MÔ HÌNH ARMA CHO CHUỖI THỜI GIAN DỪNG

TS. Nguyễn Mạnh Hùng

Đại học Giao thông Vận tải, 2021

School of Applied Mathematics

NỘI DUNG

- ❑ Quá trình ngẫu nhiên dừng
- ❑ Phân tích Wold
- ❑ Quá trình tự hồi quy AR(p)
- ❑ Quá trình trung bình trượt MA(q)
- ❑ Quá trình trung bình trượt tự hồi quy ARMA(p,q)
- ❑ Xây dựng mô hình ARMA và các ví dụ

Quá trình ngẫu nhiên dừng

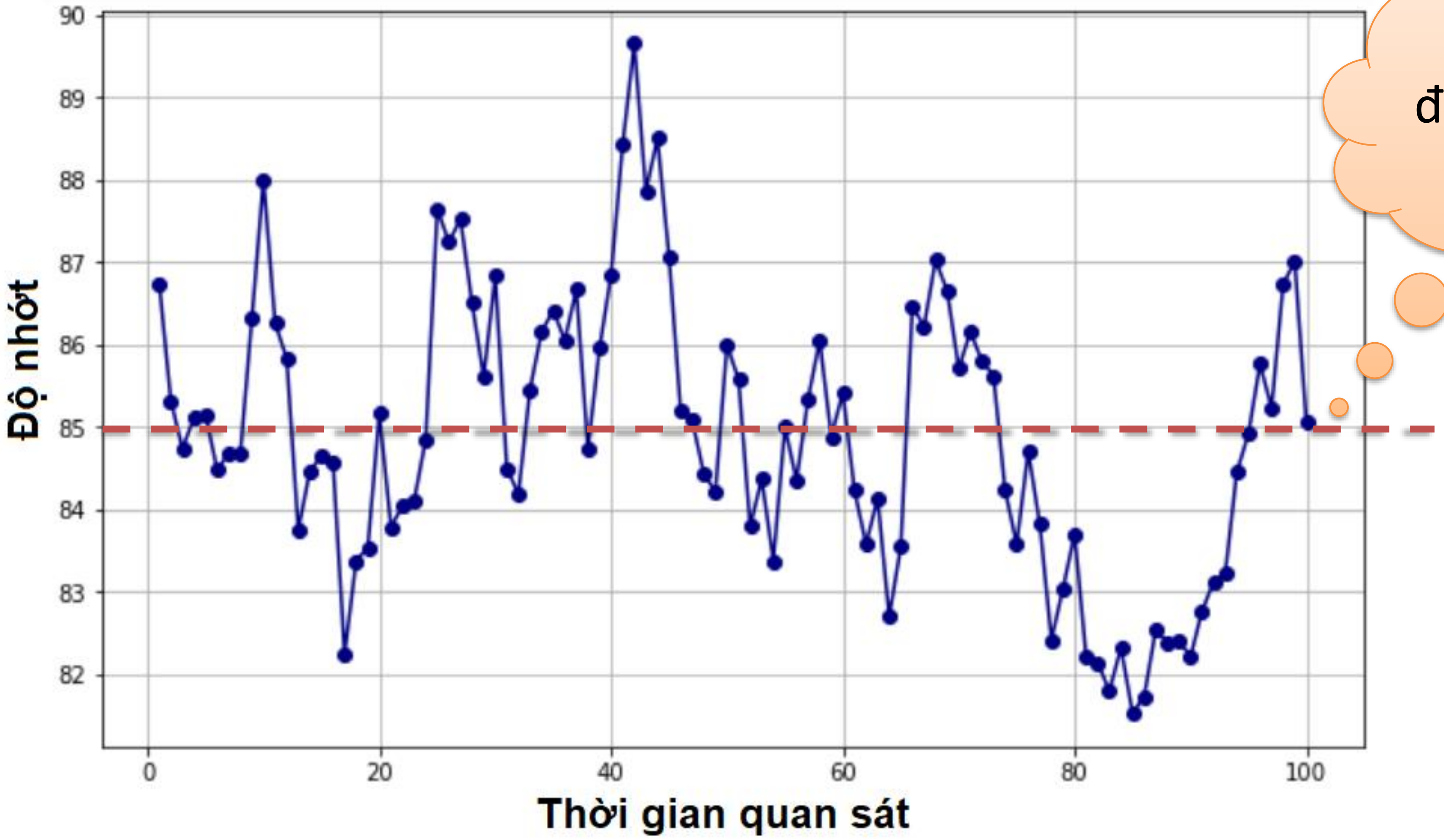
Quá trình ngẫu nhiên y_t được gọi là ***dừng yếu*** (***dừng hiệp phương sai***) nếu

$$E(y_t) = \mu; \quad Var(y_t) = \sigma_y^2$$
$$\gamma_k = Cov(y_t, y_{t-k}) = E[(y_t - \mu)(y_{t-k} - \mu)]$$
$$\rho_k = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t).Var(y_{t-k})}} = \frac{\gamma_k}{\gamma_0}$$

không phụ thuộc vào thời gian mà chỉ phụ thuộc vào độ trễ k.

School of Applied Mathematics

Độ nhớt của hoá chất trong quá trình sản xuất



Chuyển động quanh mức 85

Nhiễu trắng

Một chuỗi thời gian bao gồm các quan sát không tương quan, kỳ vọng 0 và phương sai hằng số được gọi là ***nhiễu trắng*** (***white noise***). Cụ thể, quá trình ngẫu nhiên a_t là nhiễu trắng nếu thoả mãn:

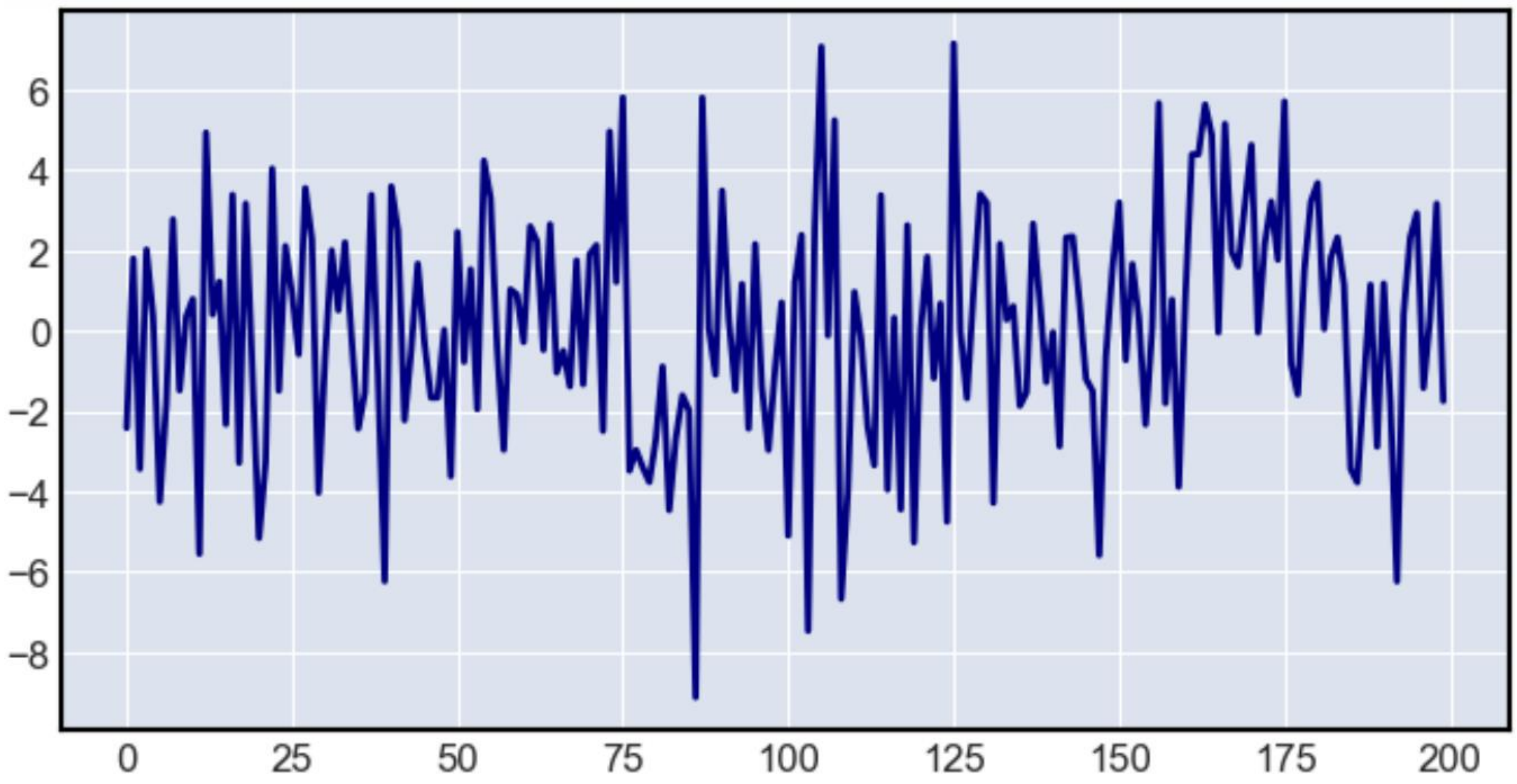
$$E(a_t) = 0;$$

$$Var(a_t) = E(a_t^2) = \sigma^2$$

$$Cov(a_t, a_\tau) = E[a_t a_\tau] = 0 \quad \text{với} \quad t \neq \tau$$

School of Applied Mathematics

Nhiễu trắng Gauss: $a_t \sim WN(0,9)$



Phân tích Wold

Mọi chuỗi thời gian dừng, thuần tuý ngẫu nhiên, $y_t - \mu$, có thể được viết dưới dạng tổ hợp tuyến tính của dãy biến ngẫu nhiên không tương quan:

$$y_t - \mu = a_t + \varphi_1 a_{t-1} + \varphi_2 a_{t-2} + \cdots$$

Dãy $a_t, t = 0, \pm 1, \pm 2, \ldots$ là các biến ngẫu nhiên không tương quan, cùng phân phối xác suất với:

$$E(a_t) = 0; \;\; Var(a_t) = E(a_t^2) = \sigma^2 < +\infty$$
$$Cov(a_t, a_{t-k}) = E(a_t a_{t-k}) = 0, \qquad k \neq 0$$

Một số tính chất *(tự chứng minh)*

(1) $E(y_t) = \mu$

(2) $\gamma_0 = \sigma^2 \sum_{j=0}^{\infty} \varphi_j^2$; $\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \varphi_j \varphi_{j+k}$

(3) $\rho_k = \dfrac{\gamma_k}{\gamma_0} = \dfrac{\sum_{j=0}^{\infty} \varphi_j \varphi_{j+k}}{\sum_{j=0}^{\infty} \varphi_j^2}$

Nếu số lượng các trọng số-φ là vô hạn thì các trọng số phải thoả mãn điều kiện khả tổng tuyệt đối, tức là:

$$\sum_{j=0}^{\infty} |\varphi_j| < +\infty$$

Quá trình tự hồi quy bậc nhất – AR(1)

- Không mất tính tổng quát, chọn $\mu = 0$.
- Quá trình AR(1) có dạng:

$$y_t - \phi y_{t-1} = a_t \quad \text{hay} \quad (1 - \phi B) y_t = a_t$$

với B là toán tử lùi $(B y_t = y_{t-1})$. Khi đó:

$$\begin{aligned} y_t &= (1 - \phi B)^{-1} a_t = (1 + \phi B + \phi^2 B^2 + \cdots) a_t \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots \end{aligned}$$

- Chuỗi này hội tụ khi $|\boldsymbol{\phi}| < \mathbf{1}$, và cũng là **điều kiện dừng** của quá trình y_t.

ACF của quá trình AR(1)

- Phương sai:

$$\gamma_0 = Var(y_t) = \frac{\sigma^2}{1-\phi^2}$$

- Tự hiệp phương sai:

$$\gamma_k = \phi^k \gamma_0$$

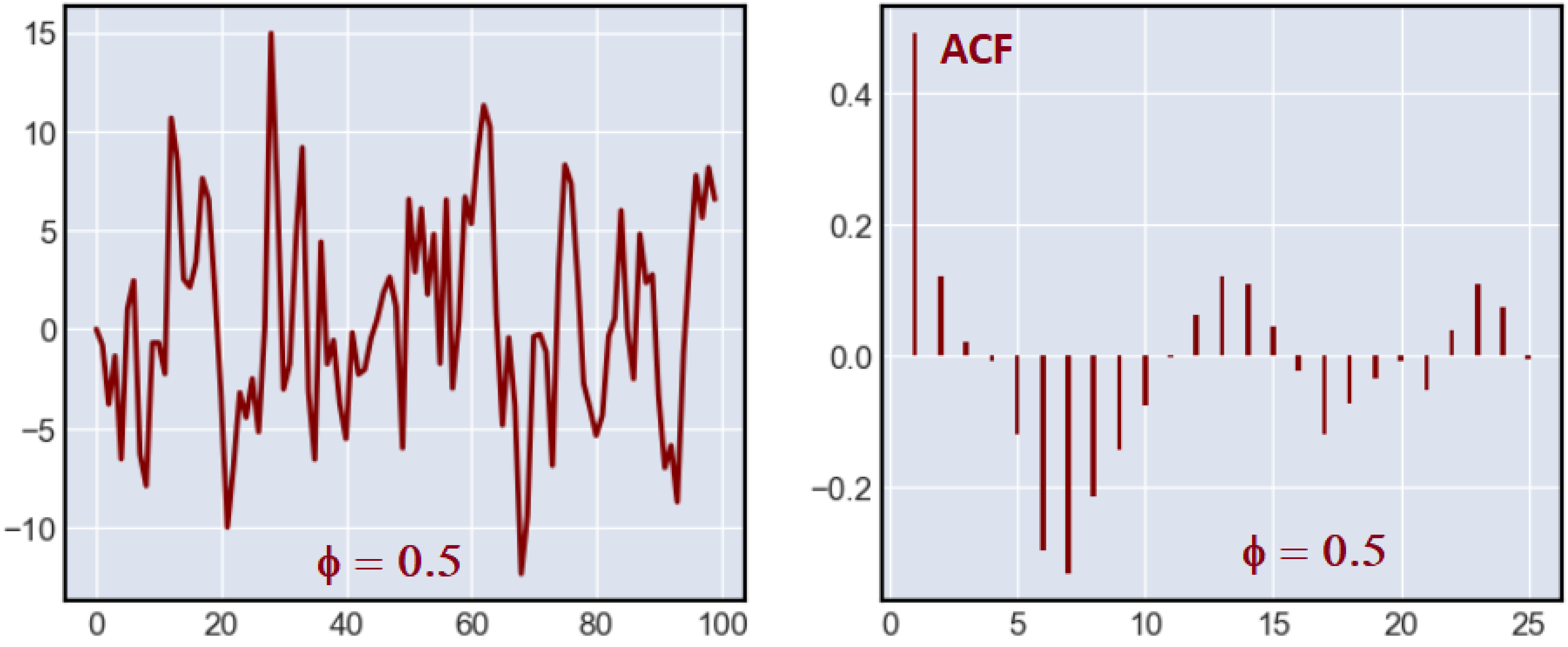
- Tự tương quan:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$$

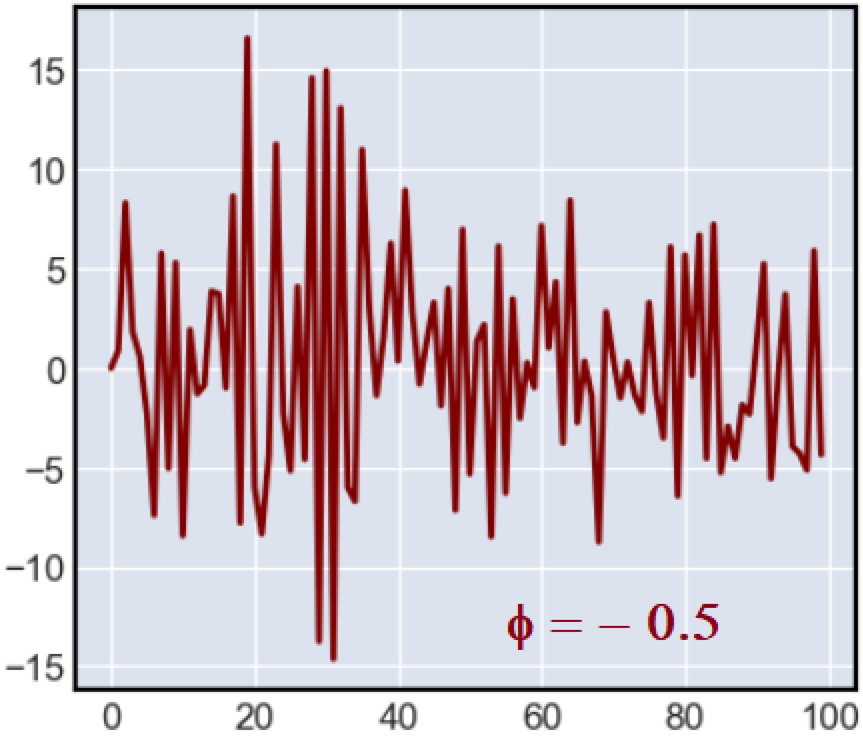
Do đó, nếu $\phi > 0$ thì hàm ACF giảm dần về 0 theo hàm mũ. Nếu $\phi < 0$ thì ACF giảm dần về 0 theo kiểu dao động tắt dần.

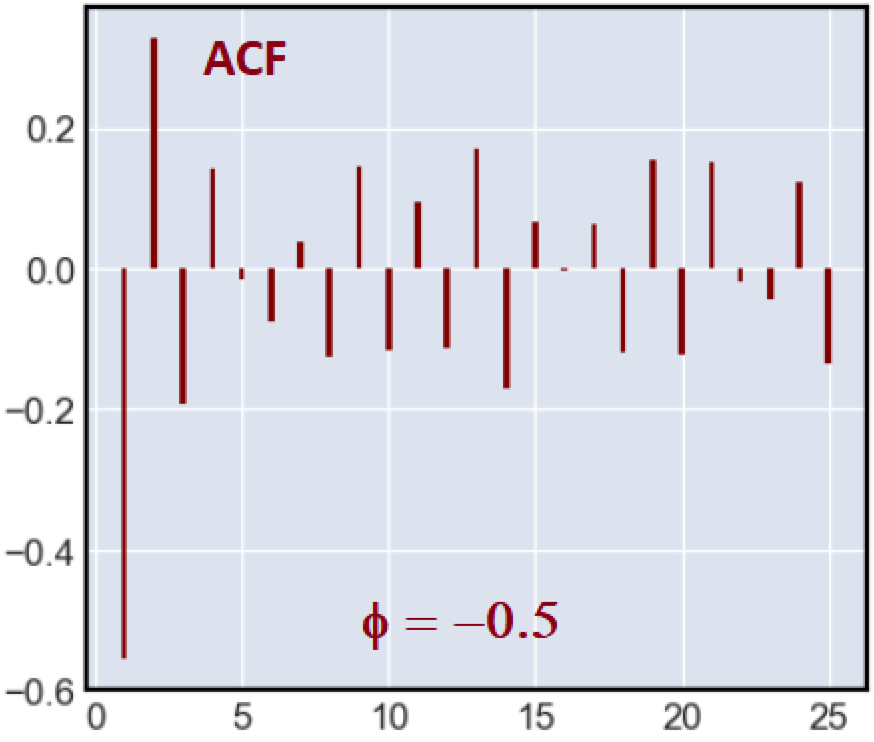
School of Applied Mathematics

Quá trình AR(1) với $\phi = 0.5$



Quá trình AR(1) với $\phi = -0.5$





Quá trình AR(p)

- Mô hình tự hồi quy bậc p, ký hiệu bởi AR(p), có dạng sau:

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = a_t$$

hay dưới dạng toán tử lùi B:

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right) y_t = \phi(B) y_t = a_t$$

- Điều kiện dừng của quá trình AR(p) là nghiệm của phương trình đặc trưng:

$$\phi(B) = (1 - g_1 B) \ldots \left(1 - g_p B\right) = 0$$

phải thoả mãn: $|g_i| < 1$ với $i = 1,2, \ldots, p$. Bởi vì y_t có dạng:

$$y_n = C_1 g_1^n + C_2 g_2^n + \cdots + C_p g_p^n$$

School of Applied Mathematics

Ví dụ

Xét quá trình AR(2):

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \phi(B) y_t = a_t$$

Phương trình đặc trưng:

$$g^2 - \phi_1 g - \phi_2 = 0$$

có 2 nghiệm:

$$g_1, g_2 = \frac{1}{2}\left(\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}\right)$$

có thể cùng là số thực hoặc số phức liên hợp. Điều kiện dừng $|g_1| < 1$ và $|g_2| < 1$ có thể quy về điều kiện với ϕ_1 và ϕ_2 sau:

$$\phi_1 + \phi_2 < 1 \; ; \quad -\phi_1 + \phi_2 < 1 \quad ; \; -1 < \phi_2 < 1$$

Thực hành: kiểm tra tính dừng của quá trình

$$y_t - 0.5 y_{t-1} - 0.3 y_{t-2} = a_t$$

ACF của quá trình AR(p)

- Tự hiệp phương sai:

$$\gamma_k = \sum_{i=1}^{p} \phi_i \gamma_{k-i} + \begin{cases} \sigma^2 & ,k = 0 \\ 0 & ,k > 0 \end{cases}$$

- Do đó, hàm ACF thoả mãn phương trình **Yule-Walker**:

$$\phi(B)\rho_k = 0$$

Nghiệm có dạng:

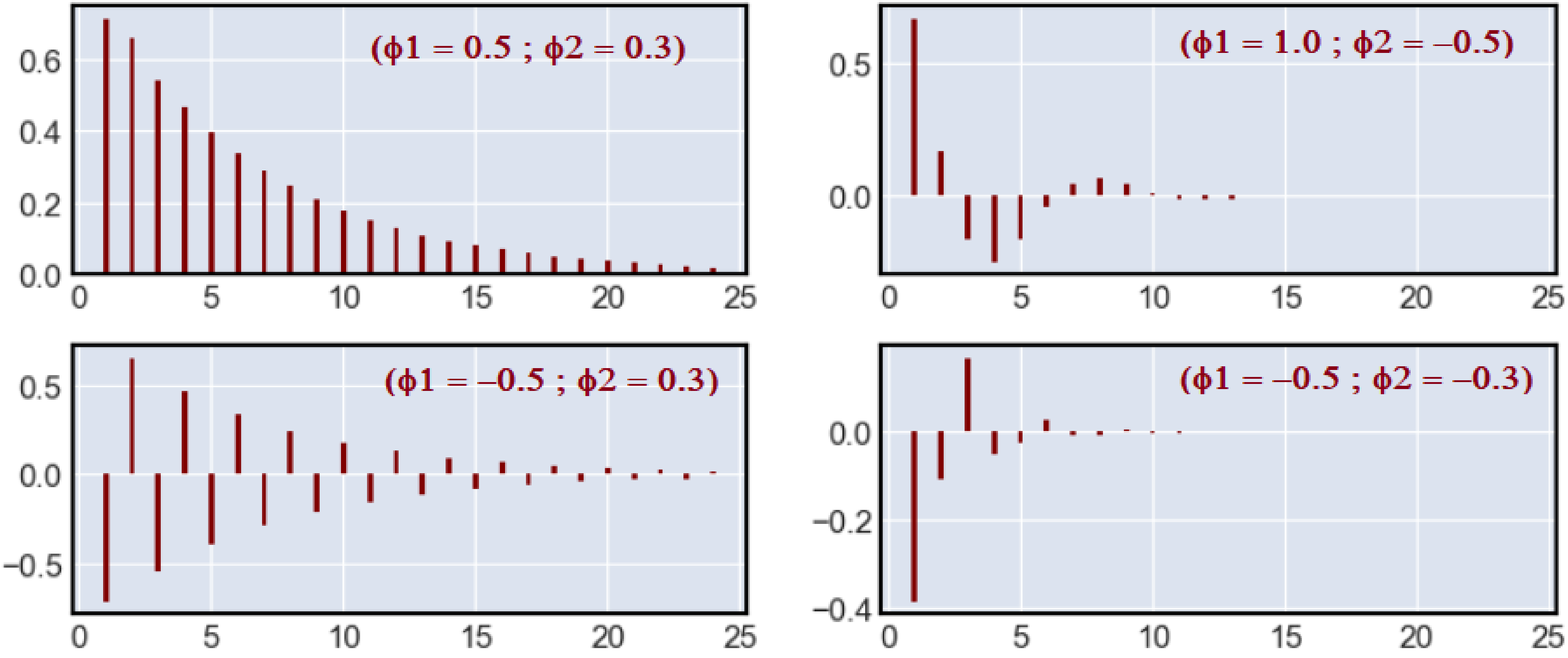
$$\rho_k = A_1 g_1^k + A_2 g_2^k + \cdots + A_p g_p^k$$

Vì $|g_i| < 1$ nên ACF giảm dần về 0 theo dạng hàm mũ (với g_i thực) kết hợp với dạng hàm sin (với g_i phức).

Thực hành: Hàm ACF của quá trình sau giảm dần về 0 theo dạng nào

$$y_t - 0.5y_{t-1} - 0.3y_{t-2} = a_t$$

ACF của một số quá trình AR(2)



School of Applied Mathematics

Hàm tự tương quan riêng phần (PACF)

- Hàm ACF của quá trình AR(p) giảm dần, tuy nhiên việc phân biệt bậc của AR(p) là khó khăn => sử dụng **hàm PACF**
- Hệ số tự tương quan riêng phần thứ k là hệ số ϕ_{kk} trong quá trình AR(k):

$$y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + a_t$$

- Phương trình Yule-Walker $\phi(B)\rho_m = 0$ $(m > 0)$, suy ra:

$$\rho_1 = \phi_{k1} + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1}$$
$$\rho_2 = \phi_{k1}\rho_1 + \phi_{k2} + \cdots + \phi_{kk}\rho_{k-2}$$
$$\ldots$$
$$\rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}$$

- Ký hiệu ma trận:

$$P_k = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{pmatrix}; \phi_k = \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \phi_{k3} \\ \vdots \\ \phi_{kk} \end{pmatrix}; \rho_k = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{pmatrix}$$

$$\Rightarrow \phi_k = P_k^{-1}\rho_k$$

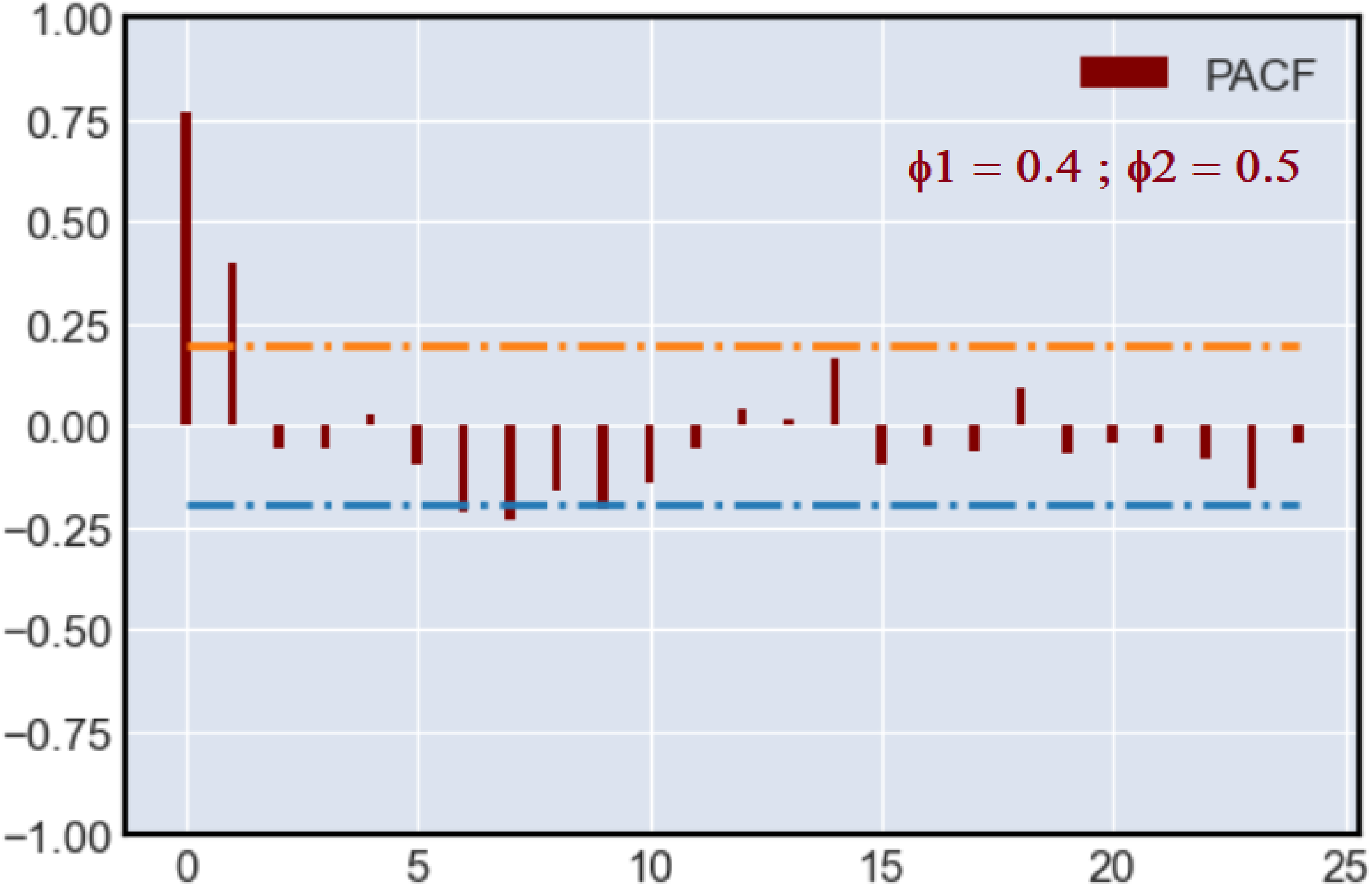
- Tính chất của hàm PACF của quá trình AR(p):

$$\phi_{11} \neq 0, \phi_{22} \neq 0, \dots, \phi_{pp} \neq 0\ ; \quad \phi_{kk} = 0 \text{ với } k > p$$

Ước lượng PACF

- Ước lượng $\hat{\phi}_{kk}$ được giải ra từ hệ phương trình:
$$P_k.\phi_k = \rho_k$$
trong đó ta thay $\rho_i \approx r_i$ (là hệ số tự tương quan mẫu).
- Với mẫu kích thước N quan sát về quá trình AR(p), hệ số $\hat{\phi}_{kk}\ (k > p)$ có phân phối xấp xỉ chuẩn với:
$$E(\hat{\phi}_{kk}) \approx 0 \text{ và } Var(\hat{\phi}_{kk}) \approx \frac{1}{N}$$
- Với độ tin cậy 95%, $E(\hat{\phi}_{kk})\ (k > p)$ nằm trong khoảng $\pm\frac{2}{\sqrt{N}}$. Ta dùng giới hạn này để xác định xem $\hat{\phi}_{kk}$ có khác biệt thống kê đáng kể với 0 hay không, từ đó xác định bậc của quá trình AR.

Quá trình AR(2): $y_t = 0.4y_{t-1} + 0.5y_{t-2} + a_t$



Quá trình trung bình trượt – MA(1)

- Quá trình MA(1) có dạng:

$$y_t = a_t - \theta a_{t-1} \quad \text{hay} \quad y_t = (1 - \theta B)a_t$$

- Nếu quá trình MA(1) là khả nghịch thì:

$$a_t = (1 - \theta B)^{-1} y_t = (1 + \theta B + \theta^2 B^2 + \cdots) y_t$$

Điều kiện khả nghịch của quá trình y_t là $|\boldsymbol{\theta}| < \mathbf{1}$.

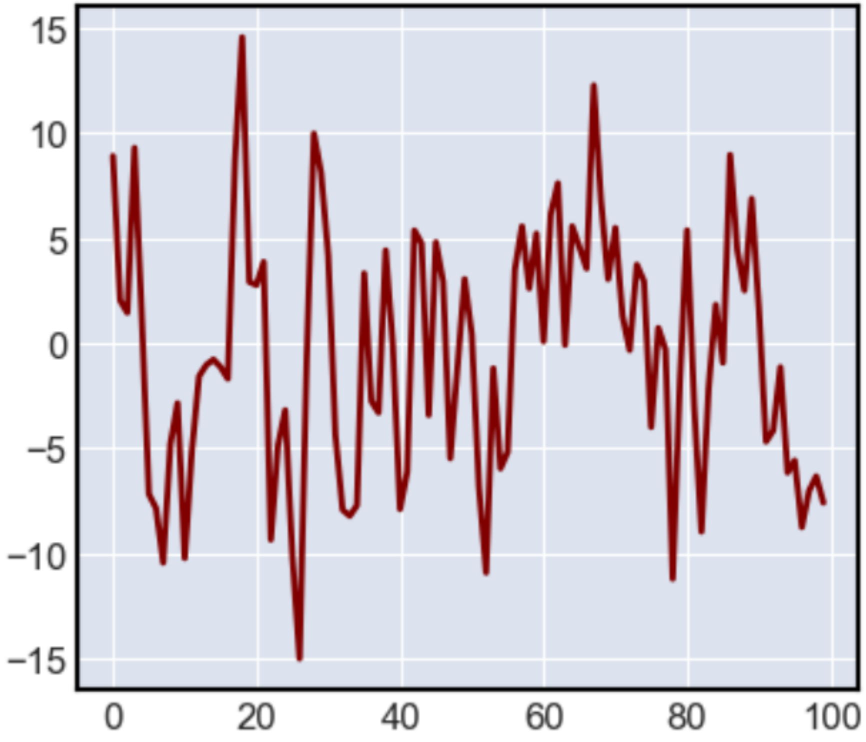
- Tự hiệp phương sai và tự tương quan:

$$\gamma_0 = \sigma^2(1 + \theta^2); \quad \gamma_1 = -\sigma^2\theta; \quad \gamma_k = 0 \quad (k > 1)$$

$$\rho_1 = -\frac{\theta}{1 + \theta^2} \; ; \quad \rho_k = 0 \quad (k > 1)$$

School of Applied Mathematics

Quá trình MA(1) với $\theta = -0.8$



Quá trình MA(q)

- Mô hình trung bình trượt bậc q, ký hiệu MA(q), có dạng:

$$y_t = a_t - \theta_1 a_{t-1} - \cdots \theta_q a_{t-q} = \theta(B) a_t$$

- Quá trình MA(q) là khả nghịch, tức là $a_t = [\theta(B)]^{-1} y_t$, nếu phương trình đặc trưng

$$\theta(B) = (1 - h_1 B) \ldots (1 - h_q B) = 0$$

có nghiệm thoả mãn $|h_i| < 1$ với $i = 1,2, \ldots, q$.

- Hàm tự tương quan ACF được tính bởi công thức:

$$\rho_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, \qquad k = 1,2, \ldots, q$$

$$\rho_k = 0\,, \qquad k > q$$

Thực hành: Kiểm tra tính khả nghịch của quá trình MA(2)

$$y_t = a_t - 0.7 a_{t-1} - 0.28 a_{t-2}$$

Tổng kết

- Hàm ACF của một quá trình AR(p) kéo dài đến vô hạn, nhưng hàm PACF bằng 0 sau độ trễ p.

- Trái lại hàm ACF của một quá trình MA(q) bằng 0 sau độ trễ q, nhưng hàm PACF lại kéo dài đến vô hạn.

Quá trình ARMA(p,q)

- Quá trình ARMA(p,q) là sự kết hợp của quá trình AR(p) và MA(q):
$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$
hay viết theo toán tử lùi:
$$\phi(B)y_t = \pi(\theta)a_t$$
- Điều kiện dừng của ARMA(p,q) là điều kiện dừng của AR(p) và điều kiện khả nghịch của ARMA(p,q) là điều kiện khả nghịch của MA(q).
- ACF của ARMA(p,q) biến thiên tương tự ACF của AR(p) sau $q-p$ giá trị đầu tiên $\rho_1, \ldots, \rho_{q-p}$, và PACF của nó biến thiên tương tự PACF của MA(q) khi $k > q-p$.

School of Applied Mathematics

Xây dựng mô hình ARMA và ước lượng

1. Ước lượng các số đặc trưng của quá trình như giá trị trung bình, phương sai, tự tương quan: $\bar{y}, s_y^2, r_k$ và vẽ biểu đồ tương quan ACF.

2. Nếu $\rho_k = 0$ với mọi k, thì r_k có phân phối xấp xỉ chuẩn với trung bình 0 và phương sai được tính bởi (Bartlett, 1946):

$$Var(r_k) = \frac{1}{T}\left(1 + 2\sum_{j=0}^{k-1} {r^*}_j^2\right)$$

$$r_j^* = \begin{cases} r_j & \rho_j \neq 0 \\ 0 & \rho_j = 0 \end{cases}$$

Xây dựng mô hình ARMA và ước lượng

3. Tính các hệ số PACF mẫu, $\hat{\phi}_{kk}$. Nếu dữ liệu được sinh bởi quá trình AR(p) thì $\hat{\phi}_{kk} \sim N\left(0, \frac{1}{T}\right)$ với $k > p$.
4. Xây dựng mô hình ARMA bằng p.p. Box-Jenkins (1970):
   – Giai đoạn 1: Nhận dạng quá trình ARMA bằng ACF, PACF
   – Giai đoạn 2: Ước lượng các tham số của mô hình (các trọng số ϕ, θ và các tham số μ, σ^2).
   – Giai đoạn 3: Kiểm tra phần dư

$$\hat{a}_t = y_t - \hat{\phi}_1 y_{t-1} - \cdots - \hat{\phi}_p y_{t-p} + \hat{\theta}_1 \hat{a}_{t-1} + \cdots + \hat{\theta}_q \hat{a}_{t-q}$$

có phải nhiễu trắng hay không? Có thể dùng thống kê:

$$Q(k) = T(T+2)\sum_{i=1}^{k} \frac{r_{a,i}^2}{T-i} \sim \chi^2(k-p-q)$$

Tiêu chuẩn lựa chọn mô hình ARMA(p,q)

- **Akaike's Information Criteria** (AIC) năm 1974:

$$AIC(p,q) = \log(\hat{\sigma}^2) + \frac{2}{T}(p+q)$$

- **Bayesian Information Criterion** (BIC) năm 1978:

$$BIC(p,q) = \log(\hat{\sigma}^2) + \frac{\log(T)}{T}(p+q)$$

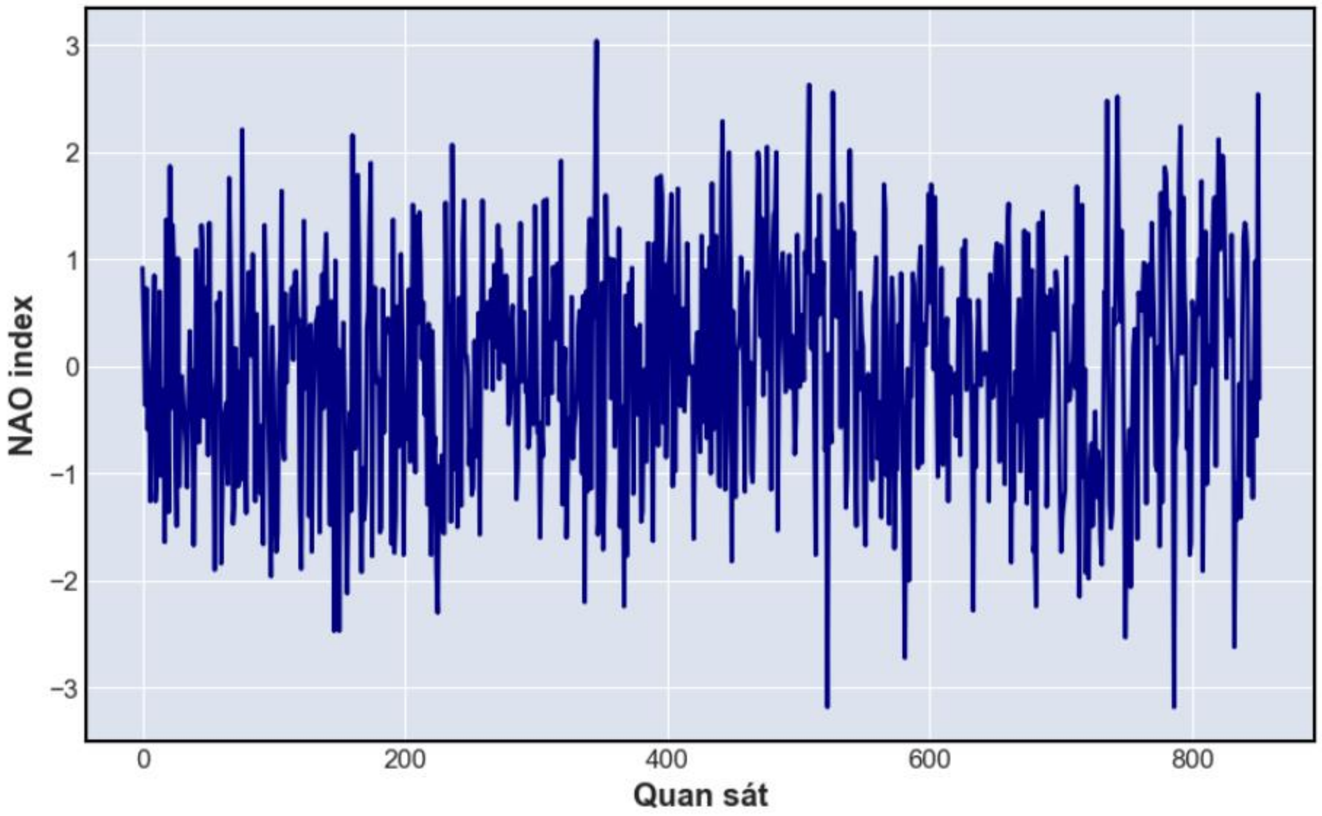
Các tiêu chuẩn này được dùng để chọn bậc p, q sao cho AIC, hoặc BIC có giá trị thấp nhất.

Thực hành: dữ liệu chỉ số NAO

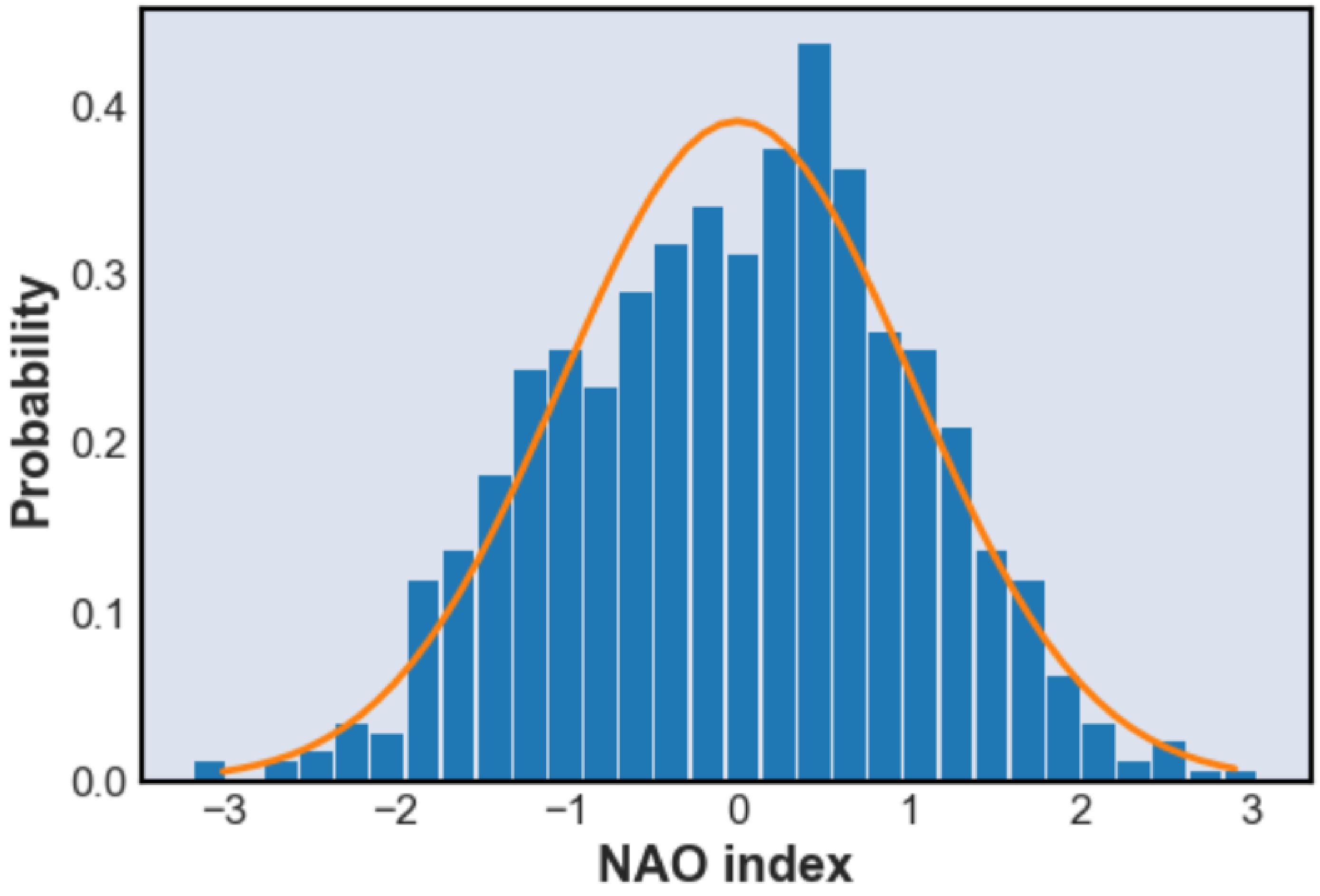
Chỉ số NAO (North Atlantic Oscillation), nguồn: https://www.noaa.gov/

có dao động khá gần với nhiễu trắng,

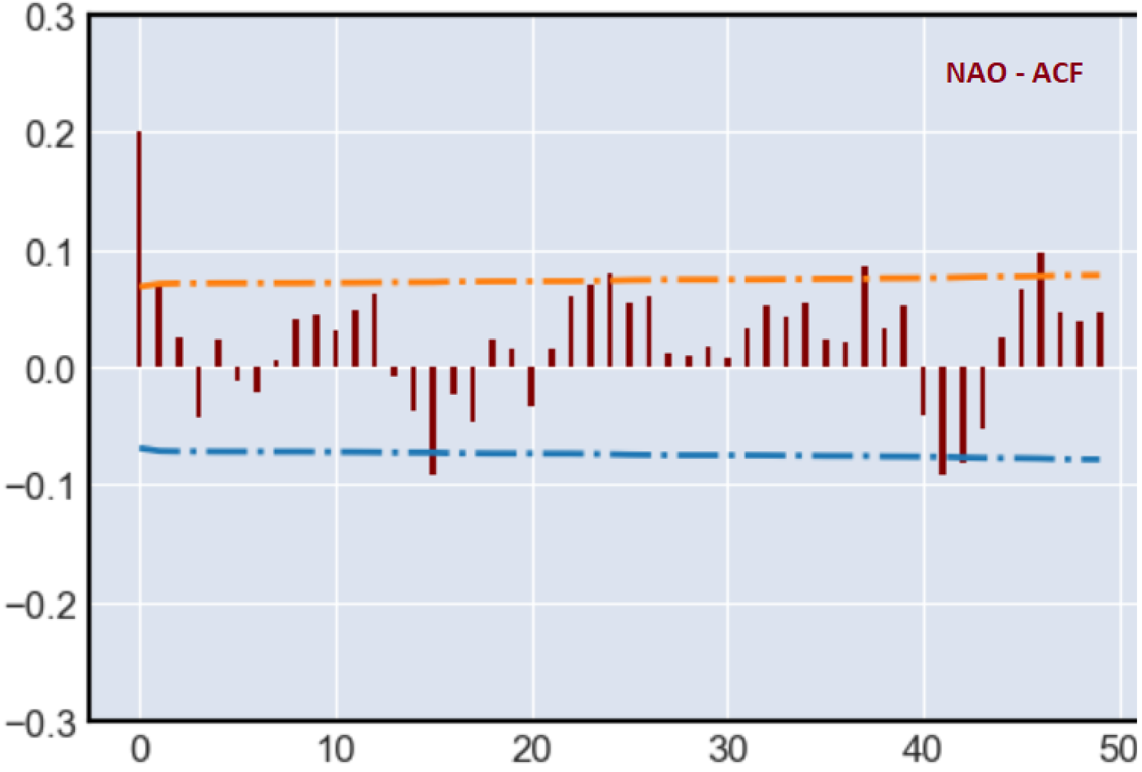
=> có thể có bậc tự tương quan thấp

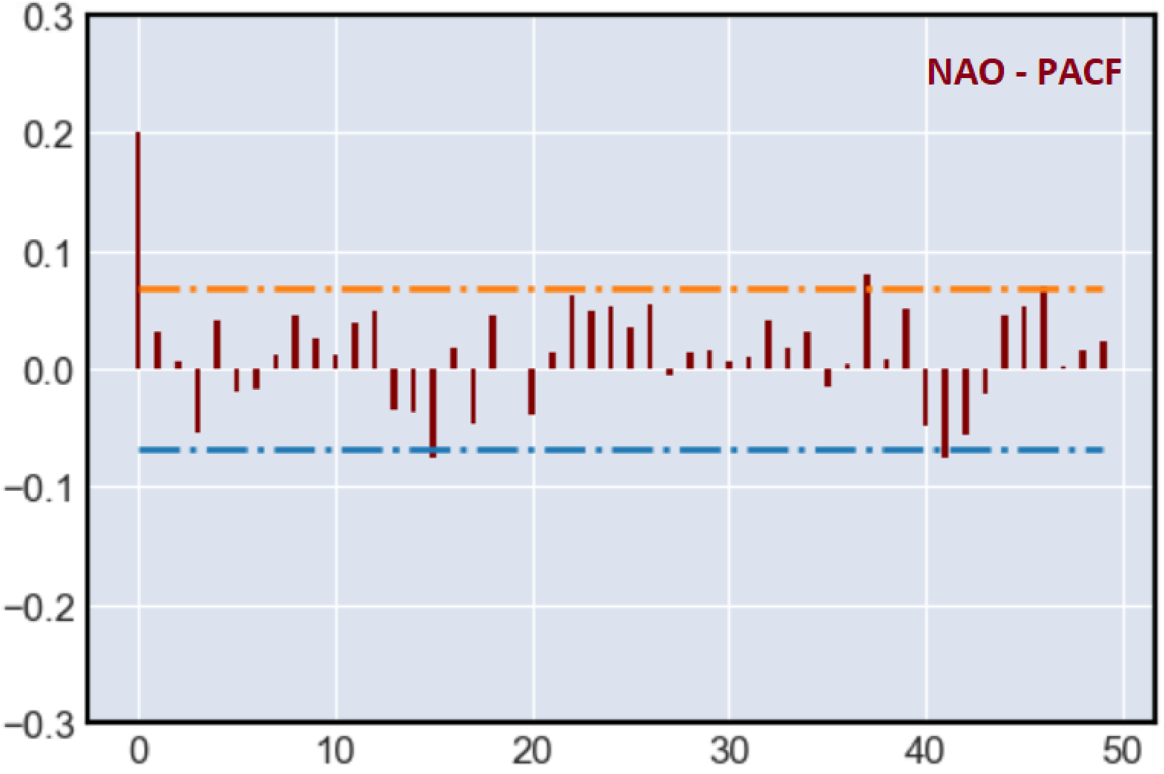


Histogram



ACF và PACF


NAO - ACF
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
10
20
30
40
50


NAO - PACF
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
10
20
30
40
50

Ước lượng và đánh giá mô hình

- Các giá trị của ACF và PACF khá gần 0 sau độ trễ $k = 1$. Do đó, có thể dùng mô hình AR(1) hoặc MA(1) để mô tả chỉ số NAO.
- Ước lượng mô hình AR(1):

$$y_t = 0.200195\, y_{t-1} + a_t$$

với $\hat{\sigma} = 1.001632$; $se(\phi) = 0.033590$; $Q(12) = 8.571582$ (p-value = 0.6614) ; $R^2 = 0.04$

- Ước lượng mô hình MA(1):

$$y_t = \hat{a}_t + 0.193787\, \hat{a}_{t-1}$$

với $\hat{\sigma} = 1.004262; se(\theta) = 0.033679; Q(12) = 12.065169$ (p-value = 0.3588) ; $R^2 = 0.036$

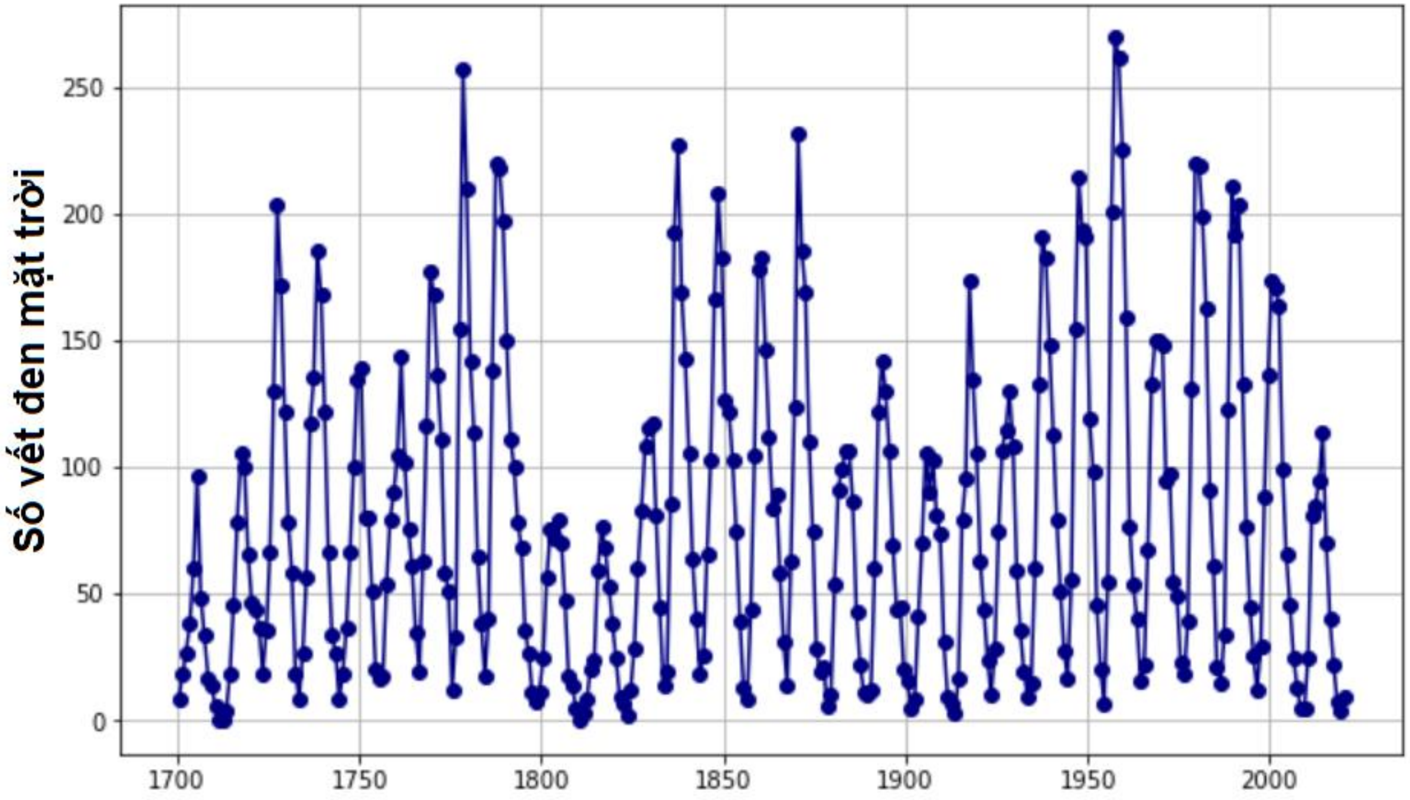
Thực hành: dữ liệu số vết đen mặt trời

Chuỗi có chuyển động với chu kỳ khoảng 11 năm.

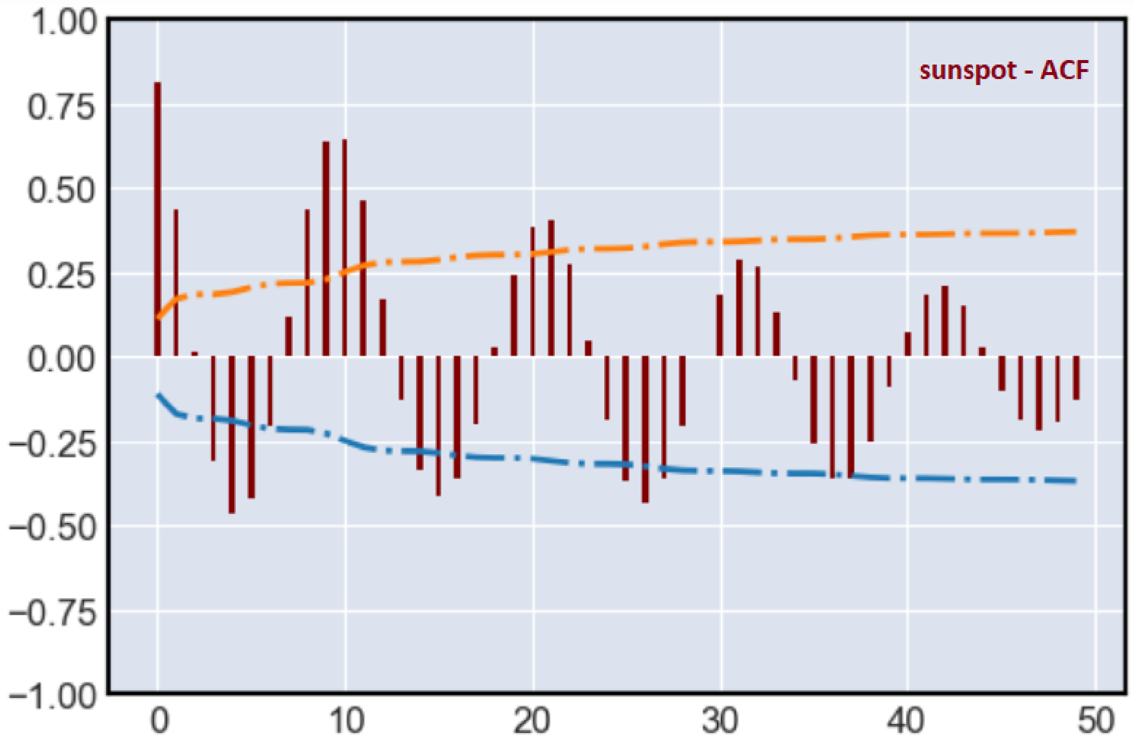
$$n = 321$$

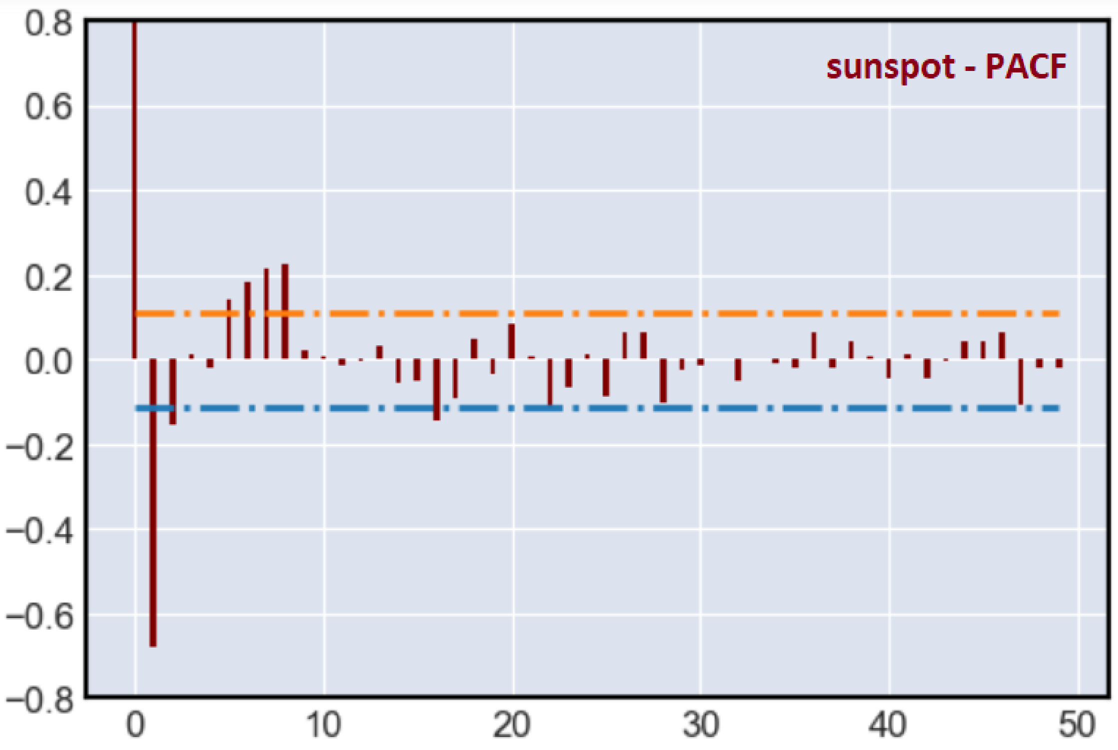
$$\bar{y} = 78.518069$$

$$s_y = 61.993548$$



ACF và PACF





School of Applied Mathematics

AIC và BIC

- Bảng bên thể hiện các tiêu chuẩn AIC và BIC được tính cho mô hình AR đến bậc 20.
- Giá trị nhỏ nhất đạt được với $k = 9$

k	AIC	BIC
1	7.155748	7.167524
2	6.514755	6.538361
3	6.505185	6.540676
4	6.516872	6.564303
5	6.529642	6.589068
6	6.521701	6.593179
7	6.469803	6.553388
8	6.432203	6.527953
9	6.393805	6.501776
10	6.406524	6.526774

k	AIC	BIC
11	6.418680	6.551268
12	6.430566	6.575550
13	6.442517	6.599956
14	6.451650	6.621604
15	6.460571	6.643100
16	6.468929	6.664093
17	6.455640	6.663501
18	6.452950	6.673568
19	6.464154	6.697591
20	6.475383	6.721703

School of Applied Mathematics

Ước lượng mô hình AR(9)

$\hat{\sigma} = 23.761$

$Q(20) = 17.1634$

(p-value = 0.1031)

$R^2 = 0.8594$

=> Các tham số từ $\hat{\theta}_3$ đến $\hat{\theta}_8$ không có sự khác biệt thống kê với 0.

	param	se(param)
0	12.225337	4.125892
1	1.170218	0.056108
2	-0.419324	0.087584
3	-0.132726	0.090388
4	0.102342	0.089935
5	-0.068406	0.089848
6	0.002390	0.089828
7	0.021711	0.089866
8	-0.050242	0.086892
9	0.221874	0.055706

Mô hình AR(9) rút gọn

$\hat{\sigma} = 23.696$;

$Q(20) = 18.1088$

(p-value = 0.3176)

$R^2 = 0.8574$

	param	se(param)
0	8.600032	3.101629
1	1.223042	0.043502
2	-0.522565	0.044239
9	0.192725	0.025591

School of Applied Mathematics

Bài tập

Bài 1: Giả sử một quá trình ngẫu nhiên dừng y_t có biểu diễn Wold như sau:

$$y_t - \mu = a_t + \varphi_1 a_{t-1} + \varphi_2 a_{t-2} + \cdots$$

ở đó $a_t \sim WN(0, \sigma^2)$. Tính kỳ vọng, tự hiệp phương sai và tự tương quan của y_t.

Bài 2: Xét quá trình AR(2):

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = a_t$$

Chứng minh rằng điều kiện dừng cho quá trình y_t có thể quy về điều kiện với ϕ_1 và ϕ_2 như sau:

$$\phi_1 + \phi_2 < 1 \;\; ; \quad -\phi_1 + \phi_2 < 1 \;\; ; \;\; -1 < \phi_2 < 1$$

Bài tập

Bài 3: Xét quá trình AR(p):

$$y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} = a_t$$

Chứng minh rằng hàm ACF của y_t thoả mãn phương trình Yule-Walker:

$$\phi(B)\rho_k = 0$$

Bài 4: Kiểm tra tính dừng của các quá trình AR sau đây:

a) $y_t = 0.95y_{t-1} + a_t$

b) $y_t = 0.8y_{t-1} - 0.6y_{t-2} + a_t$

c) $y_t = 0.4y_{t-1} - 0.5y_{t-2} + 0.2y_{t-3} + a_t$

Bài tập

Bài 5: Hãy tính hàm ACF của quá trình trung bình trượt MA(q)

$$y_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

Bài 6: Cho quá trình trung bình trượt MA(2):

$$y_t = a_t - 0.8a_{t-1} + 0.3a_{t-2}$$

a) Kiểm tra điều kiện khả nghịch của quá trình.

b) Mô phỏng và vẽ hàm ACF mẫu. Từ đó nhận xét về bậc của quá trình MA.

School of Applied Mathematics

School of Applied Mathematics

Bài tập

Bài 7: Xây dựng mô hình ARMA cho dữ liệu về doanh số bán thuốc '*N05C*' trong file '*salesmonthly.csv*'.

Bài 8: Xây dựng mô hình ARMA cho dữ liệu về độ nhớt của hoá chất trong quá trình sản xuất trong file '*viscosity.csv*'.