

Giới Thiệu Học Phần Dữ Liệu Chuỗi Thời Gian

Giới Thiệu Giảng Viên

- Giảng viên phụ trách Đỗ Hoàng Phương.
- Quá trình đào tạo: cử nhân ĐH Ngoại Thương, Thạc sĩ tại University of Trento (Italy), and PhD in Economics tại Virginia Tech (USA).
- Lĩnh vực nghiên cứu chính: kinh tế học vĩ mô, chính sách công nghiệp và kinh tế chính trị Việt Nam.
- Giờ văn phòng: thứ năm 15:30-17:00, hoặc email.
- Địa chỉ email: hoangphuong.do@fulbright.edu.vn

Phát Họa Nội Dung Phần 2 – Dữ Liệu Chuỗi Thời Gian (Time Series)

- Nội dung chủ yếu:
 - Các khái niệm cơ bản liên quan đến chuỗi thời gian.
 - Các mô hình cơ bản liên quan tới chuỗi thời gian.
 - Dự báo trong chuỗi thời gian.
 - Một số vấn đề thực tế xử lý dữ liệu và nghiên cứu.
- Phần mềm sử dụng: STATA.
- Cấu trúc: buổi ngày thứ ba và năm tập trung vào bài giảng, buổi ngày thứ sáu tập trung vào thực hành với STATA.

Đề Cương

- Tuần 9: Các khái niệm cơ bản về chuỗi thời gian.
- Tuần 10+11: Các mô hình xoay quanh phương pháp OLS cho dữ liệu chuỗi dừng.
- Tuần 11+12: Các mô hình liên quan đến chuỗi không dừng. Dự báo trong dữ liệu chuỗi thời gian.
- Tuần 13+14: Một số vấn đề thực hành và ôn tập.

Mục Tiêu Môn Học và Đánh Giá

- Mục tiêu:
 - Trang bị kiến thức căn bản về dữ liệu chuỗi thời gian.
 - Đọc hiểu những nghiên cứu sử dụng dữ liệu chuỗi thời gian.
 - Kỹ năng căn bản xử lý và chạy mô hình dữ liệu chuỗi thời gian với STATA.
 - Có khả năng thực hiện những dự báo cơ bản phục vụ cho các nghiên cứu chính sách.
- Đánh giá:
 - Bài tập 3: phát 30/03, hạn nộp 25/04 8am.
 - Bài tập 4: phát 25/04, hạn nộp 05/05 8am.
 - Thi cuối kì: phát 11/05, hạn nộp 16/05 8am.
 - Đánh giá trên lớp: ngẫu nhiên.

Sách Tham Khảo

- *Basic Econometrics*, 5th edition của Gujarati Damodar N. và Dawn C. Porter, NXB McGraw-Hill, 2008. **GP5**
- *Introductory Econometrics: A Modern Approach*, 5th edition của Jeffrey M. Wooldridge. **JW**
- *Introduction to Econometrics*, 4th edition của James H. Stock và Mark W. Watson, Princeton University, NXB Pearson. **SW4**

Các Khái Niệm Cơ Bản Về Chuỗi Thời Gian (1)

Nội Dung Buổi Học

- Dữ liệu chuỗi thời gian
- Các thành tố cơ bản (patterns) của chuỗi thời gian.
- Quá trình ngẫu nhiên (stochastic processes).
- Tự tương quan và correlogram.

Dữ liệu chuỗi thời gian

Dữ liệu chuỗi thời gian

Ba loại dữ liệu thời gian chủ yếu:

- Dữ liệu chéo.
- Dữ liệu bảng.
- Dữ liệu chuỗi thời gian.
 - Dữ liệu kinh tế vĩ mô, dữ liệu tài chính, dữ liệu kinh doanh.

Ví dụ:

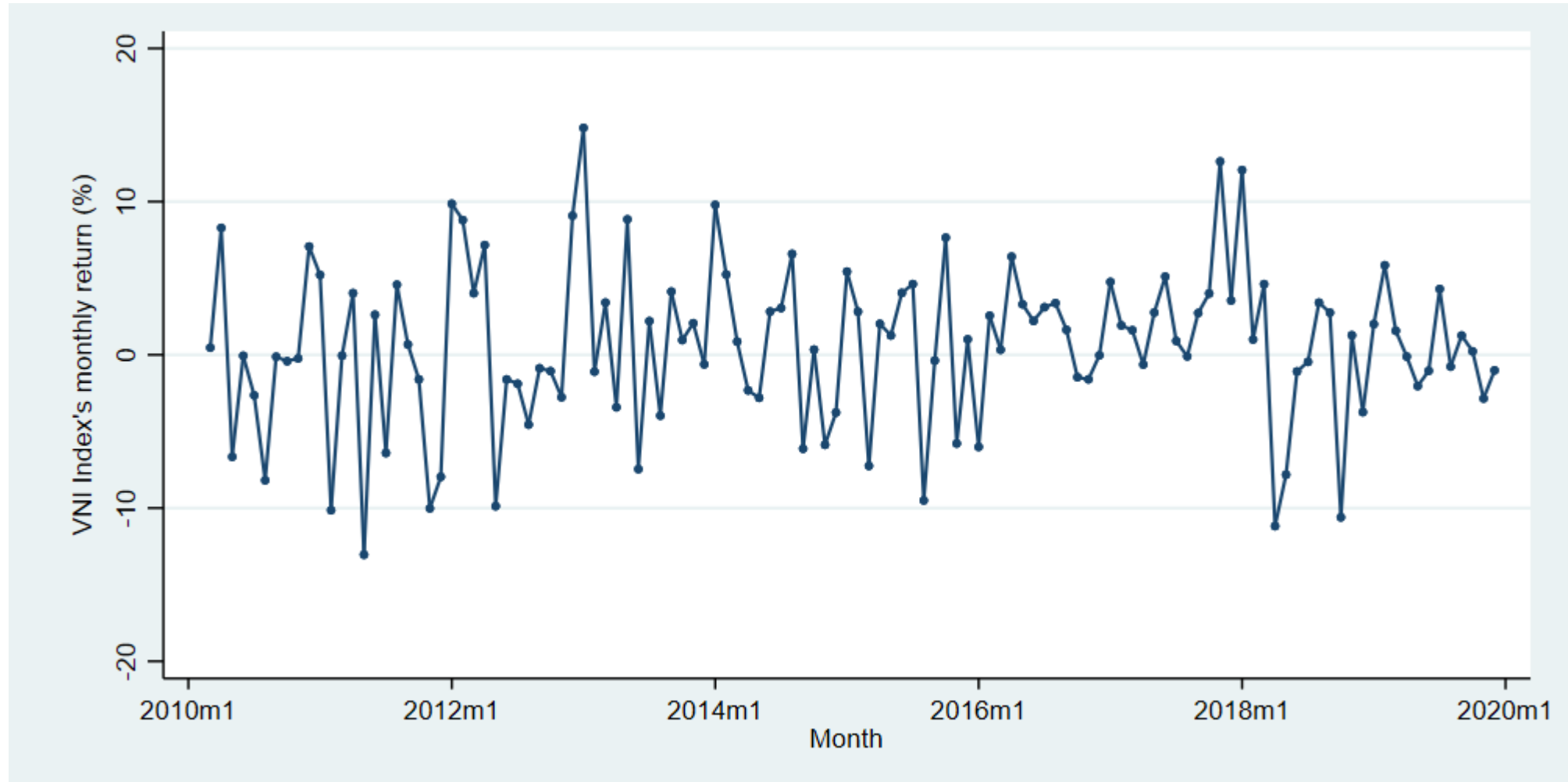
Dữ liệu chuỗi thời gian

	▲ Year ▼	CPI ▼
1	2003	103.18
2	2004	107.71
3	2005	108.29
4	2006	107.48
5	2007	108.30
6	2008	122.97
7	2009	106.88
8	2010	109.19
9	2011	118.58
10	2012	109.21
11	2013	106.60
12	2014	104.09
13	2015	100.63
14	2016	102.66
15	2017	103.53
16	2018	103.54
17	2019	102.79
18	2020	103.23

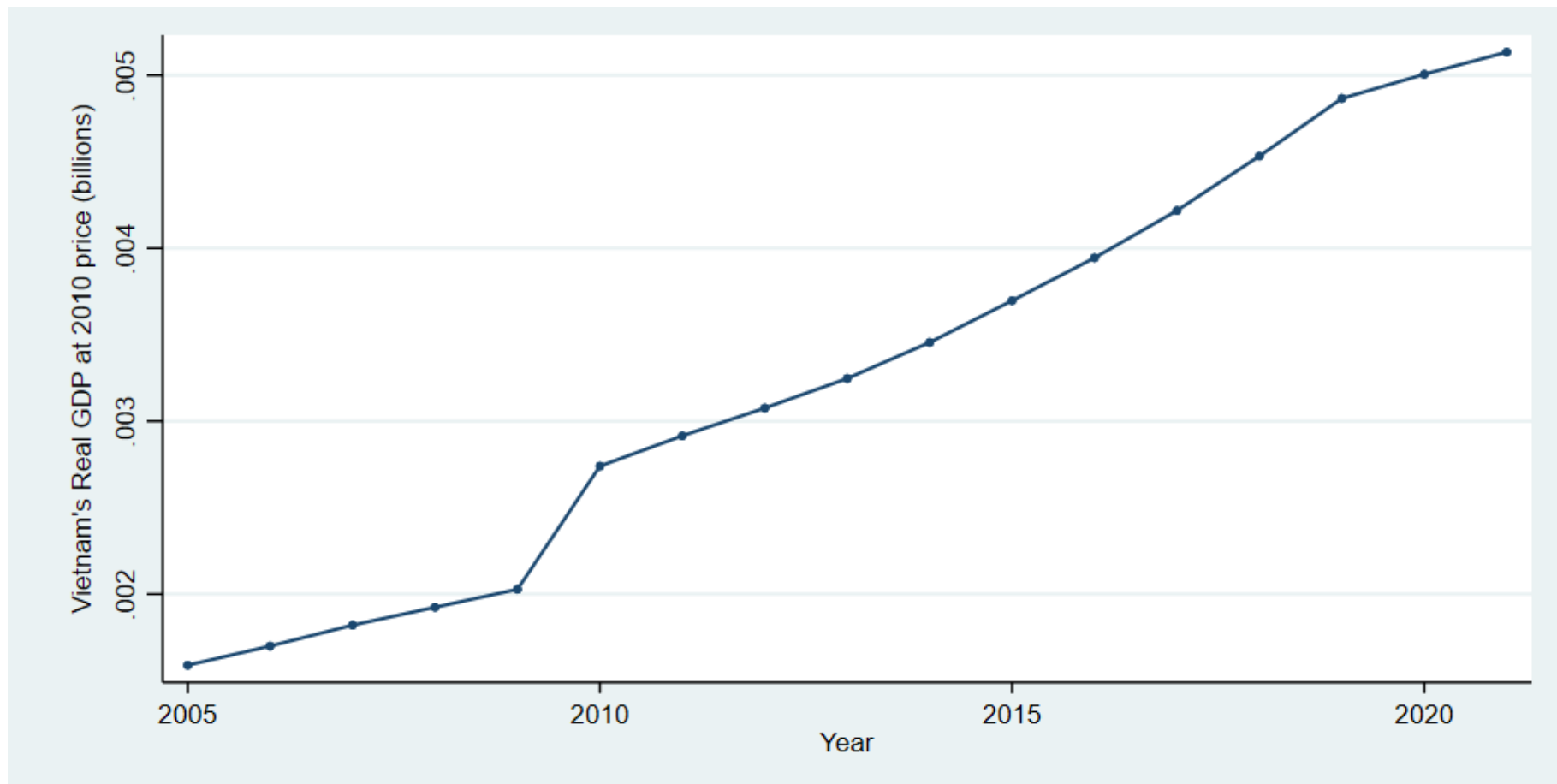
Dữ liệu chéo

	▲ country ▼	lifeExp ▼	pop ▼	gdpPercap ▼	continent ▼
1	Afghanistan	43.828	31889923	974.5803	Asia
2	Albania	76.423	3600523	5937.0295	Europe
3	Algeria	72.301	33333216	6223.3675	Africa
4	Angola	42.731	12420476	4797.2313	Africa
5	Argentina	75.320	40301927	12779.3796	Americas
6	Australia	81.235	20434176	34435.3674	Oceania
7	Austria	79.829	8199783	36126.4927	Europe
8	Bahrain	75.635	708573	29796.0483	Asia
9	Bangladesh	64.062	150448339	1391.2538	Asia
10	Belgium	79.441	10392226	33692.6051	Europe
11	Benin	56.728	8078314	1441.2849	Africa
12	Bolivia	65.554	9119152	3822.1371	Americas
13	Bosnia and Herzegovina	74.852	4552198	7446.2988	Europe
14	Botswana	50.728	1639131	12569.8518	Africa
15	Brazil	72.390	190010647	9065.8008	Americas
16	Bulgaria	73.005	7322858	10680.7928	Europe
17	Burkina Faso	52.295	14326203	1217.0330	Africa
18	Burundi	49.580	8390505	430.0707	Africa
19	Cambodia	59.723	14131858	1713.7787	Asia

Minh họa (1)



Minh họa (2)

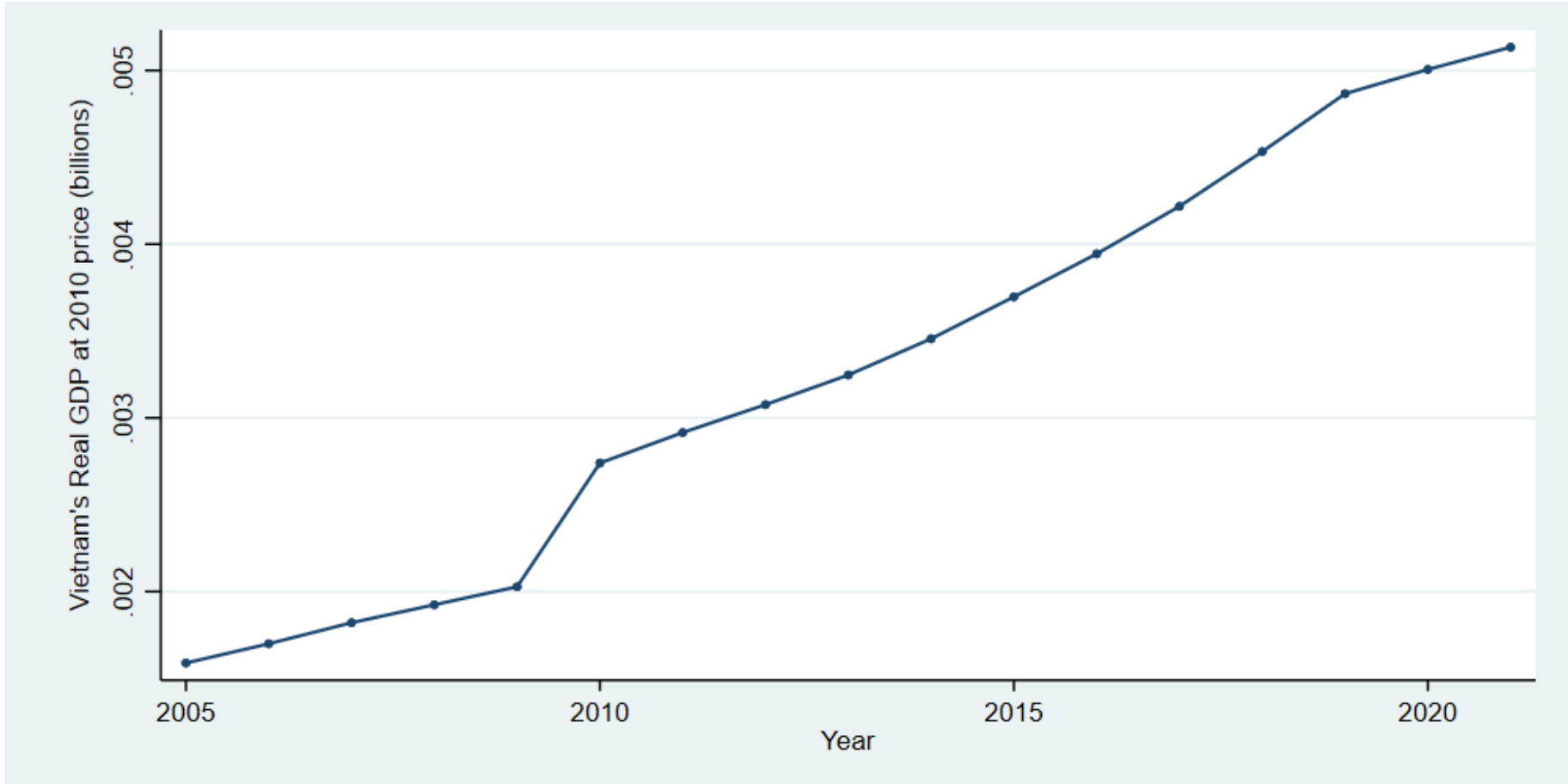


Các Dạng Mẫu Cơ Bản (Patterns) Chuỗi Thời Gian

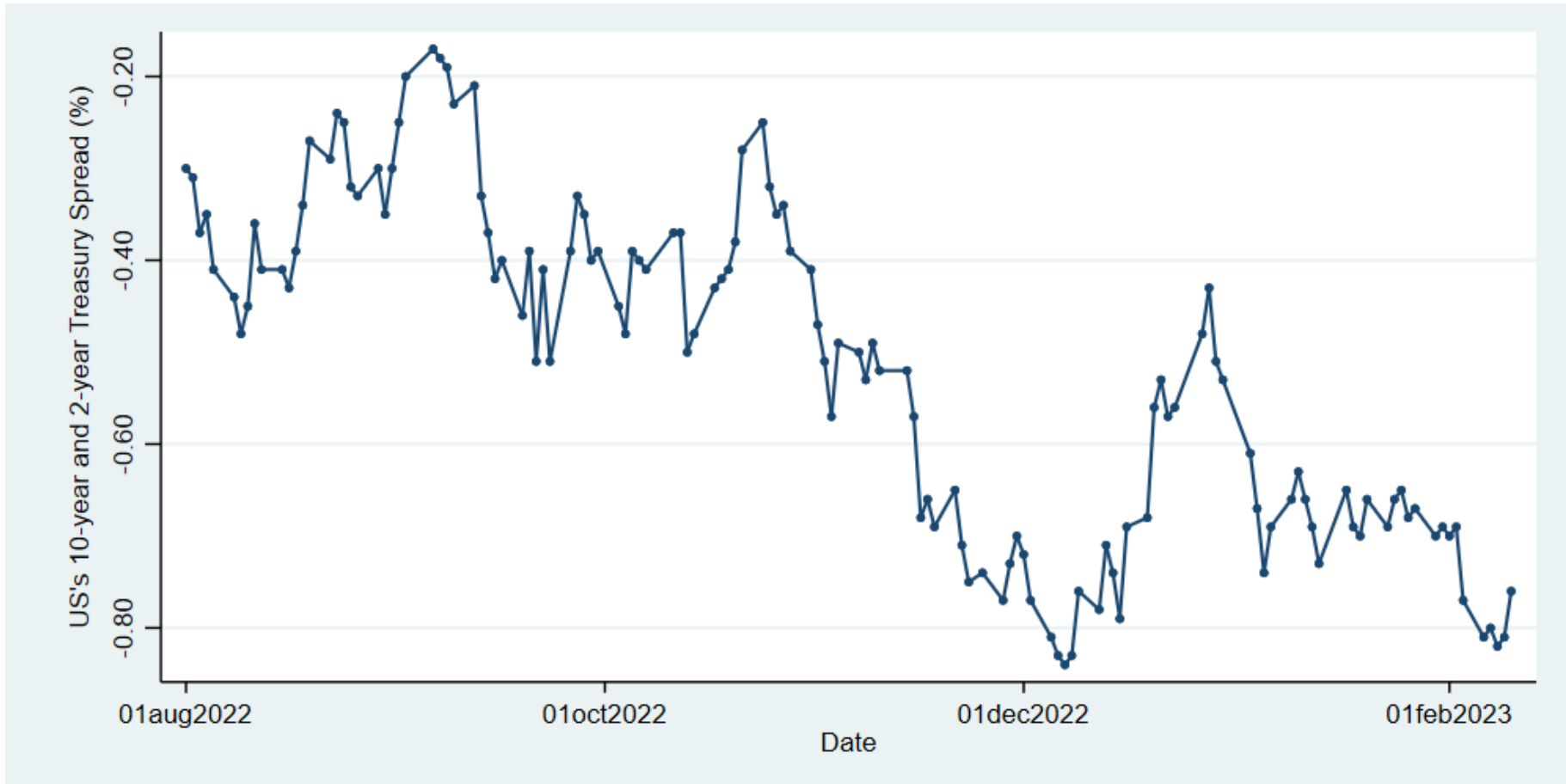
Các Dạng mẫu Cơ Bản (Patterns) Của Chuỗi Thời Gian

- Xu hướng/xu thế (trend): dạng mẫu tồn tại dưới dạng xu hướng tăng hay giảm dài hạn.
- Mùa vụ theo chu kì (seasonal): dạng mẫu tồn theo chu kì cố định (theo tuần, tuần tháng hoặc theo quý)
- Tuần hoàn bất chu kì (cyclic): dạng mẫu tăng **giảm không theo một chu kì cố định** nào (thường kéo dài trên 1 năm).

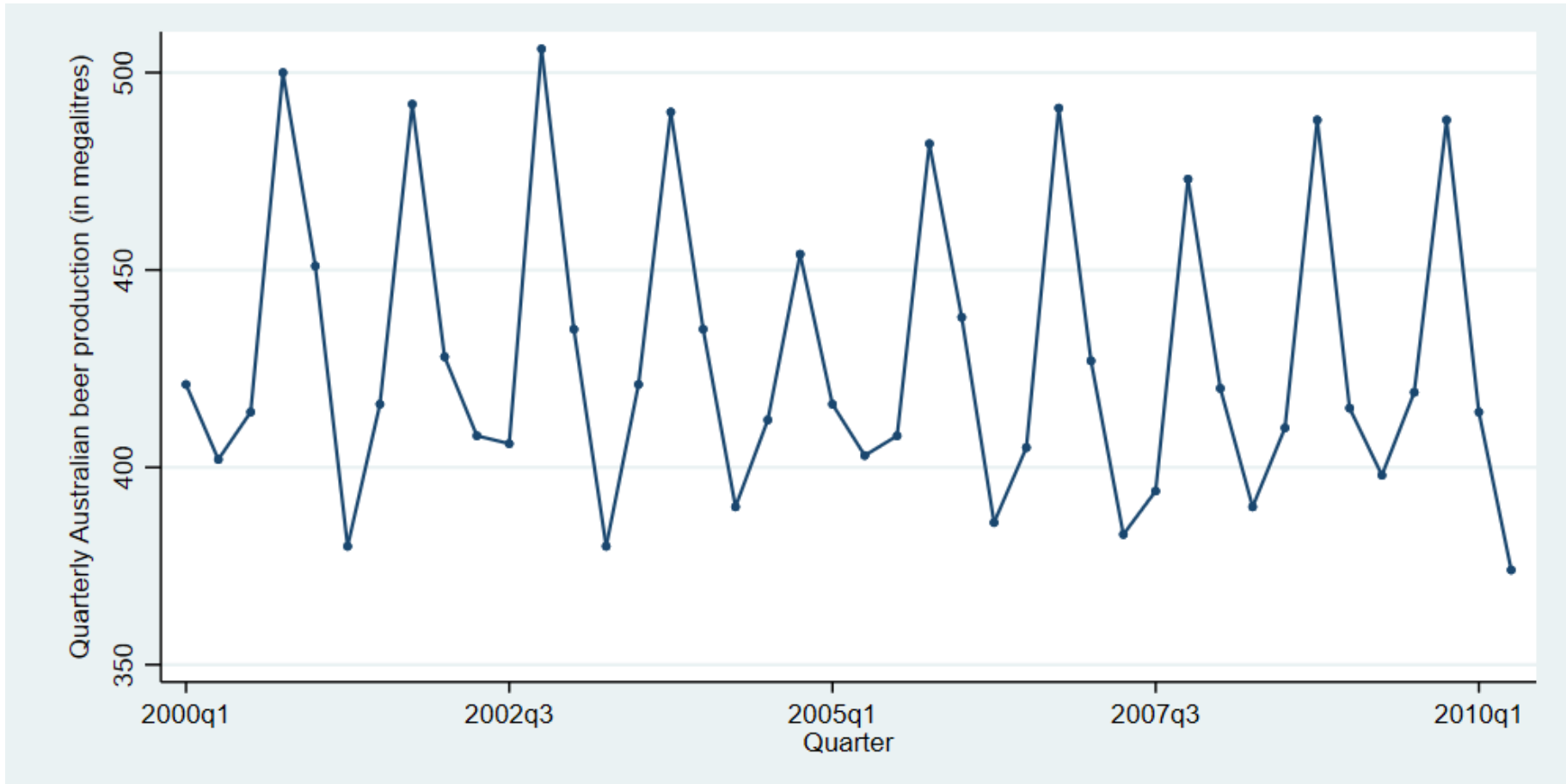
Minh Họa (1)



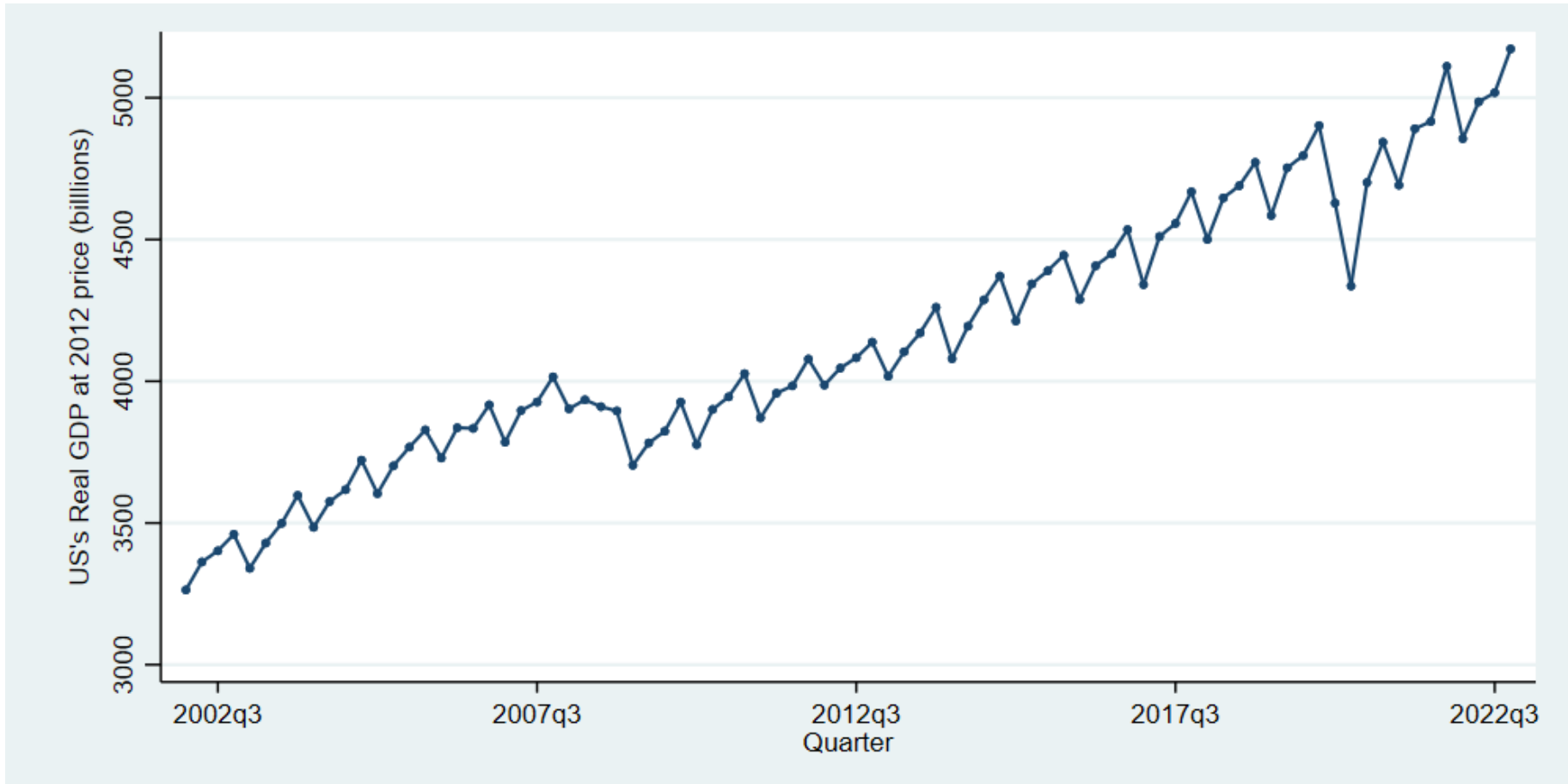
Minh Họa (2)



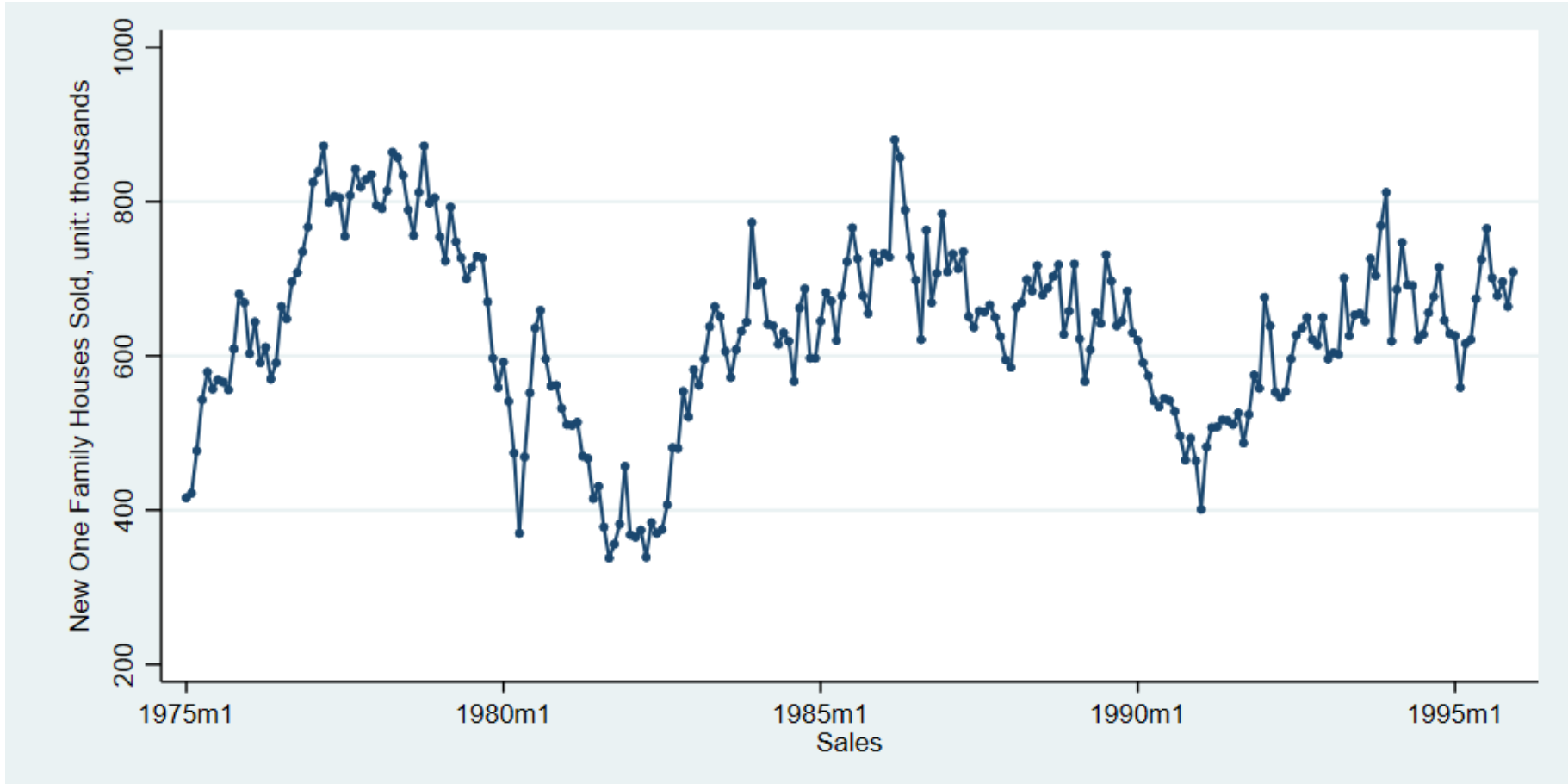
Minh Họa (3)



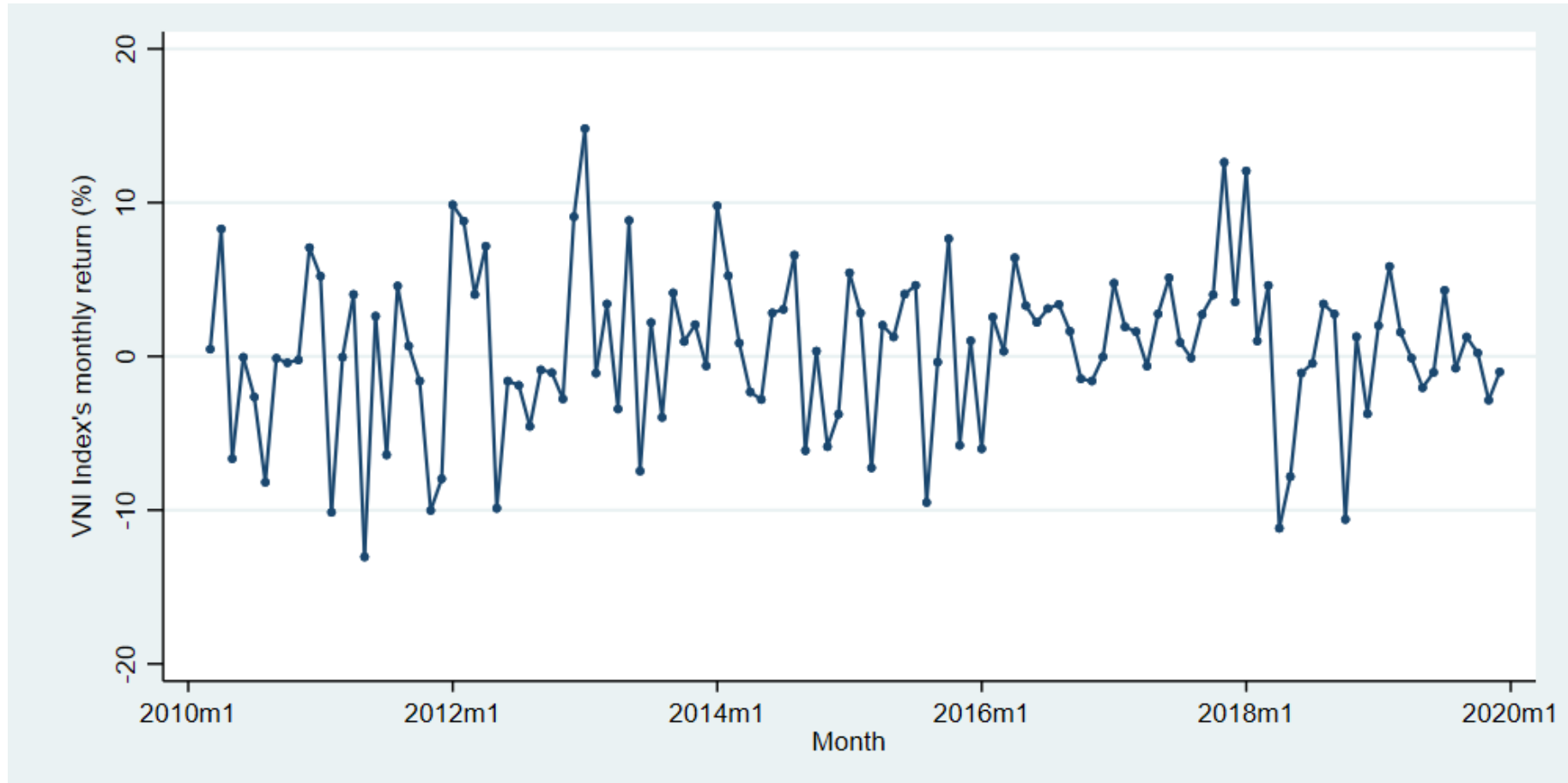
Minh Họa (4)



Minh Họa (5)



Minh Họa (6)



Quá Trình Ngẫu Nhiên (Stochastic Processes)

Quá Trình Ngẫu Nhiên (Stochastic Processes)

- Stochastic processes (quá trình ngẫu nhiên): là một tập hợp các biến ngẫu nhiên theo trình tự thời gian.
- Về mặt toán học:

$$y_1, y_2, y_3 \dots, y_n$$

- Với mỗi y_t ($i = 1, 2, \dots, n$) được xem như là một biến ngẫu nhiên.
- Một giá trị thực là được xem như là hiện thực hóa (realization) của biến ngẫu nhiên y_t .

Tương Quan Chuỗi (Autocorrelation)

Hệ Số Tương Quan (Correlation)

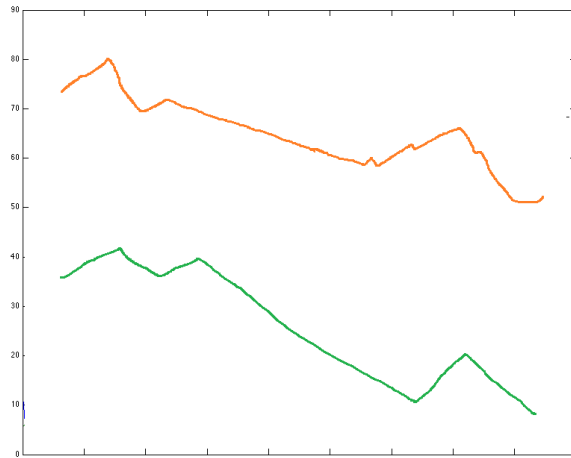
- Hệ số tương quan là một chỉ số đo lường liên hệ tuyến tính thống kê giữa hai biến ngẫu nhiên.

$$\rho_{xy} = \text{corr}(x, y) = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

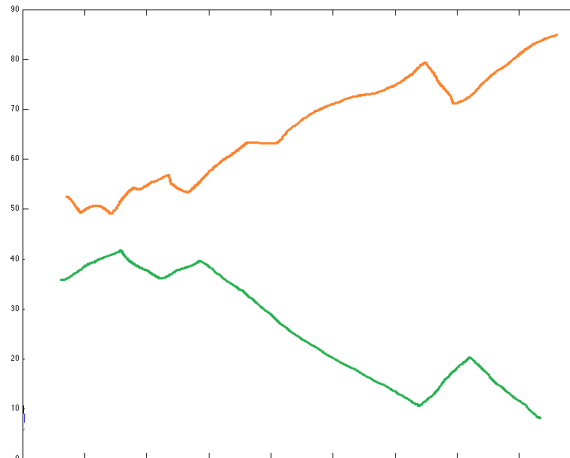
- Hệ số tương quan mang giá trị giữa -1 và 1:
 - Hệ số gần 1 hoặc -1: tương quan rất cao.
 - Hệ số gần 0: Hầu như không có mối liên hệ.

Minh họa

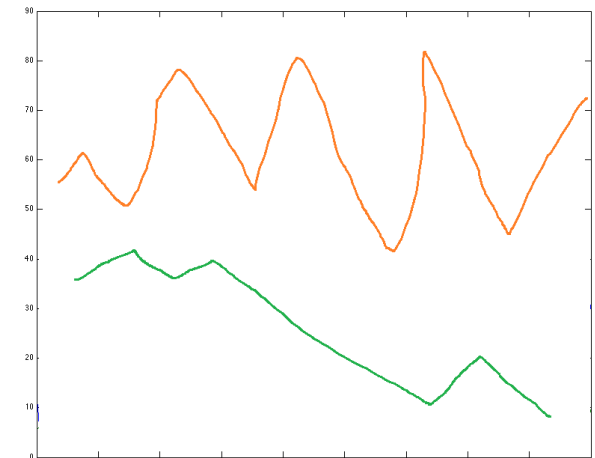
Tương quan rất cao
dương ($\approx +1$)



Tương quan rất
cao âm (≈ -1)



Hầu như không
tương quan (≈ 0)



Absolute value of correlation

$r < 0.25$

$0.25 < r < 0.5$

$0.5 < r < 0.75$

$r > 0.75$

Strength of relationship

No relationship

Weak relationship

Moderate relationship

Strong relationship

Tương Quan Cao Hay Thấp?

- Thu nhập và chi tiêu hằng tháng của từng cá nhân.
- Nhiệt độ trung bình hằng tháng và doanh số trà đá vỉa hè Hà Nội.
- Thời gian xem Tiktok hằng ngày và nhiệt độ hằng tháng.
- Số lượng stories đăng lên Instagram mỗi ngày và số lượng crush của một bạn nữ.
- Chiều cao của đứa trẻ 5 tuổi và dân số Việt Nam trong vòng 10 năm tiếp theo.

Biến Trễ (Lagged Variables)

- y_t là chuỗi thời gian.
- Biến trễ của y_t là $y_{t-1}, y_{t-2}, y_{t-3} \dots y_{t-k}$.
- Biến trễ dùng để nghiên cứu quá khứ ảnh hưởng tới hiện tại thế nào trong phân tích chuỗi thời gian.

t	y_t	y_{t-1}	y_{t-2}
$t = 1$	5	.	.
$t = 2$	8	5	.
$t = 3$	6	8	5
$t = 4$	4	6	8
$t = 5$	3	4	6
$t = 6$	3	3	4
$t = 7$	7	3	3

Tự Tương Quan (Autocorrelation) - 1

- **Tự tương quan (autocorrelation):** đo lường mức độ tương quan (correlation) giữa một biến và biến trễ của nó.
- Hệ số này dùng để đo lường mức độ giống nhau (đồng dạng) giữa một chuỗi thời gian và biến trễ của nó
- Kí hiệu toán học:

$$\text{corr}(y_t, y_{t-i}) \quad i = 1, 2, 3, \dots$$

Tự Tương Quan (Autocorrelation) - 2

Một loạt tự tương quan với độ trễ tăng dần được tính toán:

$$\rho_1 = \text{corr}(y_t, y_{t-1})$$

$$\rho_2 = \text{corr}(y_t, y_{t-2})$$

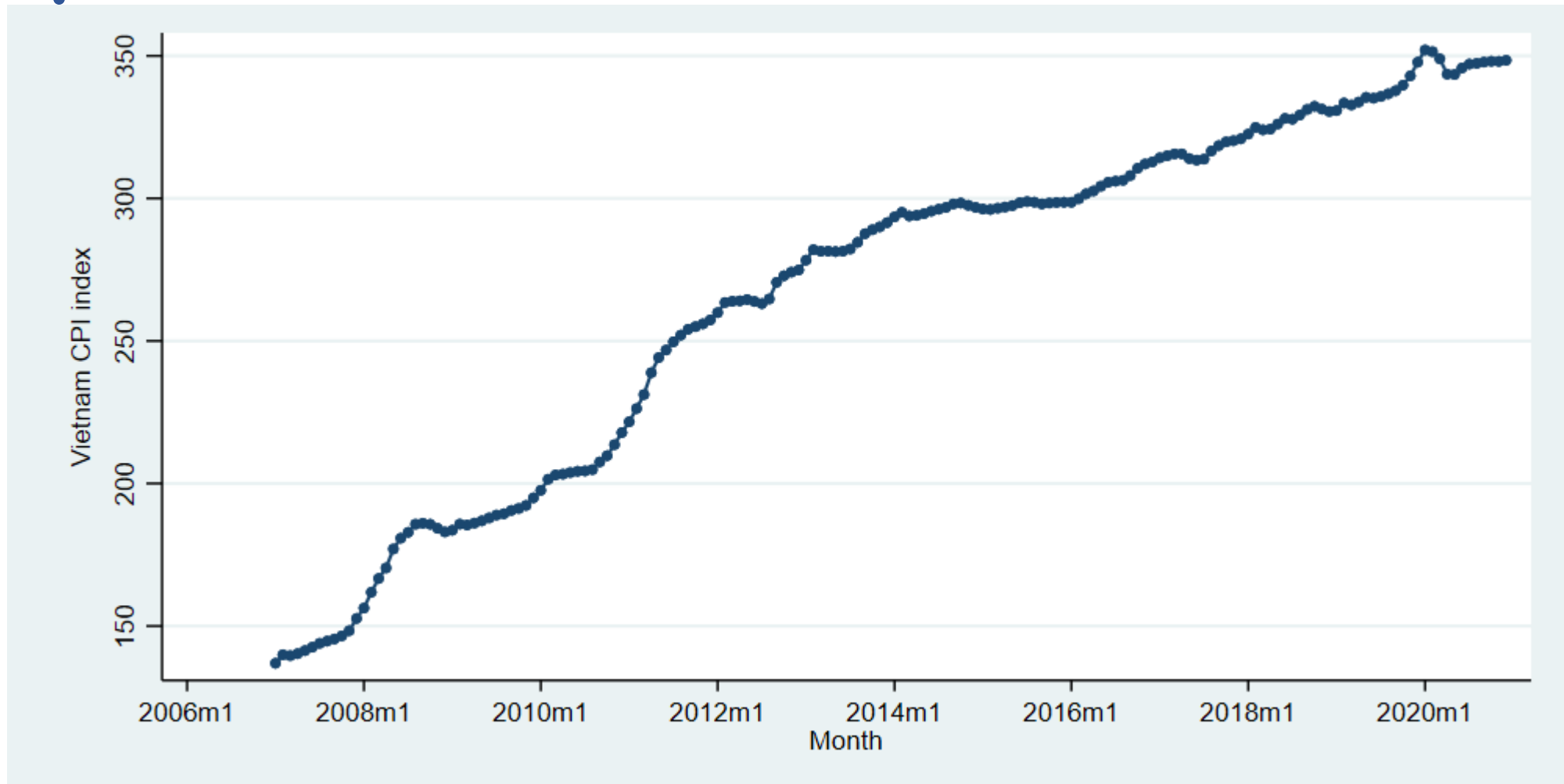
$$\rho_3 = \text{corr}(y_t, y_{t-3})$$

$$\rho_4 = \text{corr}(y_t, y_{t-4})$$

....

Correlogram là một biểu đồ tổng hợp những tự tương quan với độ trễ tăng dần.

Tự Tương Quan (Autocorrelation) - Ví dụ 1 (1)



Tự Tương Quan (Autocorrelation) - Ví dụ 1 (2)

- Các giá trị Q đến từ kiểm định Ljung – Box (Ljung – Box test).
- Kiểm định Ljung – Box kiểm tra xem tính tự tương quan của chuỗi:

H_0 : Chuỗi không có tự tương quan.

H_1 : Chuỗi có tự tương quan.

- Giá trị p -value 0.000 có ý nghĩa gì?

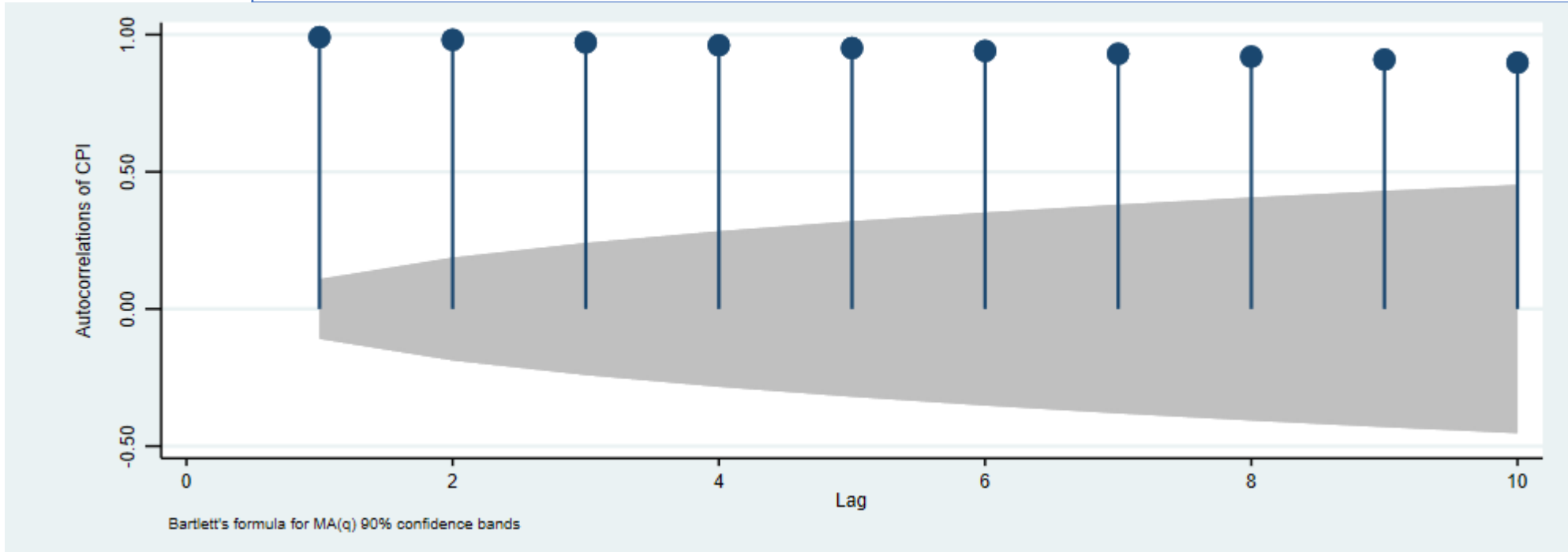
```
. * Autocorrelation
. corrgram CPI, lags(36)
```

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial autocor]		
1	0.9906	0.9995	226.71	0.0000						
2	0.9811	-0.5959	450.08	0.0000						
3	0.9713	0.0571	669.95	0.0000						
4	0.9612	-0.0530	886.26	0.0000						
5	0.9510	-0.0171	1098.9	0.0000						
6	0.9406	-0.0543	1307.9	0.0000						
7	0.9301	0.1031	1513.2	0.0000						
8	0.9196	-0.1005	1714.8	0.0000						
9	0.9090	0.0161	1912.6	0.0000						
10	0.8978	0.0106	2106.5	0.0000						
11	0.8862	-0.0095	2296.3	0.0000						
12	0.8745	-0.0425	2482	0.0000						
13	0.8629	-0.1392	2663.6	0.0000						
14	0.8517	0.2476	2841.4	0.0000						
15	0.8406	-0.0424	3015.3	0.0000						
16	0.8294	-0.0798	3185.5	0.0000						
17	0.8181	-0.1053	3351.8	0.0000						
18	0.8065	-0.0025	3514.2	0.0000						
19	0.7949	0.0063	3672.8	0.0000						
20	0.7831	-0.0461	3827.4	0.0000						
21	0.7712	0.0827	3978.1	0.0000						
22	0.7592	-0.1243	4124.8	0.0000						
23	0.7470	0.1141	4267.5	0.0000						
24	0.7349	-0.0663	4406.4	0.0000						
25	0.7228	-0.0656	4541.3	0.0000						
26	0.7107	0.1919	4672.4	0.0000						
27	0.6984	0.0230	4799.7	0.0000						
28	0.6861	-0.0748	4923.1	0.0000						
29	0.6739	-0.0047	5042.8	0.0000						
30	0.6617	0.0665	5158.7	0.0000						
31	0.6493	-0.1604	5271	0.0000						

Command

Tự Tương Quan (Autocorrelation) – Ví dụ 1 (3)

$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\rho}_4$	$\hat{\rho}_5$	$\hat{\rho}_6$	$\hat{\rho}_7$	$\hat{\rho}_8$	$\hat{\rho}_9$	$\hat{\rho}_{10}$
0.9906	0.9811	0.9713	0.9612	0.951	0.9406	0.9301	0.9196	0.909	0.8978



Lưu ý: Đồ thị này được gọi là Autocorrelation Function – Đồ thị ACF

Khoảng Tin Cây Bartlett (Bartlett's Confidence Interval)

- Khi đọc correlogram, làm sao chúng ta biết một giá trị tự tương quan $\hat{\rho}_k$ có giá trị thống kê hay không?
- Nếu giá trị $\hat{\rho}_k$ đến từ một chuỗi không có sự tương quan, phân phối của $\hat{\rho}_k$ sẽ xấp xỉ:

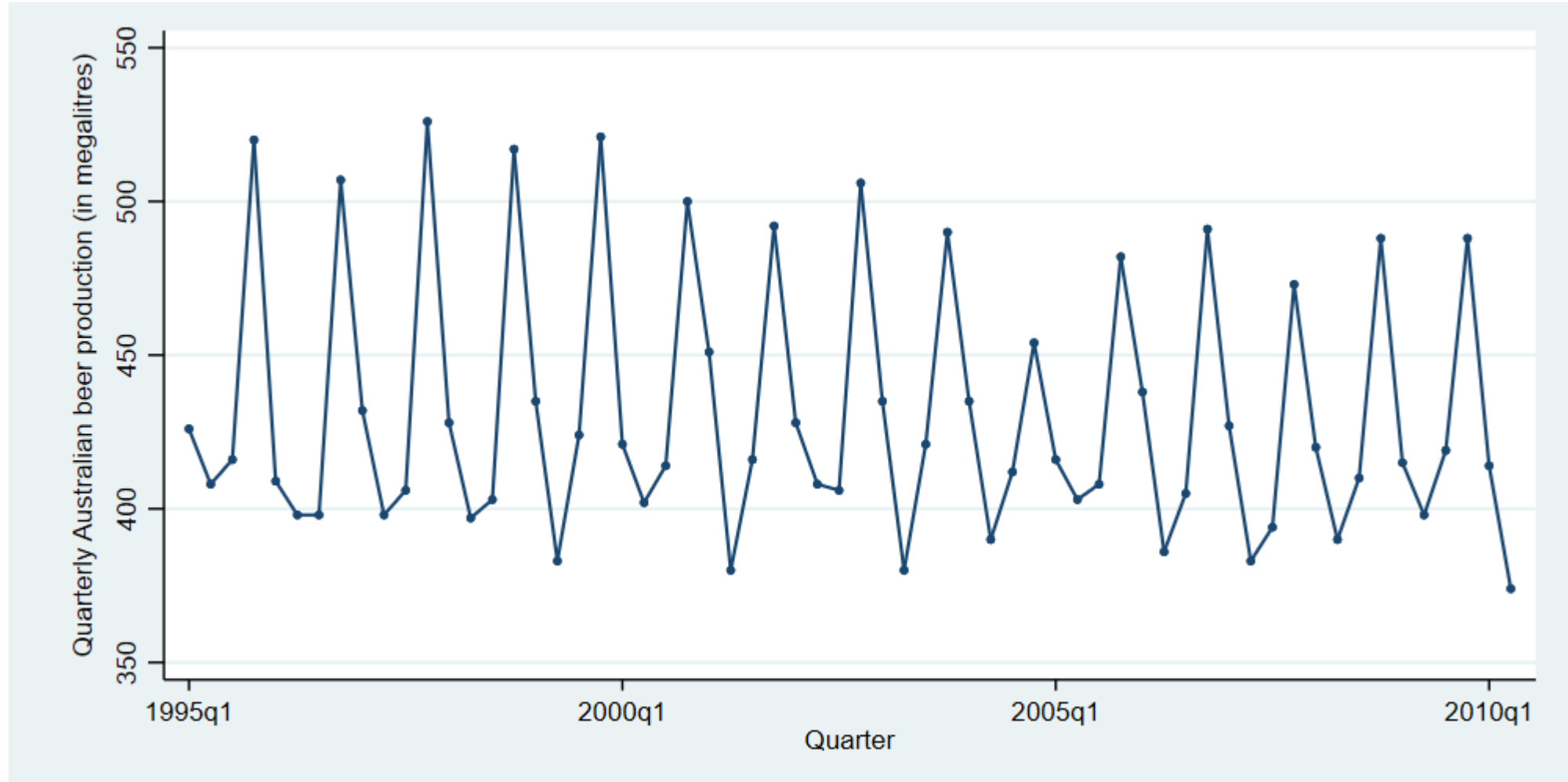
$$\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right)$$

- 95% confidence interval sẽ nằm trong khoảng:

$$\hat{\rho}_k \pm 1.96\sqrt{1/n}$$

- Ví dụ: $n = 200$, ta có $\pm 1.96\sqrt{1/200} = \pm 0.139$

Tự Tương Quan (Autocorrelation) - Ví dụ 2 (1)

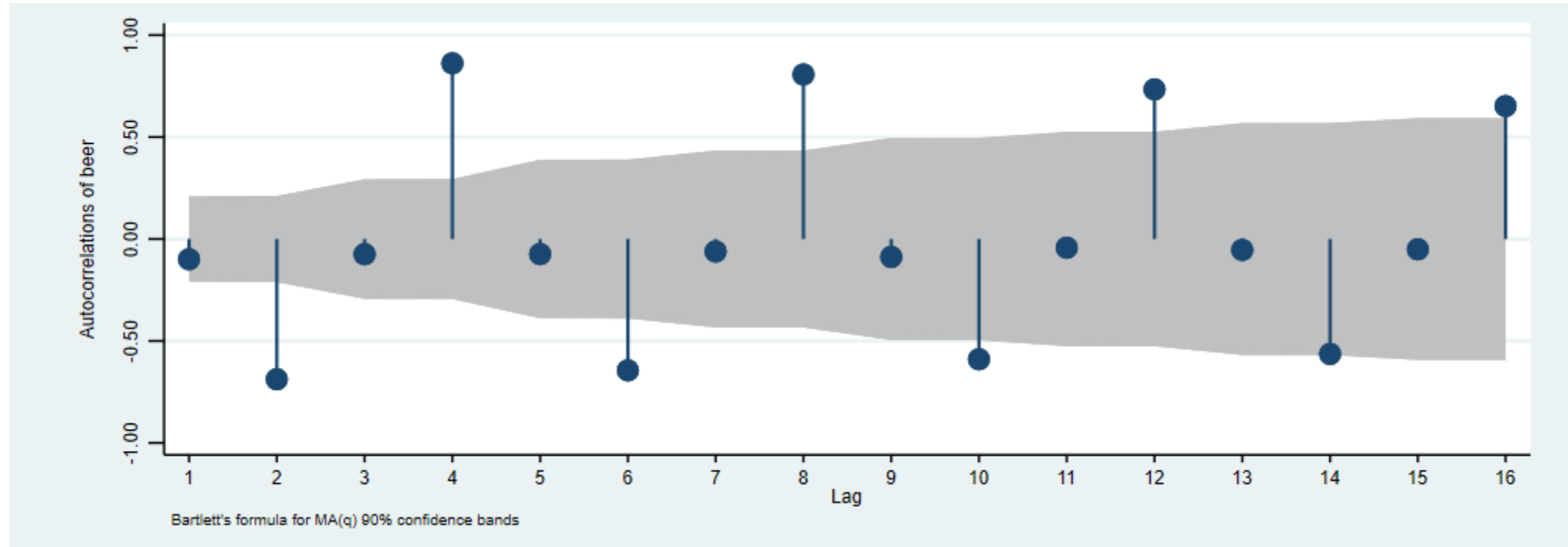


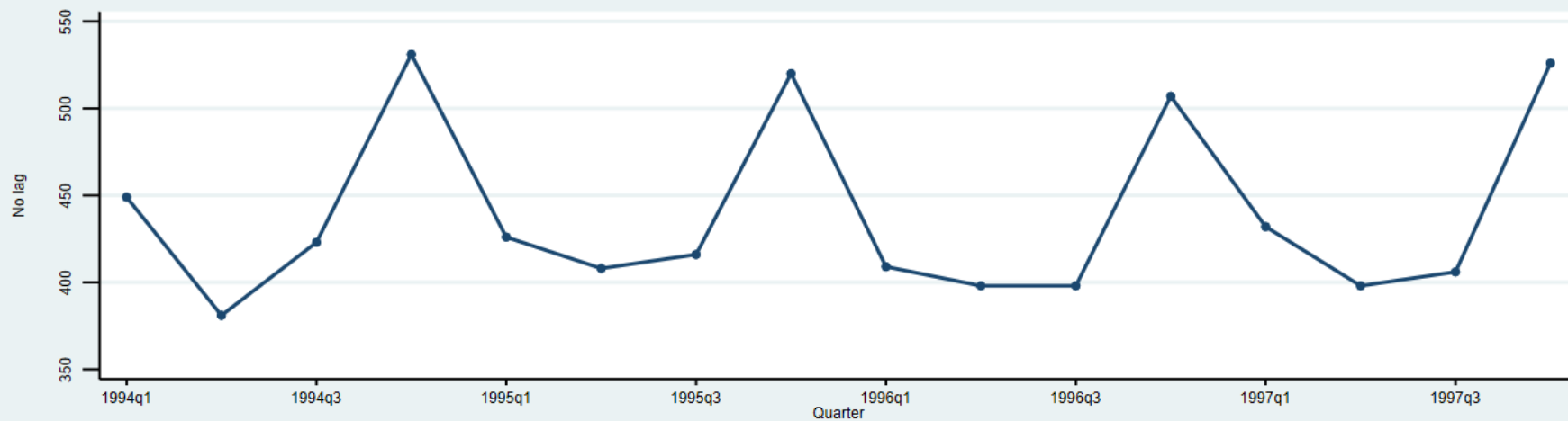
Tự Tương Quan (Autocorrelation) - Ví dụ 2 (2)

```
. corrgram beer if tin(1995q1,), lags(12)
```

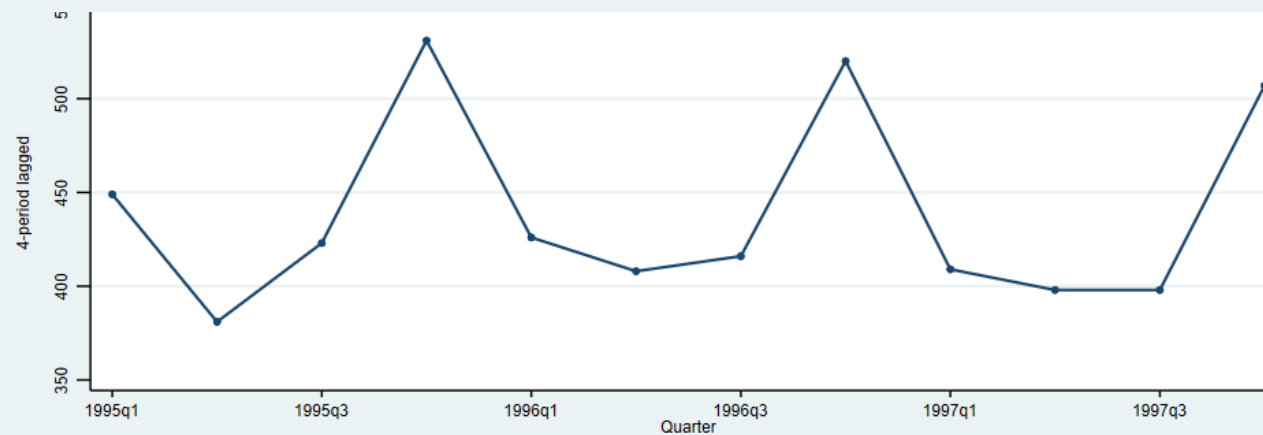
LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
					[Autocorrelation]			[Partial autocor]		
1	-0.0993	-0.1025	.64163	0.4231						
2	-0.6879	-0.7336	31.938	0.0000	—			—		
3	-0.0747	-0.6036	32.313	0.0000				—		
4	0.8616	0.7495	83.097	0.0000		—			—	
5	-0.0741	0.0975	83.479	0.0000						
6	-0.6444	-0.0221	112.9	0.0000	—					
7	-0.0610	-0.1863	113.17	0.0000				—		
8	0.8068	0.4593	161	0.0000		—			—	
9	-0.0879	-0.0801	161.58	0.0000						
10	-0.5898	0.1529	188.13	0.0000	—				—	
11	-0.0429	-0.0378	188.27	0.0000						
12	0.7337	0.4112	230.99	0.0000		—			—	

Tự Tương Quan (Autocorrelation) - Ví dụ 2 (3)

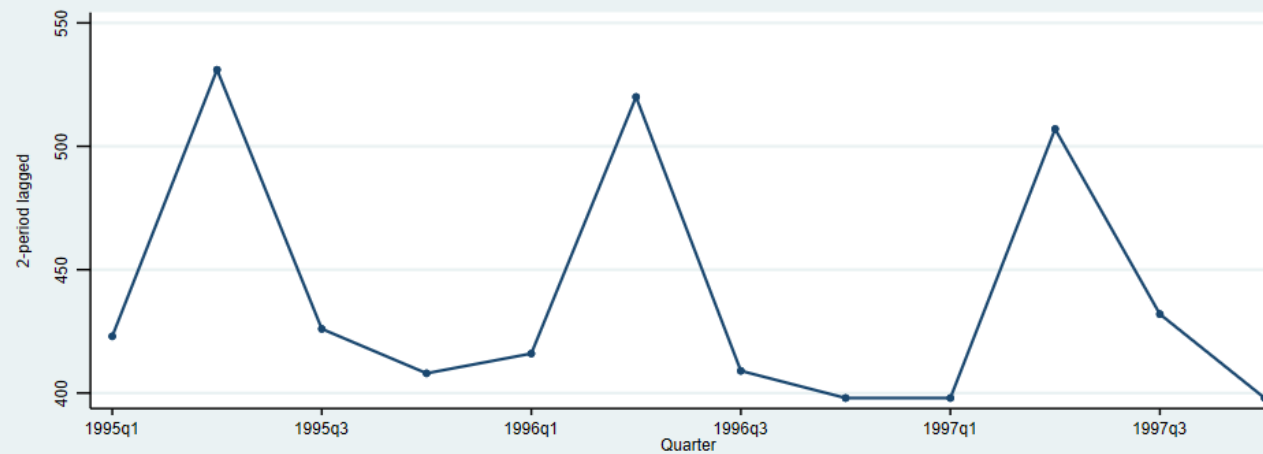




← y_t

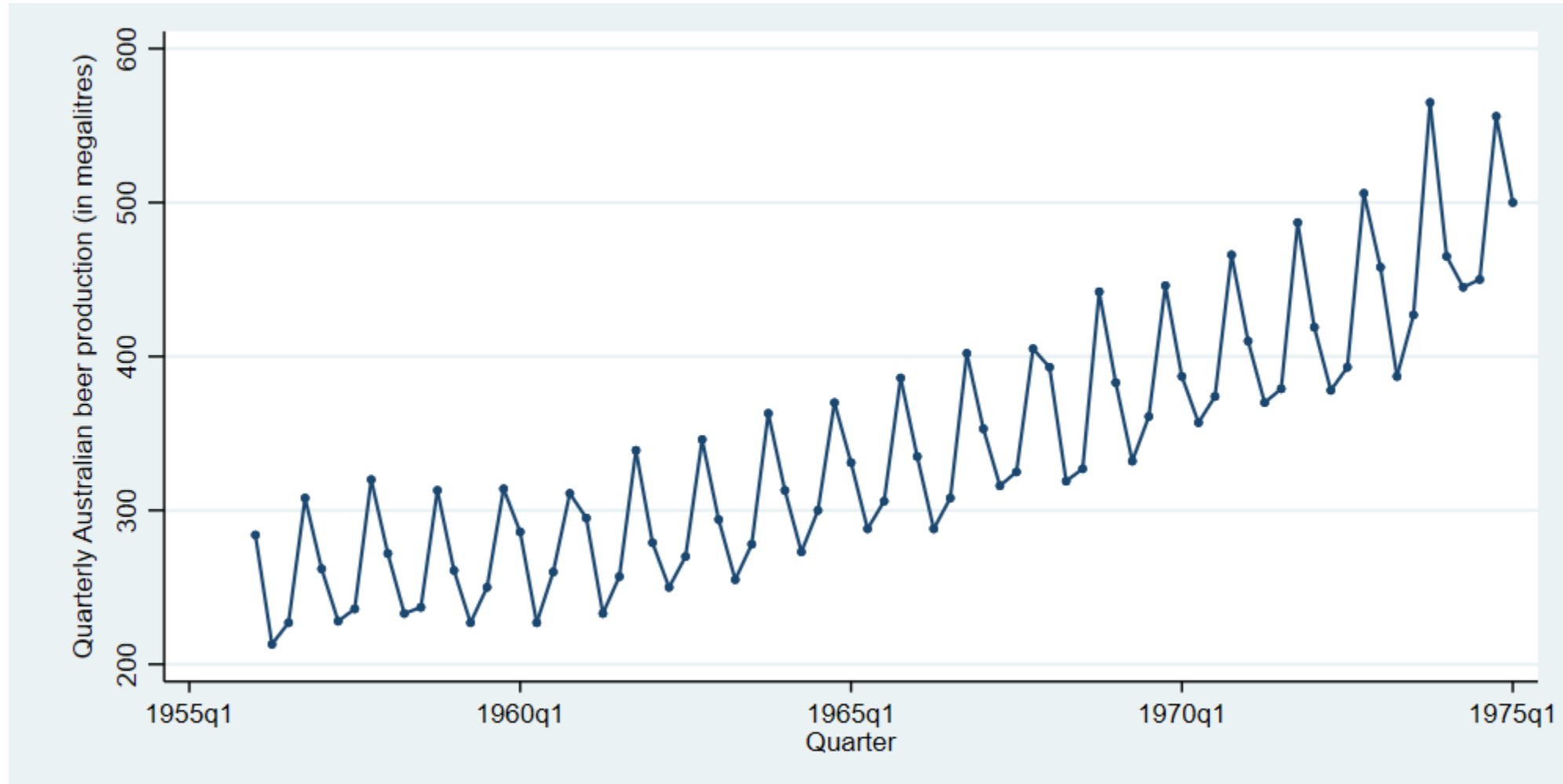


← y_{t-4}



← y_{t-2}

Tự Tương Quan (Autocorrelation) - Ví dụ 3 (1)



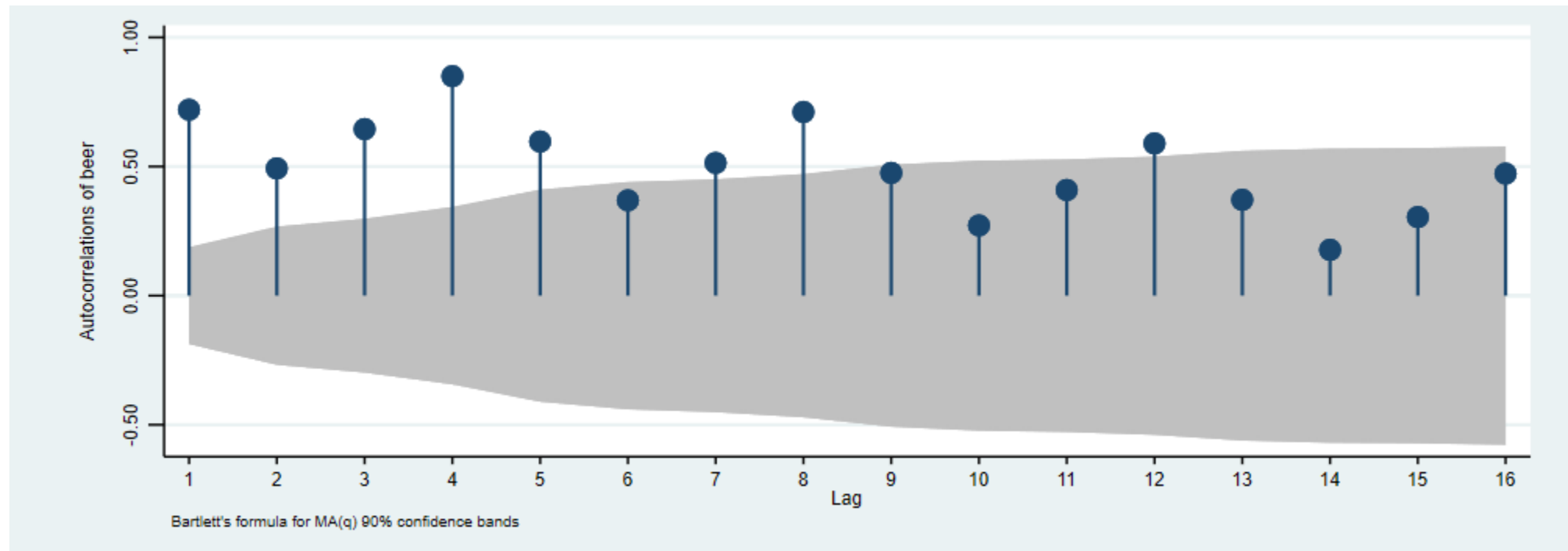
Tự Tương Quan (Autocorrelation) - Ví dụ 3 (2)

```
. corrgram beer if tin(,1975q1), lags(16)
```

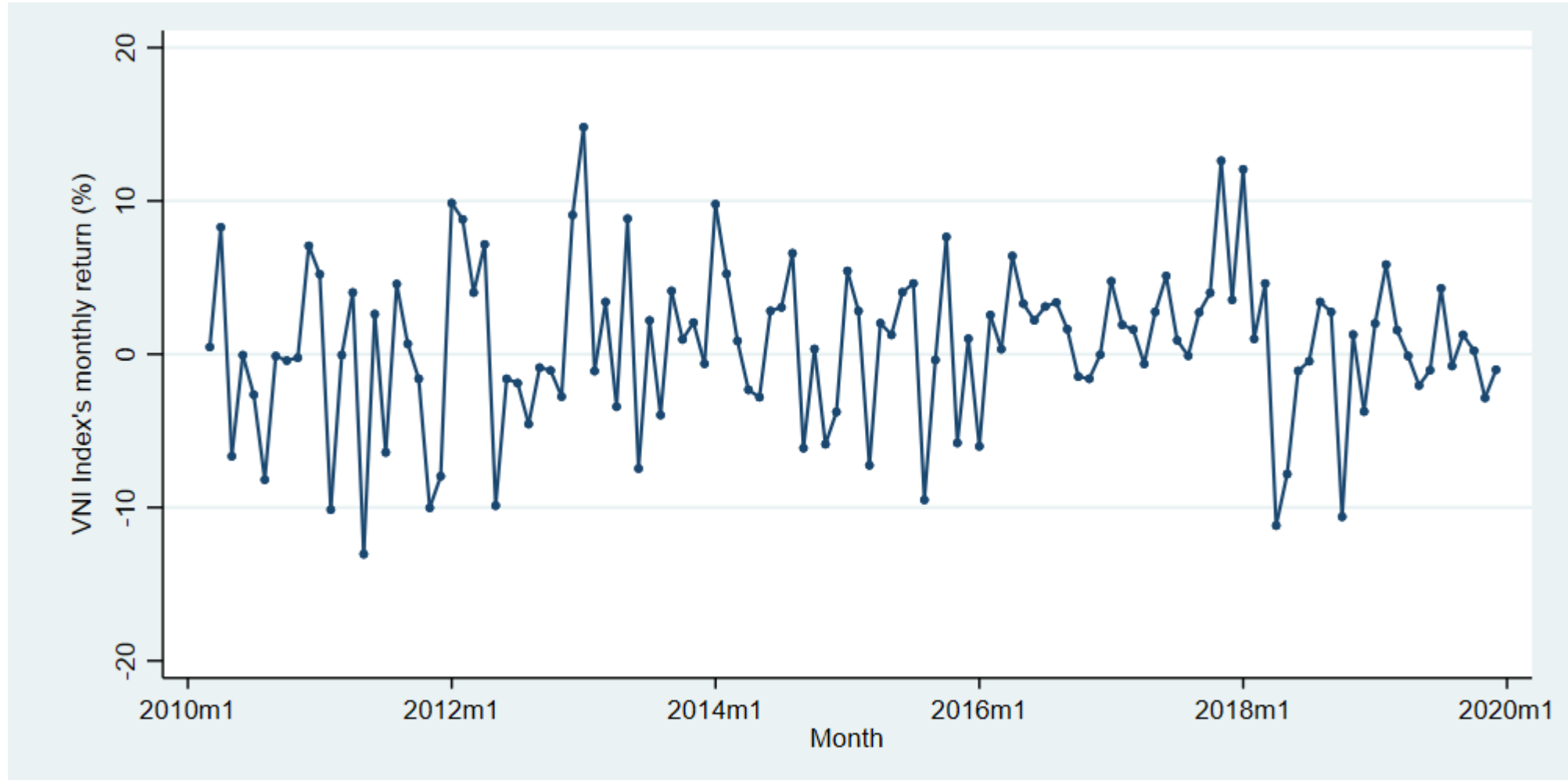
LAG	AC	PAC	Q	Prob>Q	-1 [Autocorrelation]	0 [Partial autocor]	1 [Partial autocor]
1	0.7201	0.7581	41.506	0.0000			
2	0.4927	0.0149	61.194	0.0000			
3	0.6449	0.7492	95.377	0.0000			
4	0.8499	0.9749	155.57	0.0000			
5	0.5968	0.2966	185.66	0.0000			
6	0.3690	0.0511	197.32	0.0000			
7	0.5140	-0.3524	220.28	0.0000			
8	0.7112	0.3477	264.87	0.0000			
9	0.4754	0.2874	285.08	0.0000			
10	0.2719	0.1140	291.79	0.0000			
11	0.4092	-0.1003	307.22	0.0000			
12	0.5893	0.2115	339.73	0.0000			
13	0.3711	0.0169	352.82	0.0000			
14	0.1775	0.0448	355.86	0.0000			
15	0.3044	-0.2721	364.95	0.0000			
16	0.4725	0.2261	387.22	0.0000			

```
.  
end of do-file
```

Tự Tương Quan (Autocorrelation) - Ví dụ 3 (3)



Tự Tương Quan (Autocorrelation) - Ví dụ 4 (1)



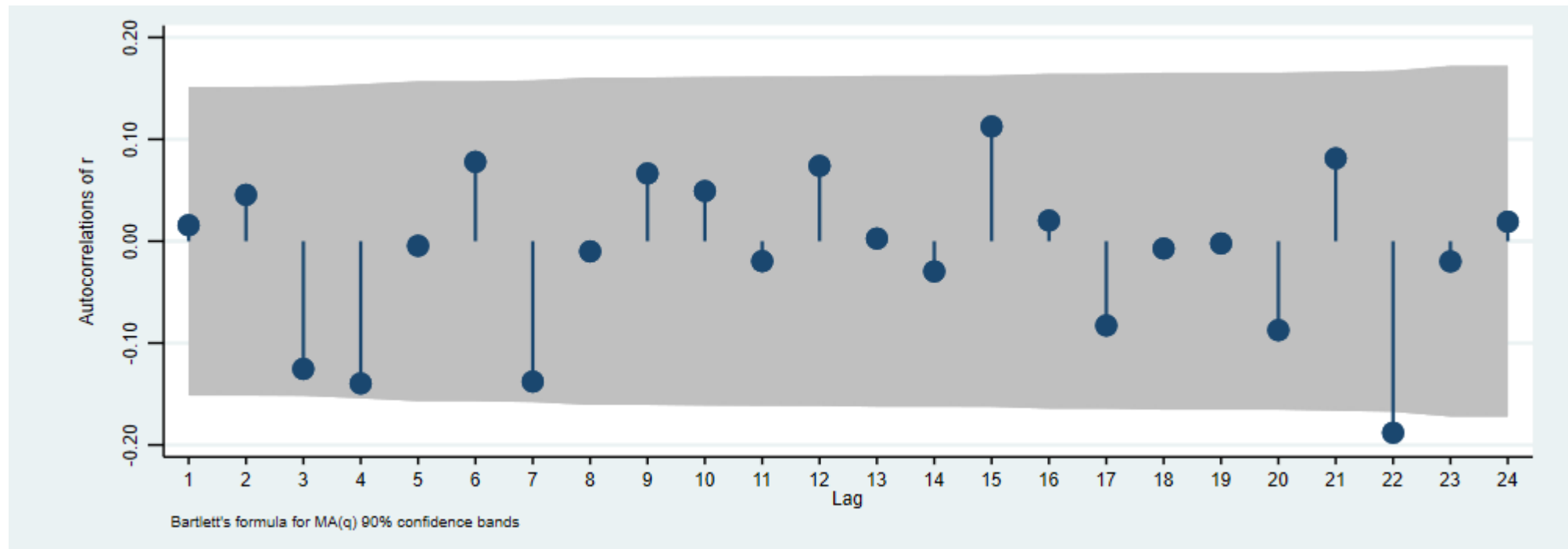
Tự Tương Quan (Autocorrelation) - Ví dụ 4 (2)

. corrgram r if tin(,2019m12)

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial autocor]
1	0.0157	0.0157	.0297	0.8632		
2	0.0455	0.0456	.28245	0.8683		
3	-0.1254	-0.1284	2.2176	0.5285		
4	-0.1397	-0.1388	4.6407	0.3262		
5	-0.0045	0.0126	4.6432	0.4609		
6	0.0779	0.0756	5.4099	0.4924		
7	-0.1378	-0.1843	7.833	0.3476		
8	-0.0100	-0.0333	7.8459	0.4487		
9	0.0664	0.1184	8.4192	0.4925		
10	0.0492	0.0300	8.7366	0.5573		
11	-0.0196	-0.0962	8.7877	0.6415		
12	0.0739	0.0943	9.5174	0.6582		
13	0.0026	0.0837	9.5183	0.7328		
14	-0.0296	-0.0895	9.6378	0.7882		
15	0.1126	0.1209	11.381	0.7251		
16	0.0203	0.0989	11.438	0.7817		
17	-0.0828	-0.1420	12.399	0.7754		
18	-0.0072	-0.0177	12.407	0.8256		
19	-0.0022	0.1348	12.407	0.8675		
20	-0.0873	-0.1173	13.51	0.8545		
21	0.0813	-0.0086	14.475	0.8483		
22	-0.1880	-0.1775	19.688	0.6026		
23	-0.0199	-0.0135	19.747	0.6571		
24	0.0191	-0.0302	19.802	0.7080		
25	0.1040	0.0684	21.448	0.6674		
26	-0.1451	-0.2545	24.689	0.5366		
27	-0.0659	-0.1898	25.365	0.5540		
28	0.0208	0.2299	25.433	0.6042		
29	0.0786	0.0612	26.416	0.6032		
30	0.0149	-0.2725	26.452	0.6519		
31	-0.0525	-0.0612	26.9	0.6772		
32	-0.2024	-0.0469	33.646	0.3877		

Command

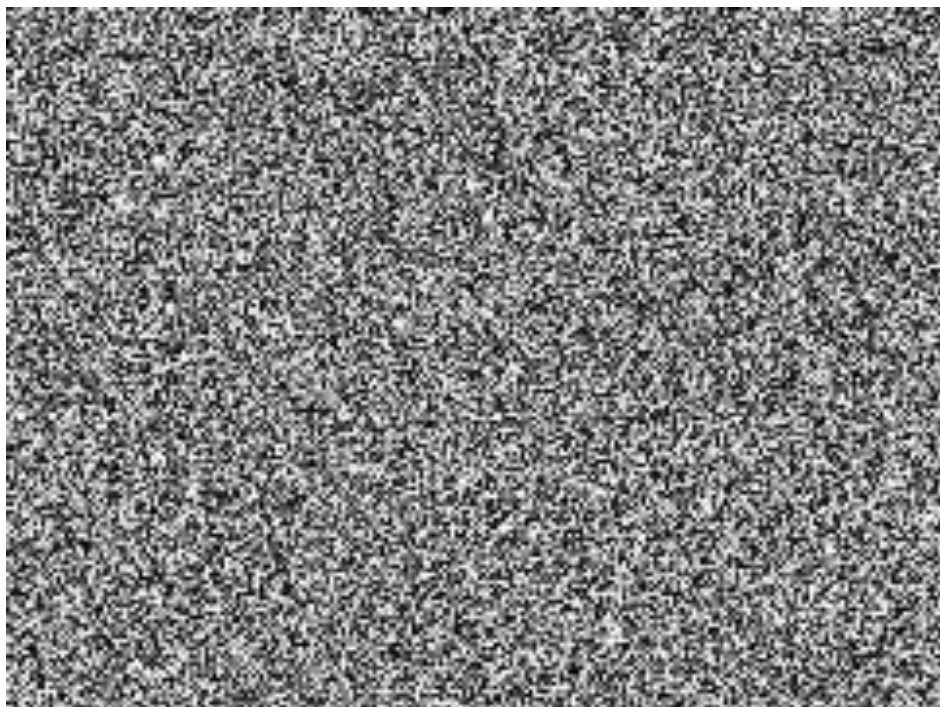
Tự Tương Quan (Autocorrelation) - Ví dụ 4 (3)



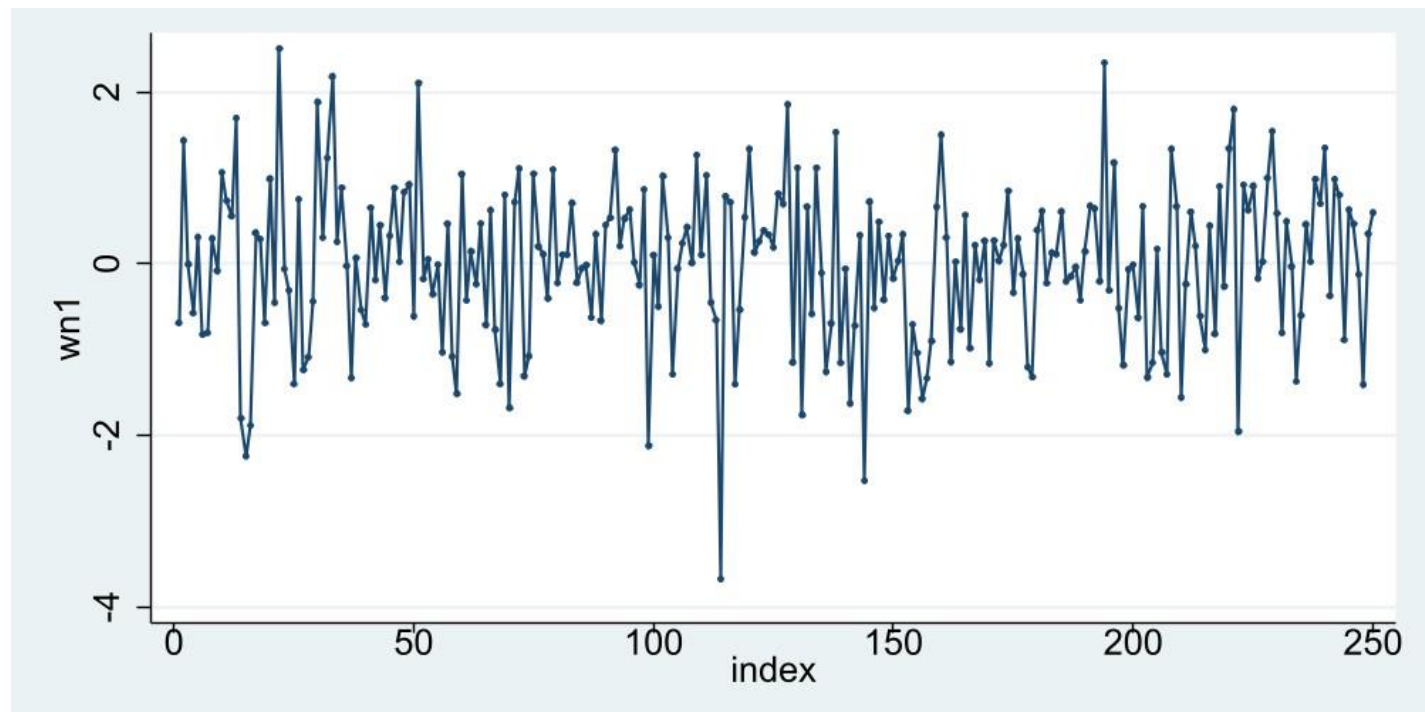
Nhiều Trắng (White Noise)

- Nhiều trắng (tạp âm trắng trong vật lý) là một dạng chuỗi thời gian mang tính ngẫu nhiên hoàn toàn.
- Xét về mặt xác suất thống kê: mỗi y_t ($t = 1, \dots, n$) có cùng một phân phối với cùng một giá trị trung bình, và phương sai cố định, cũng như không tự tương quan với các biến trễ của nó.
- Do tính chất ngẫu nhiên hoàn toàn, chúng ta không thể dự đoán nhiều trắng.
- Nhiều trắng chuẩn (Gaussian noise): nhiều trắng có phân phối chuẩn.

Nhiều Trắng:



Nguồn ảnh: Wikipedia



Nguồn: Dữ liệu simulation