữ liệu năm, theo tổng thu nhập của nông dân Mỹ (I_t) và cung tiền của Honduras (m_t).

$$y_t = 179.9 - 0.2952 I_t - 0.0439 m_t, \quad R^2 = 0.918$$

$$(16.63) \quad (-2.32) \quad (-4.26)$$

$$Corr = 0.8858, -0.9113, -0.9445$$

- Chỉ số xuất khẩu của Mỹ (y_t), 1960-1990, dữ liệu năm, và tuổi thọ đàn ông người Úc (x_t).

$$y_t = -2943 + 45.7974 x_t, \quad R^2 = 0.916$$

$$(-16.70) \quad (17.76)$$

$$Corr = 0.9570$$

Hồi Quy Giả (Spurious Regression)

- Hồi quy giả là kết quả hồi quy có sự tương đồng cao giữa hai biến đến từ hai hiện tượng thực tế hầu như không liên quan với nhau.
- Hồi quy giả thường xuất hiện giữa các biến không dừng có cùng xu hướng tăng (hoặc giảm) theo thời gian.
- Để tránh hồi quy giả trong phương pháp OLS, **chúng ta phải sử dụng các biến dừng.**
- Chúng ta sẽ nghiên cứu sâu vấn đề này ở phần tiếp theo.

Kiểm Định Chuỗi Dừng

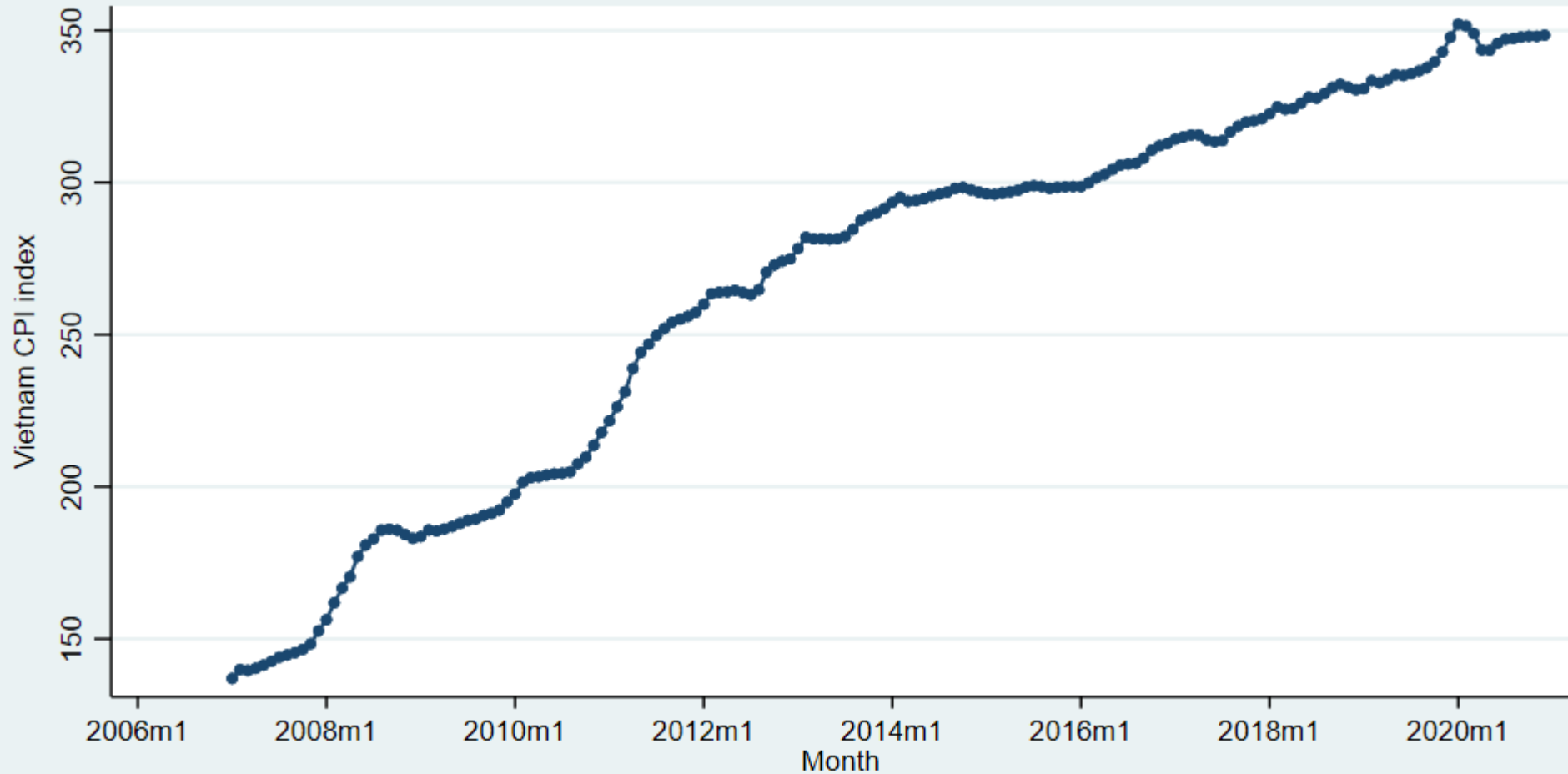
Các Cách Tiếp Cận Cơ bản

1. Sử dụng đồ thị.

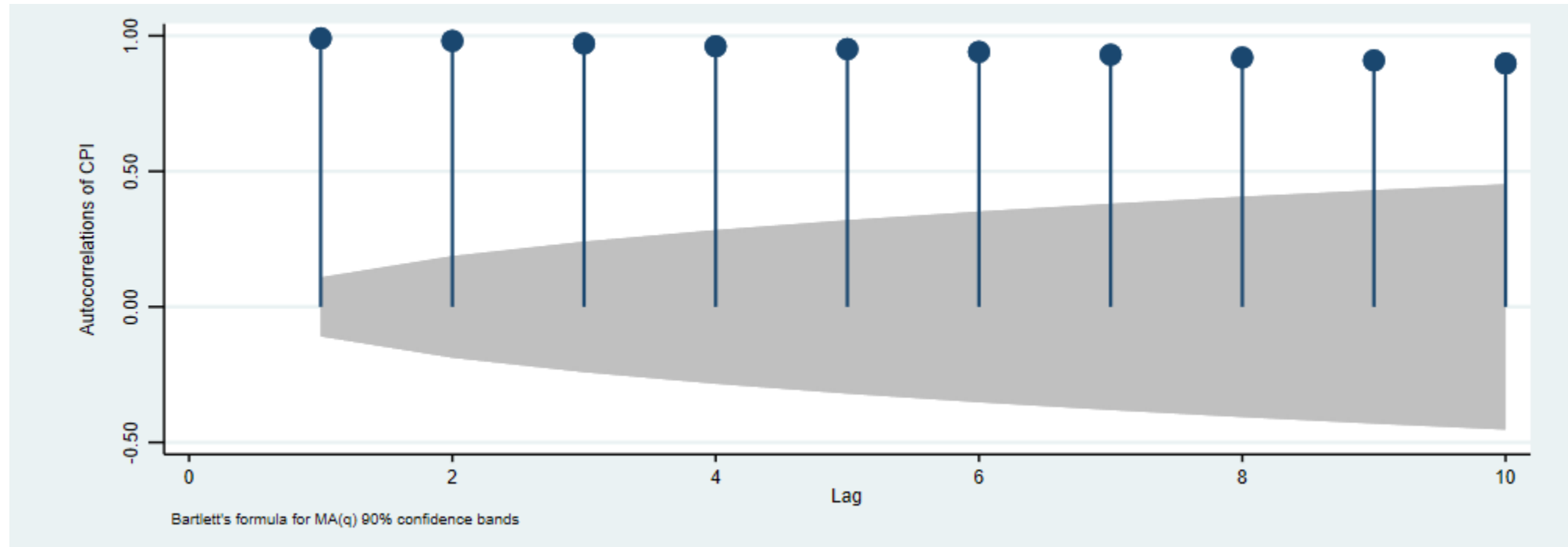
1. Sử dụng Hàm tự tương quan (ACF) và Correlogram.

2. Sử dụng Kiểm định nghiệm đơn vị (Unit root test).

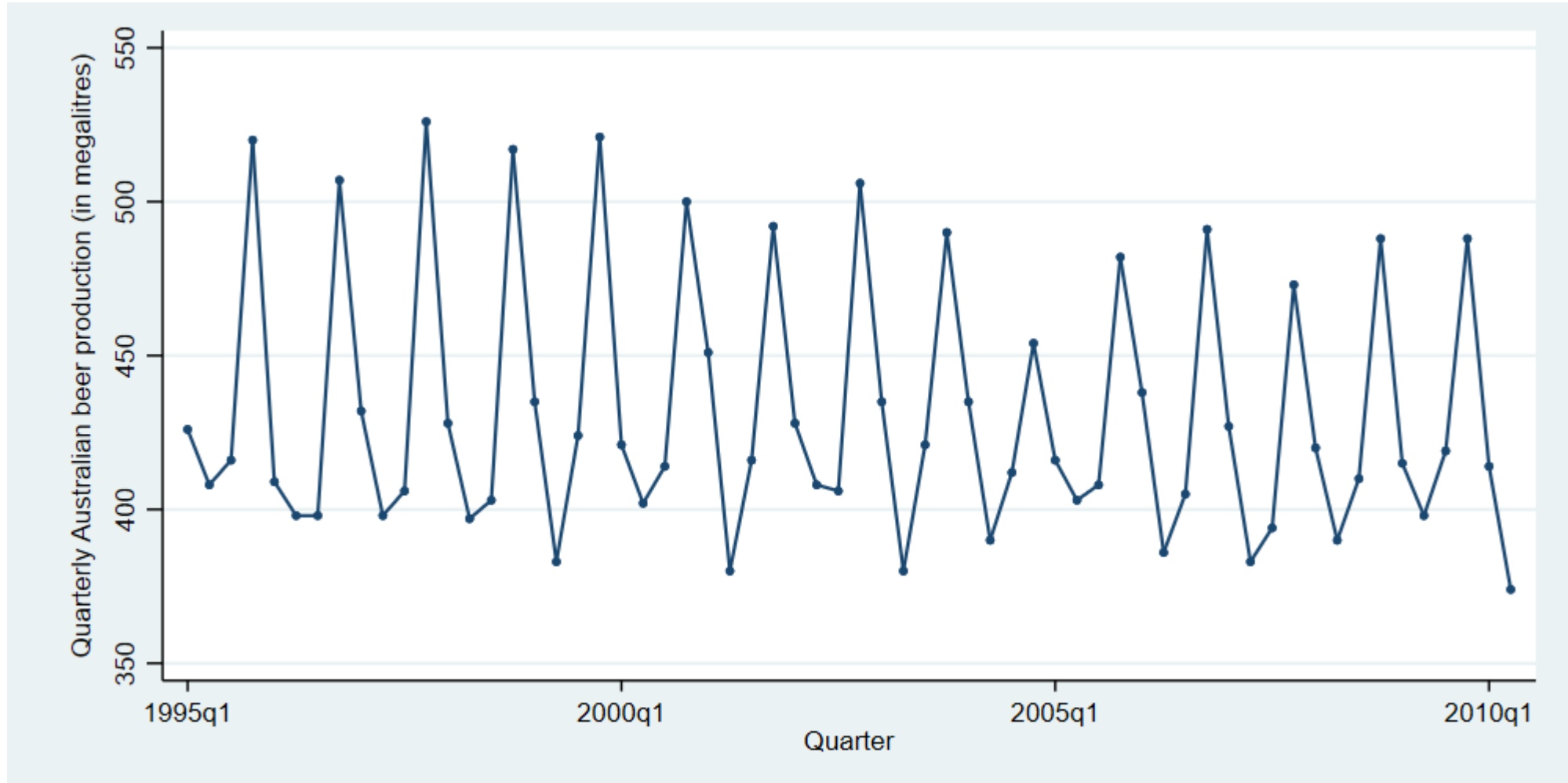
Dừng Hay Không Dừng? - Ví dụ 1 (1)



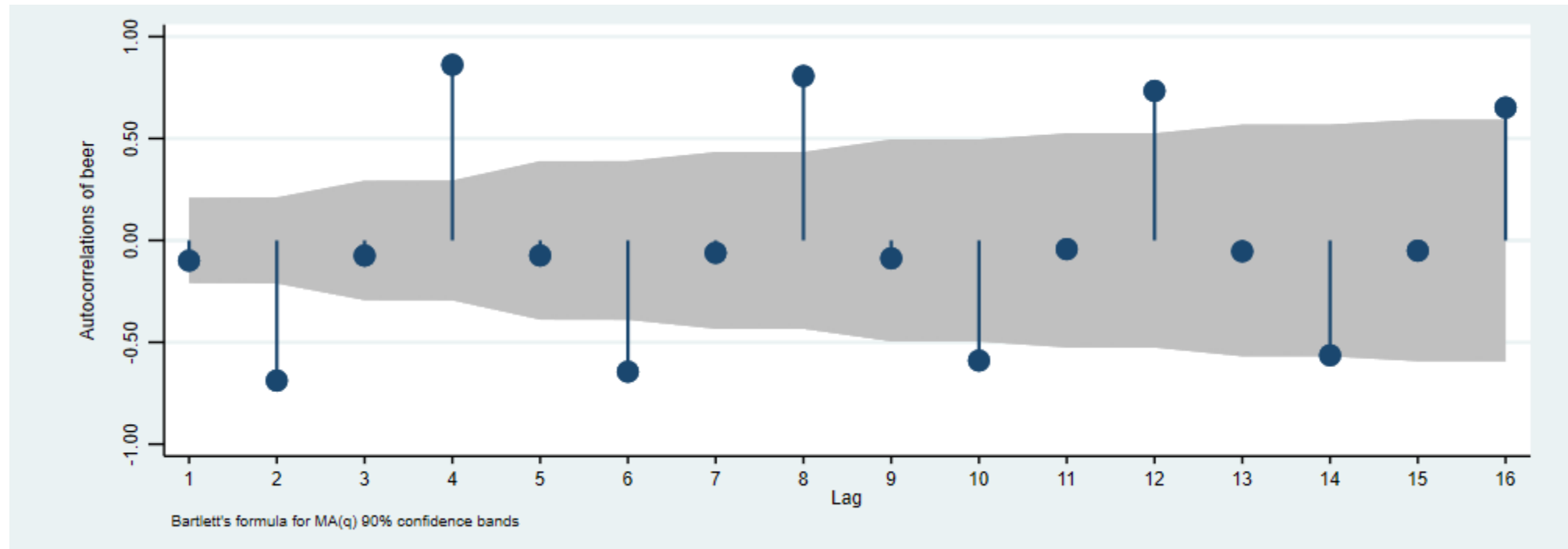
Dừng Hay Không Dừng? - Ví dụ 1 (2)



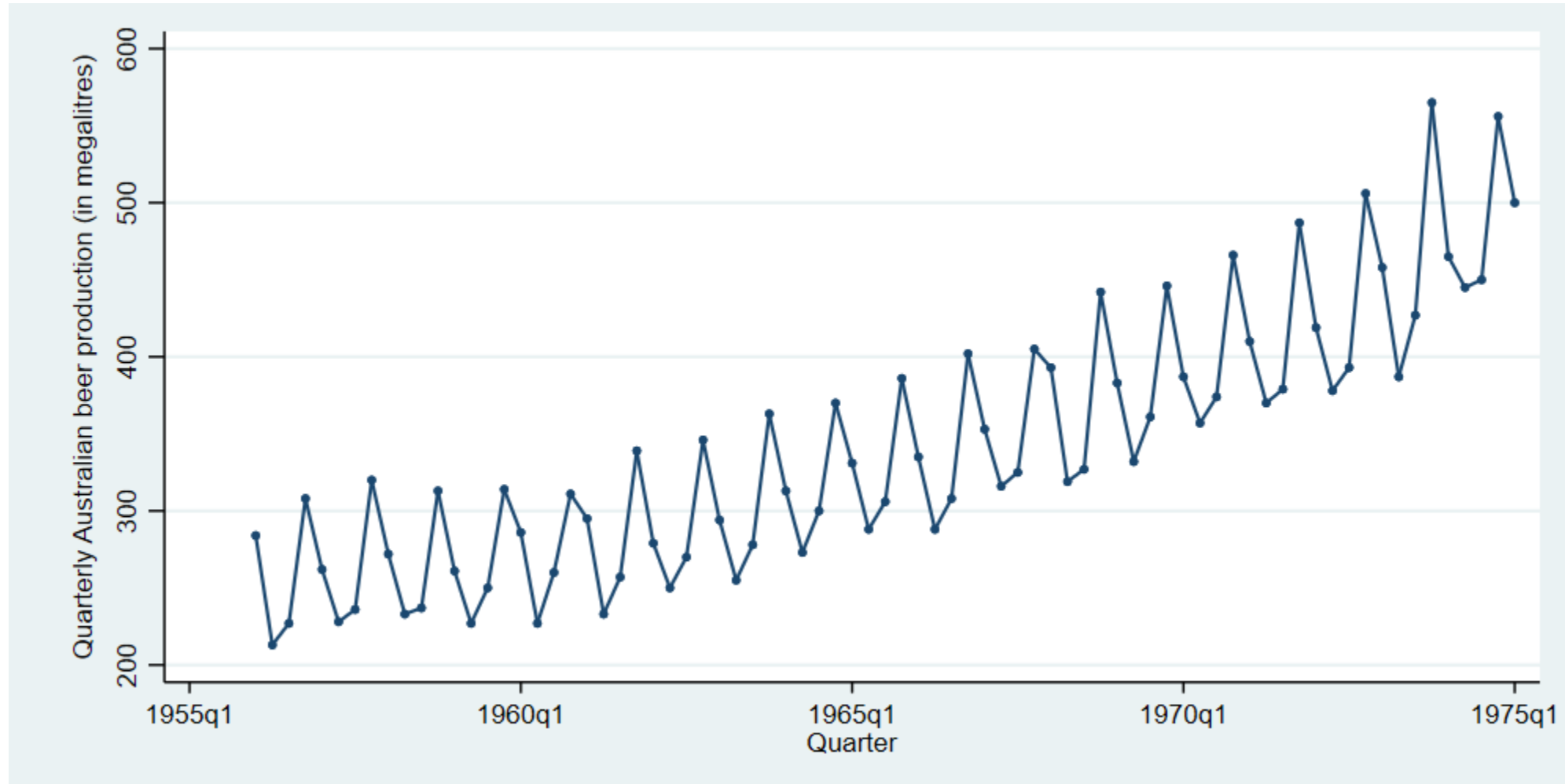
Dừng Hay Không Dừng? - Ví dụ 2 (1)



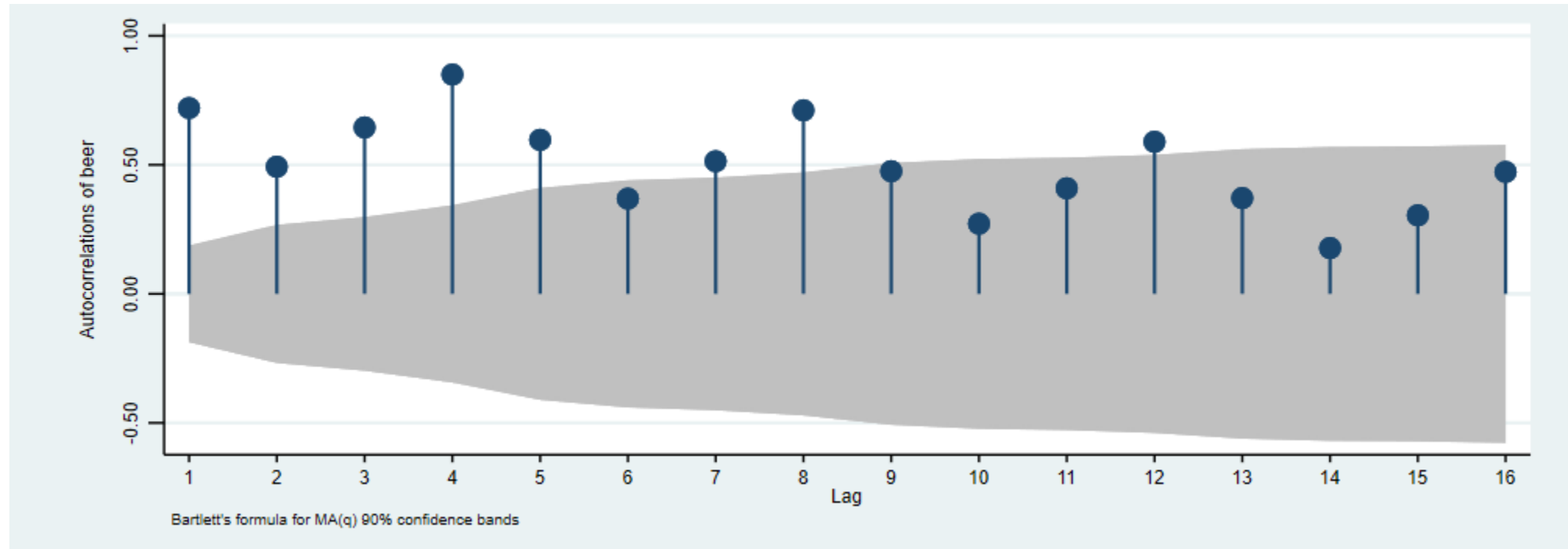
Dừng Hay Không Dừng? - Ví dụ 2 (2)



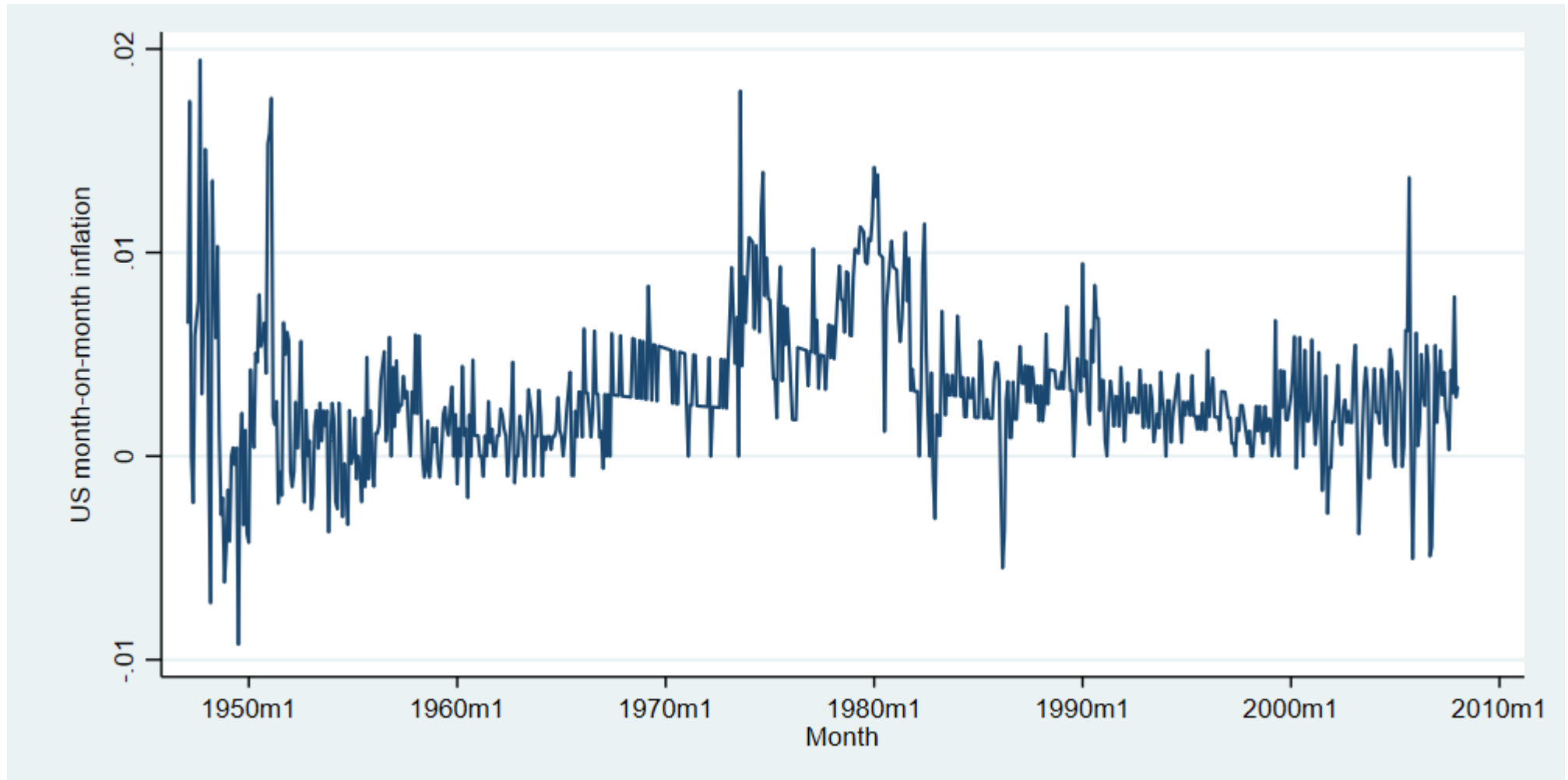
Dừng Hay không dừng? - Ví dụ 3 (1)



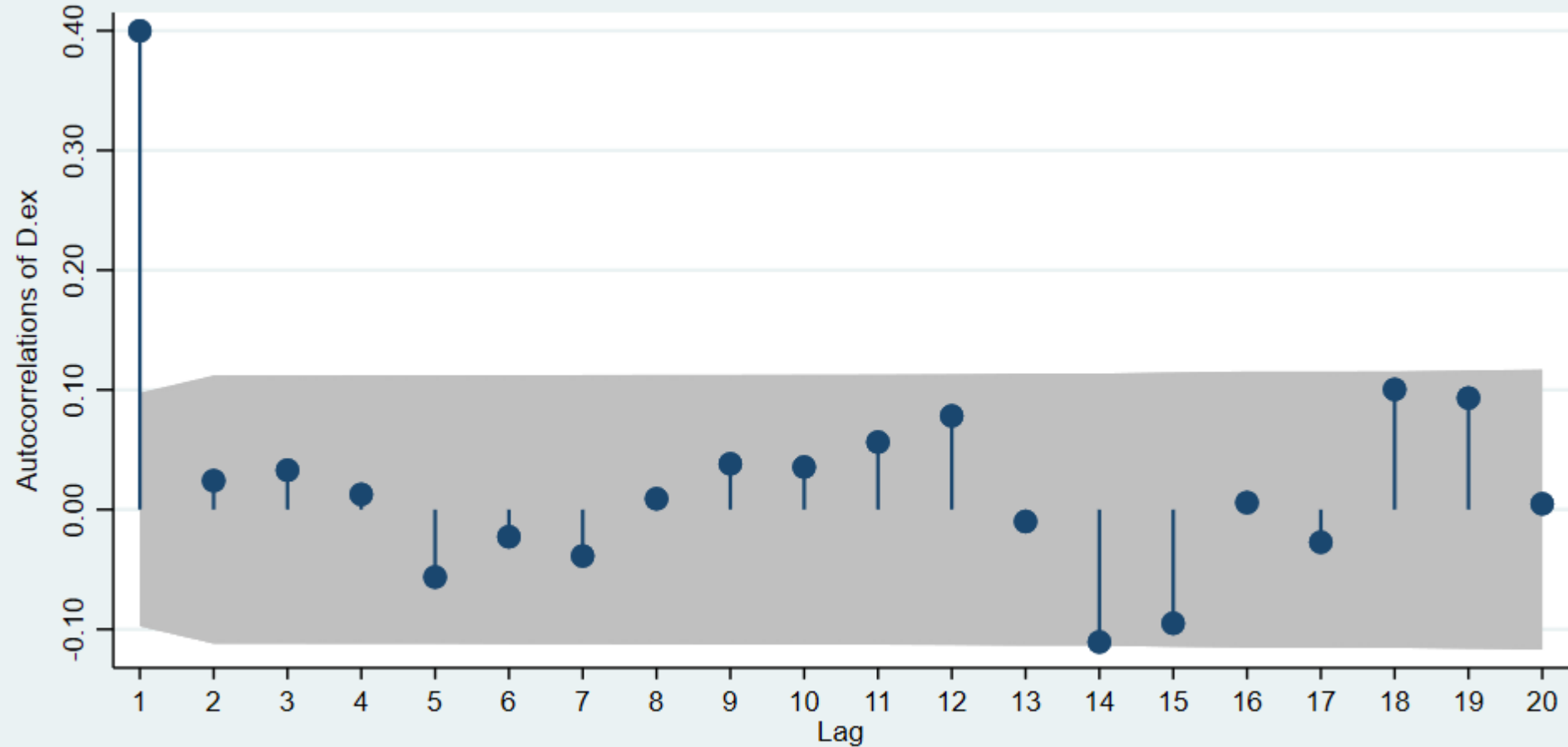
Dừng Hay Không Dừng? - Ví dụ 3 (2)



Dừng Hay Không Dừng? - Ví dụ 4 (1)

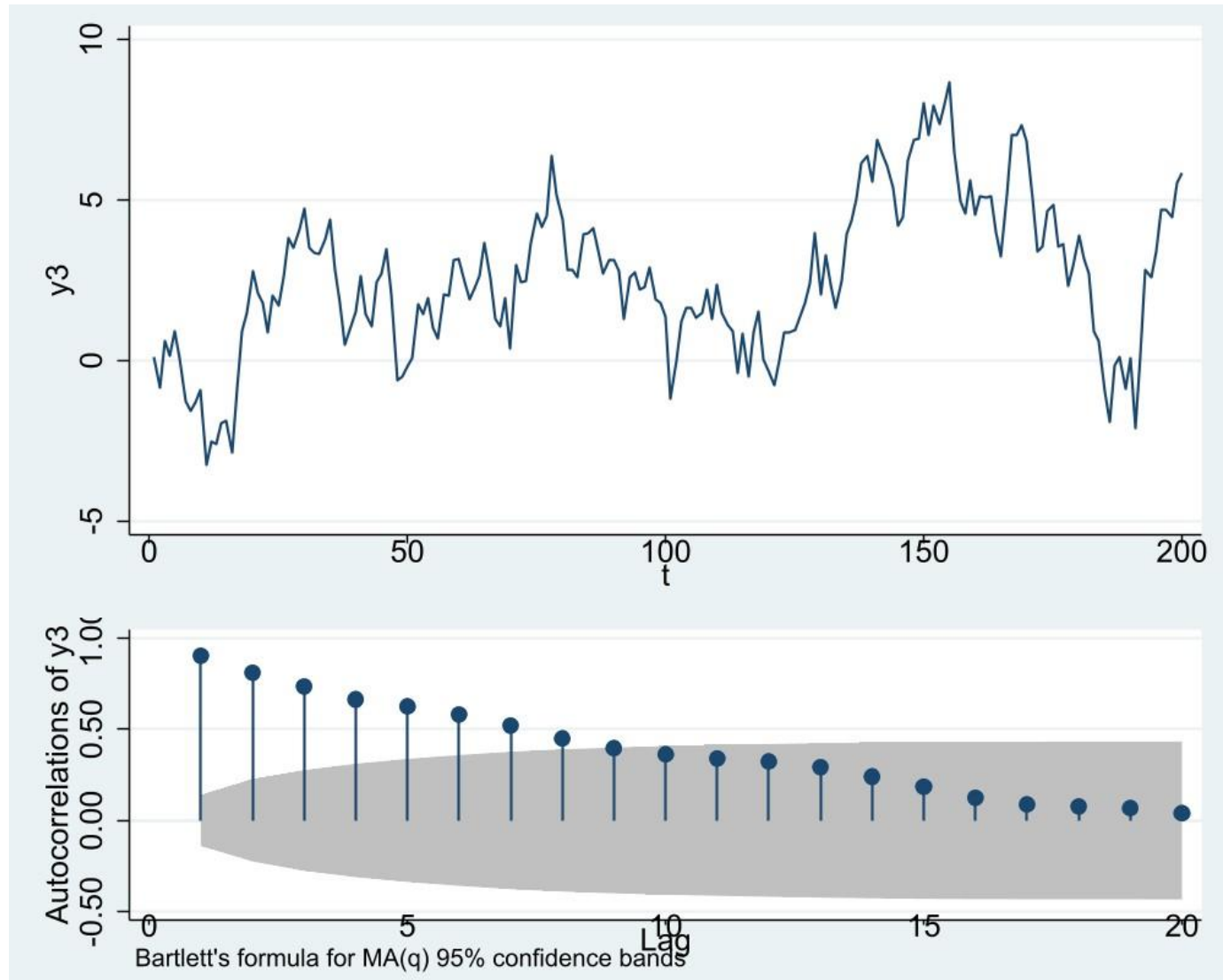


Dừng Hay Không Dừng? - Ví dụ 4 (2)



Bartlett's formula for MA(q) 90% confidence bands

Trường Hợp: $y_t = 0.3 + 0.90y_{t-1} + u_t$



Kiểm Định Dickey-Fuller (Dickey-Fuller (DF) Test)

- y_t là bước đi ngẫu (a random walk):

$$\Delta y_t = \delta y_{t-1} + u_t$$

- y_t là bước đi ngẫu nhiên không trôi dạt (a random walk with drift):

$$\Delta y_t = \beta_1 + \delta y_{t-1} + u_t$$

- y_t là bước đi ngẫu nhiên trôi theo một chiều hướng định trước (a random walk with drift around a deterministic trend):

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + u_t$$

- Giá trị của δ khi chuỗi dừng và không dừng?

Giả Thuyết Vô Hiệu và Thay Thế (Null and Alternative Hypotheses)

Trong từng trường hợp, chúng ta có:

- Giả thuyết vô hiệu:

$$H_0: \delta = 0$$

được hiểu như nghiệm đơn vị (unit root) tồn tại và *chuỗi không dừng*.

- Giả thuyết thay thế:

$$H_1: \delta < 0,$$

được hiểu như *chuỗi dừng*.

- Sử dụng t -stat để kiểm định giá trị δ . Tuy nhiên, giá trị tới hạn không dùng t -test như bình thường.

The Augmented Dickey–Fuller (ADF) Test

- The Augmented Dickey–Fuller (ADF) Test:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-i} + u_t$$

- Giả thuyết vô hiệu:

$$H_0: \delta = 0$$

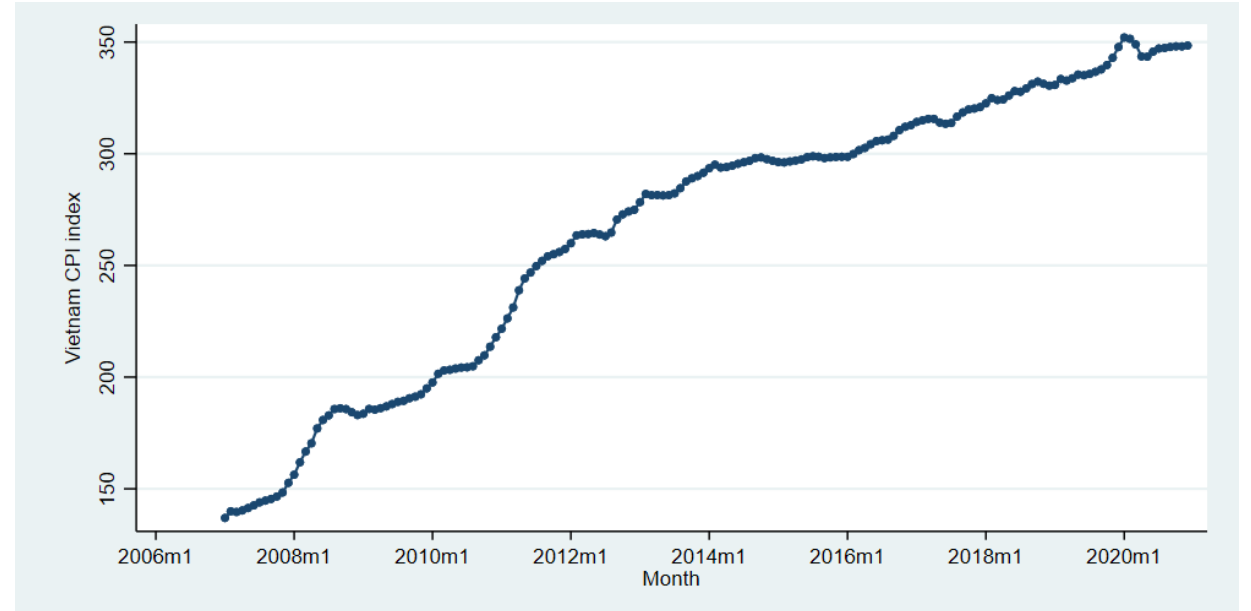
được hiểu như nghiệm đơn vị (unit root) tồn tại và *chuỗi không dừng*

Các Kiểm Định Nghiệm Đơn Vị Khác

- The Phillips–Perron (PP) Unit Root Test.
- Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test.
 - Lưu ý: H_0 của kiểm định KPSS ngược với các kiểm định DF, ADF, PP – giả thuyết vô hiệu là chuỗi dừng.

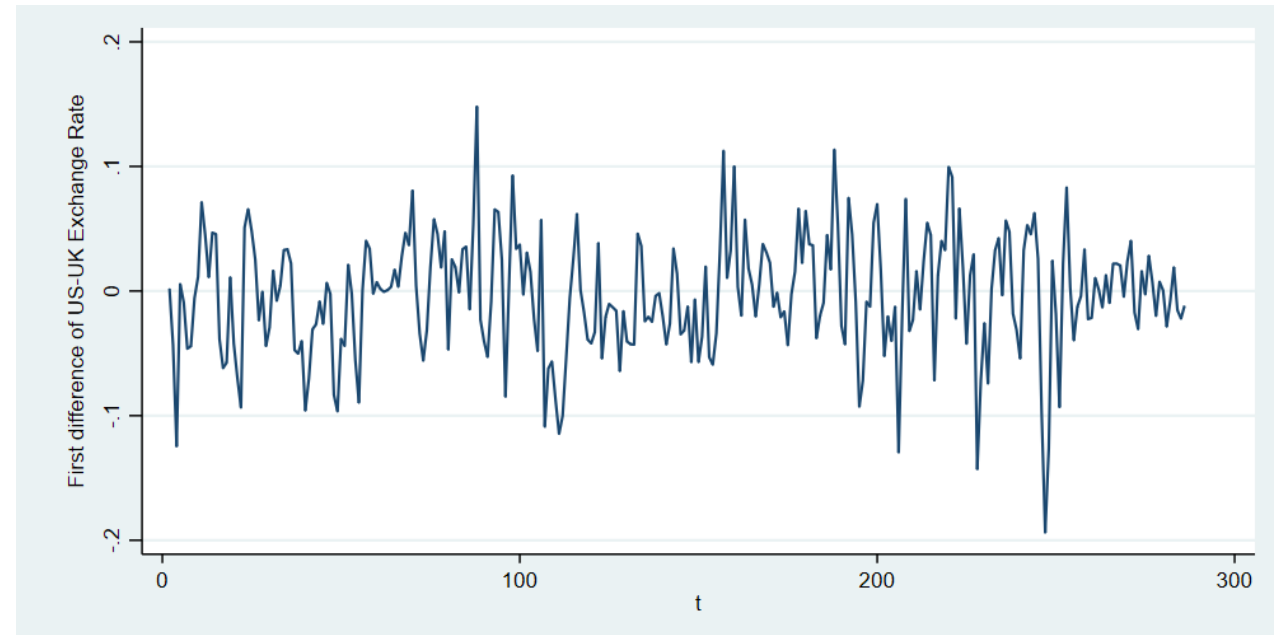
Minh Họa (1)

		Dickey–Fuller critical value		
	Test statistic	1%	5%	10%
With drift	-0.438	-3.468	-2.882	-2.572
With trend	-0.480	-3.998	-3.433	-3.133
With drift and 12 lags	-1.520	-4.002	-3.435	-3.135



Minh Họa (2)

		Dickey–Fuller critical value		
	Test statistic	1%	5%	10%
With drift	-11.155	-3.989	-3.429	-3.13
With trend	-11.153	-2.340	-1.650	-1.285
With drift and 12 lags	-4.228	-3.989	-3.429	-3.130



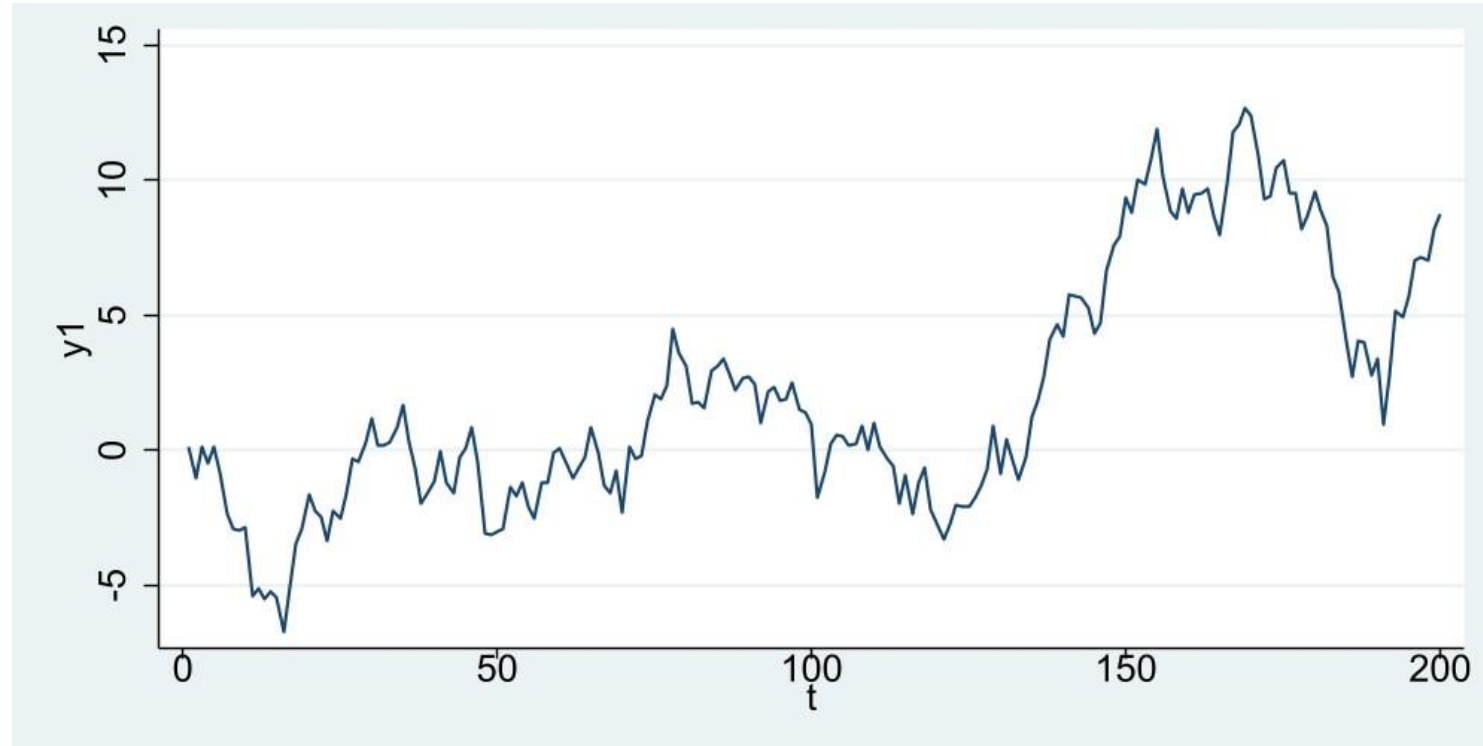
Lưu Ý Về Kiểm Định ADF

- Khi giá trị của δ rất gần 0 (giá trị nghiệm rất gần 1). Kiểm định ADF không còn hiệu lực mạnh.
 - Chúng ta có thể sử dụng kiểm định KPSS đề xuất ở trên để bổ sung tham chiếu.
- Cần thận khi lựa chọn dạng quá trình khác nhau.
- Vấn đề cấu trúc thay đổi (structural break).
- Cách lựa chọn độ trễ cho kiểm định ADF. Chúng ta sẽ nghiên cứu vấn đề này ở bài sau.

Minh Hoạ: Giá Trị Nghiệm Đơn Vị Rất Gần 1

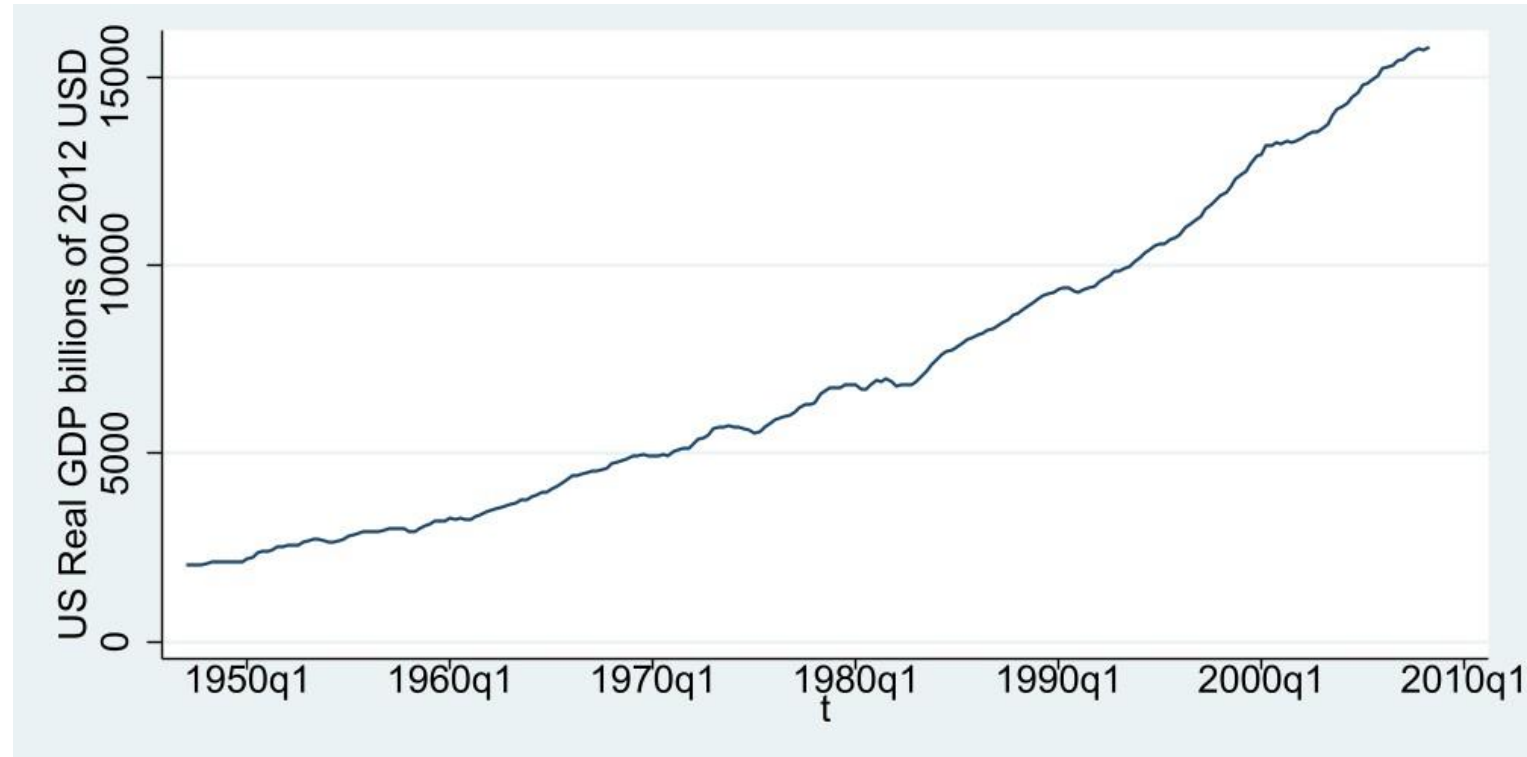
$$y_t = 0.1 + 0.98y_{t-1} + u_t$$

		ADF critical values		
	Test-stat	1%	5%	10%
With drift	-1.271	-2.345	-1.653	-1.286
With trend	-2.659	-4.007	-3.437	-3.137
With drift and 2 lags	-1.252	-2.346	-1.653	-1.286
KPSS test	1.167	2.621	1.617	1.187
Note: H_0 của kiểm định KPSS là chuỗi dừng.				

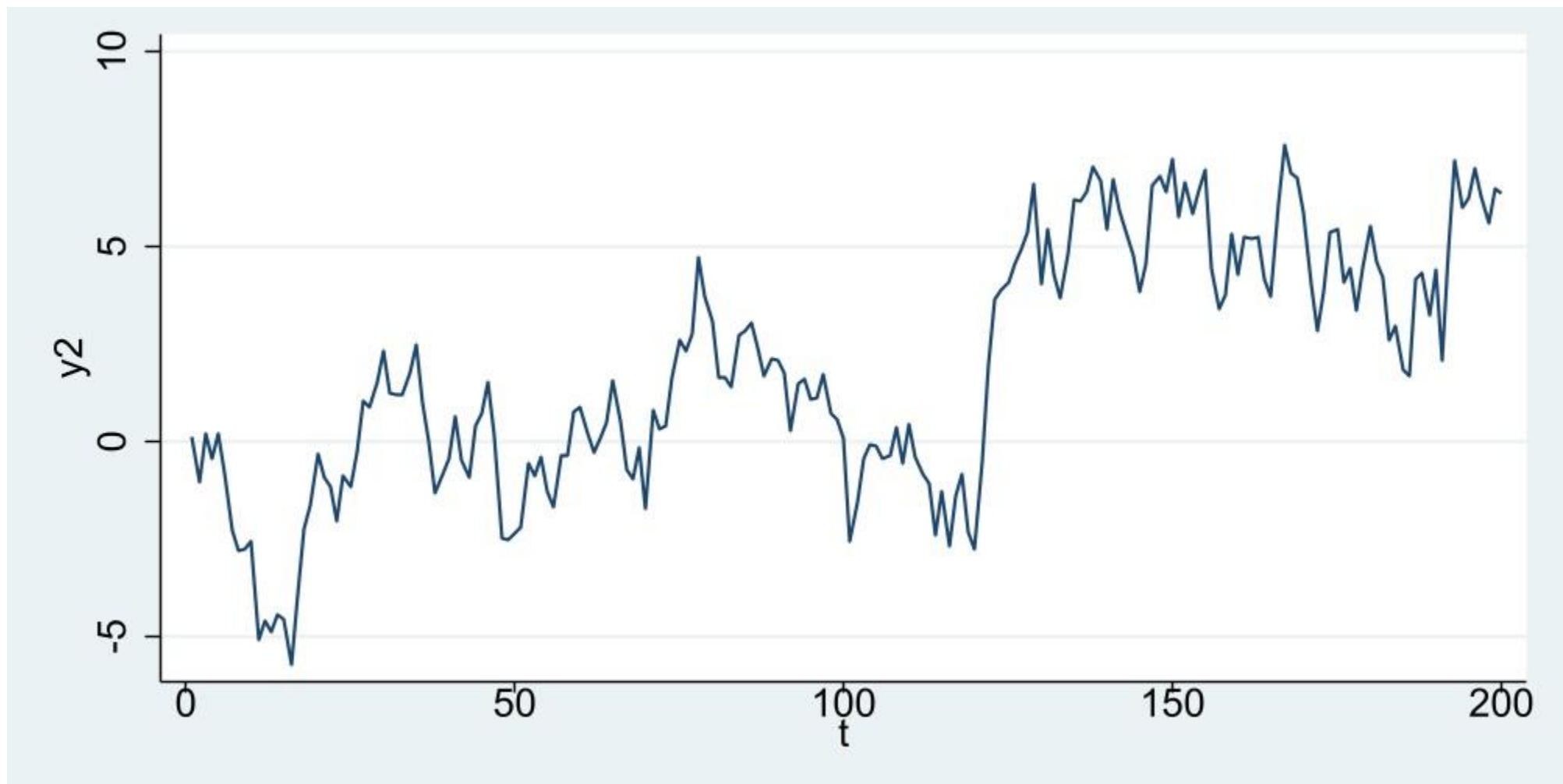


Minh Họa: GDP Thực Của Mỹ

		ADF critical values		
	Test-stat	1%	5%	10%
With drift	6.978	-2.342	-1.651	-1.285
With trend	0.240	-3.992	-3.431	-3.131
With drift and 2 lags	3.478	-2.342	-1.651	-1.285
KPSS test	7.147	2.648	1.622	1.188
Note: H_0 của kiểm định KPSS là chuỗi dừng.				



Minh Họa: Cấu Trúc Thay Đổi



**Chuyển Đổi Chuỗi Dừng
Thành Chuỗi Không Dừng**

Chuỗi Dừng Xu Thế (Trend-Stationary Processes)

- Chúng ta bắt đầu với một chuỗi không dừng như sau:

$$y_t = \beta_1 + \beta_2 t + u_t \quad (t = 1, 2, 3, \dots)$$

- Loại bỏ xu hướng thời gian:

$$\hat{u}_t = y_t - \hat{\beta}_1 + \hat{\beta}_2 t$$

- Chúng ta có chuỗi dừng $\hat{u}_t$.

Chuỗi Dừng Sai Phân (Difference-Stationary Process)

- Chúng ta bắt đầu với một chuỗi không dừng như sau:

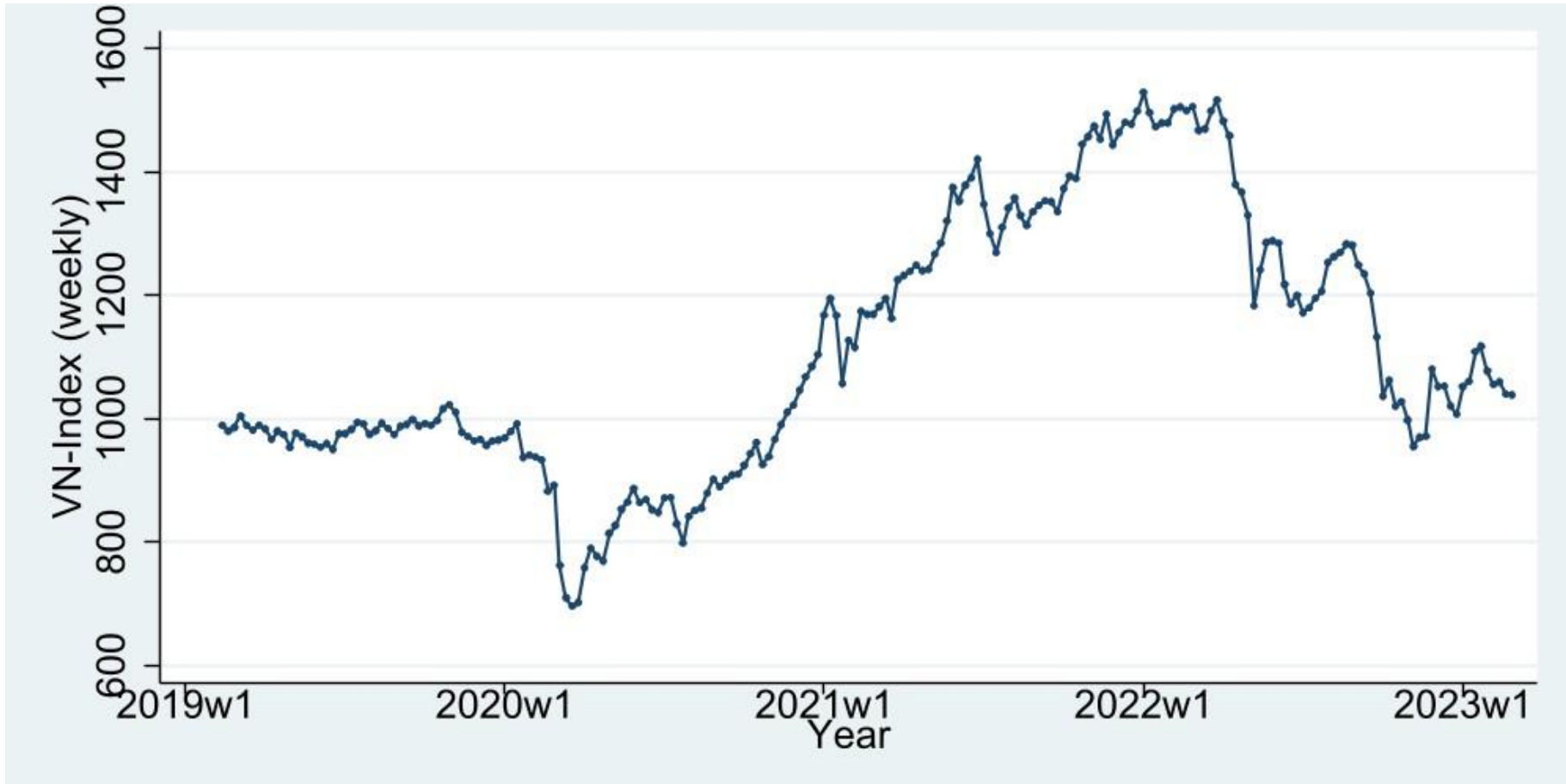
$$y_t = \delta + y_{t-1} + u_t \quad (t = 1, 2, 3, \dots)$$

- Điều gì sẽ xảy ra nếu chúng ta có:

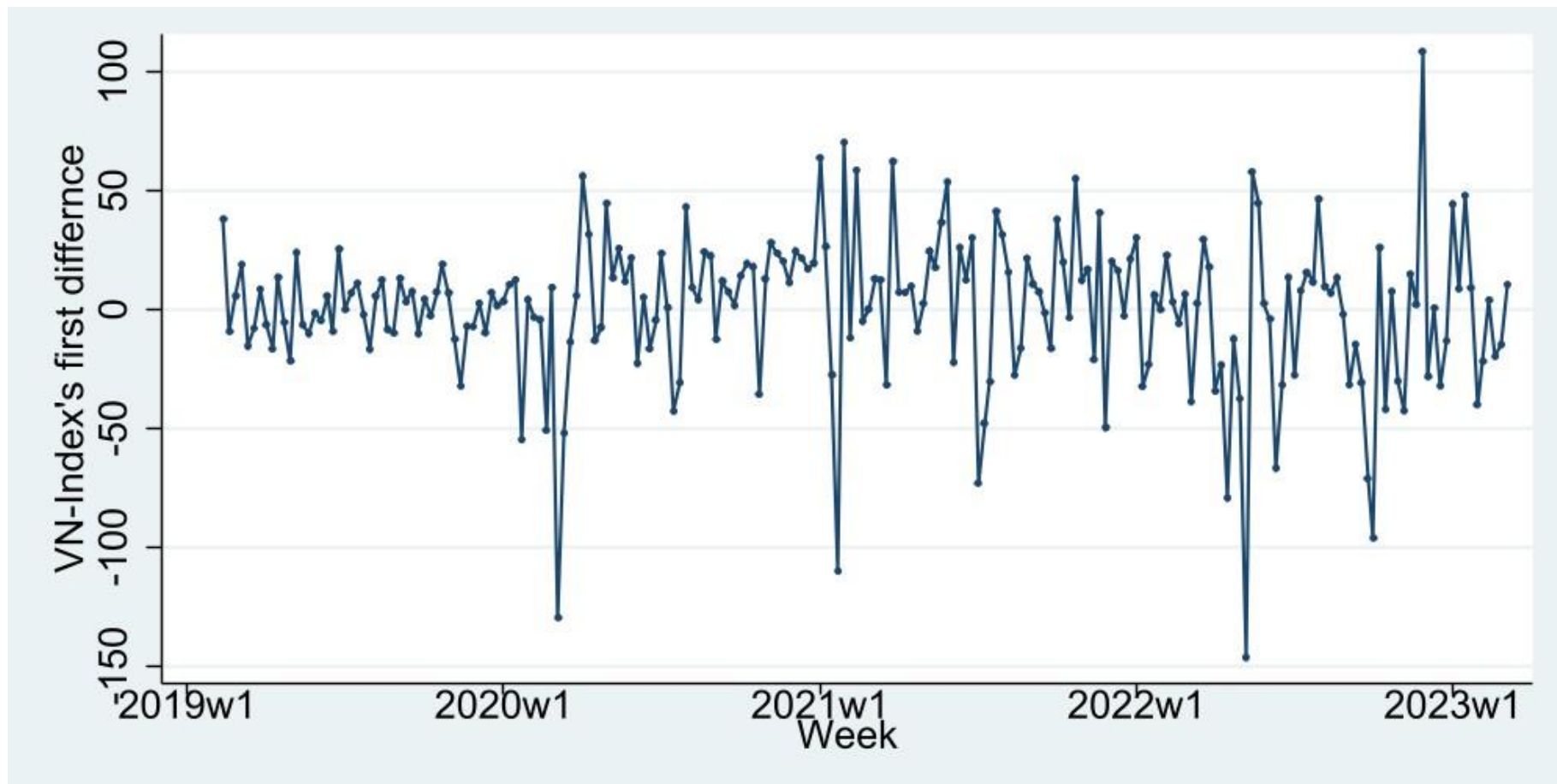
$$\begin{aligned}\Delta y_t &= y_t - y_{t-1} \\ &= (\delta + y_{t-1} + u_t) - (\delta + y_{t-2} + u_{t-1}) \\ &= u_t - u_{t-1}\end{aligned}$$

- Chuỗi Δy_t có phải là chuỗi dừng hay không?

Ví dụ: Chuỗi VN-Index (1)



Ví dụ: Chuỗi VN-Index (2) – Sai Phân Bậc 1



Chuỗi Dừng Sai Phân Hàm Log (Log-difference Form)

- Một phương pháp khác phổ biến là:

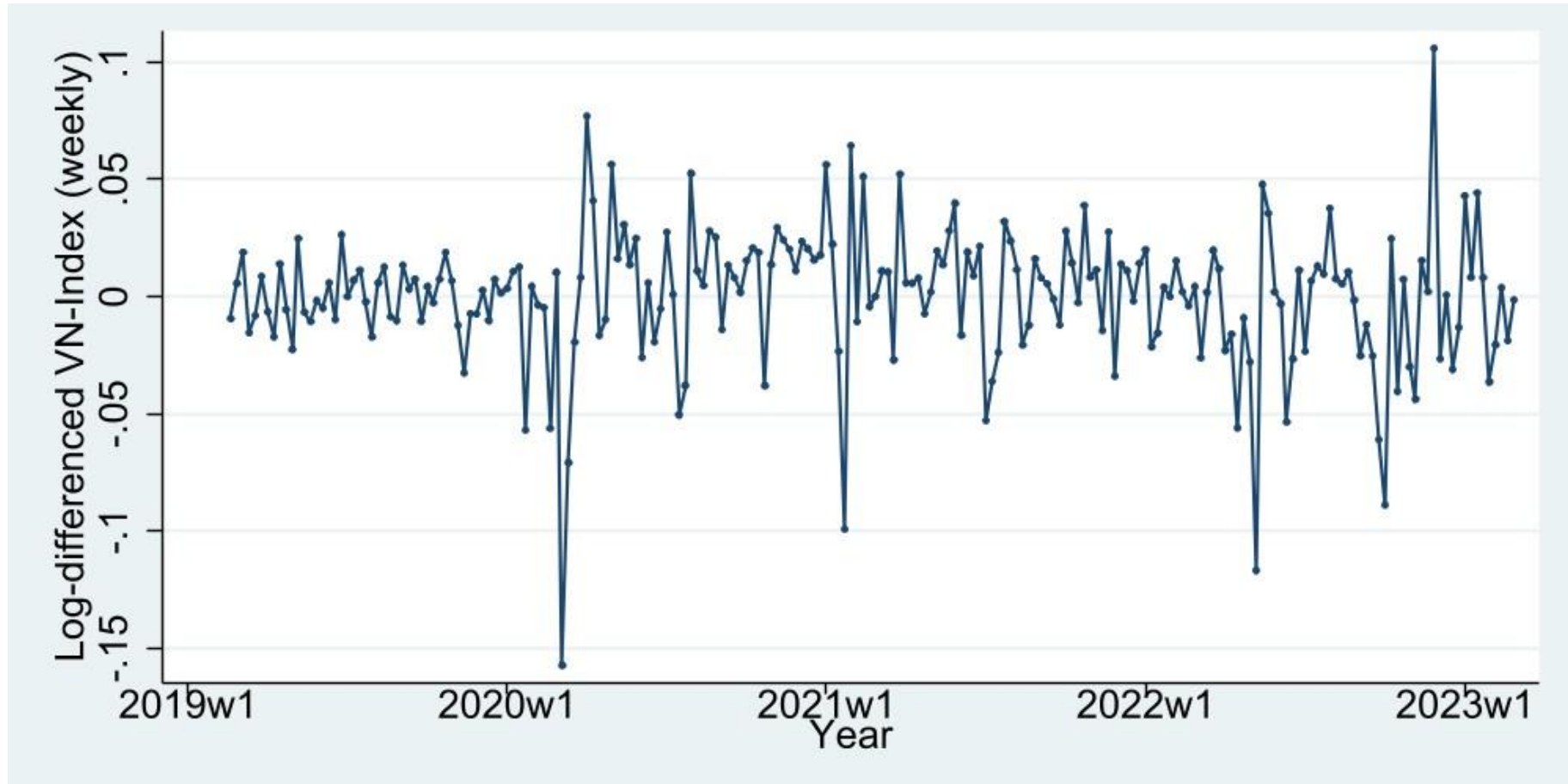
$$y'_t = \log(y_t) - \log(y_{t-1})$$

- Đây gọi là phương pháp “log differences”.
- Dễ luận giải hơn sai phân đơn thuần:

$$\log(y_t) - \log(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}} = \% \text{ change}$$

- Ví dụ: chuỗi CPI, chuỗi GDP, giá cổ phiếu...
- Ổn định phương sai.

Ví dụ: Chuỗi VN-Index (3) – Sai Phân Bậc 1 Với Hàm Log



Minh Họa: Kết Quả Kiểm Định ADF

$\Delta y_t = y_t - y_{t-1}$		Dickey–Fuller critical value		
	Test statistic	1%	5%	10%
With drift	-13.347	-2.344	-1.652	-1.286
With trend	-13.346	-4.003	-3.435	-3.135
With drift and 4 lags	-6.091	-2.345	-1.652	-1.286
$y'_t = \log(y_t) - \log(y_{t-1})$		Dickey–Fuller critical value		
	Test statistic	1%	5%	10%
With drift	-13.086	-2.344	-1.652	-1.286
With trend	-13.074	-4.003	-3.435	-3.135
With drift and 4 lags	-6.186	-2.345	-1.652	-1.286

Lưu Ý:

- Chuỗi dừng được kí hiệu $I(0)$.
- Chuỗi cần lấy sai phân 1 lần để dừng được kí hiệu $I(1)$.
- Chuỗi cần lấy sai phân 2 lần để dừng được kí hiệu $I(2)$.
- Thông thường chúng ta không cần lấy sai phân tới quá 2 lần để đạt tính dừng.
- Nếu chuỗi không dừng do tính mùa vụ? Chúng ta sẽ học cách xử lý ở phần dự báo.