



MÔ HÌNH ARIMA CHO CHUỖI THỜI GIAN KHÔNG DỪNG

TS. Nguyễn Mạnh Hùng

Đại học Giao thông Vận tải, 2021



NỘI DUNG

- ☐ Quá trình ngẫu nhiên không dừng
- ☐ Bước ngẫu nhiên
- ☐ Quá trình ARIMA và một số ví dụ xây dựng mô hình
- ☐ Kiểm định nghiệm đơn vị
- ☐ Giới thiệu về dự báo
- ☐ Các quá trình có thuộc tính mùa



1. Quá trình ngẫu nhiên không dừng

- Thực tế, nhiều chuỗi thời gian là không dừng, do có giá trị trung bình hay phương sai thay đổi theo thời gian. Để xử lý tính không dừng kiểu này, ta phân tích:

$$y_t = \mu_t + \varepsilon_t$$

trong đó, μ_t : là mức trung bình thay đổi theo thời gian và ε_t là sai số ngẫu nhiên.

- Có 2 cách để mô hình hoá mức trung bình không dừng:
 - μ_t là một đa thức của thời gian, với ε_t được giả thiết là một quá trình dừng với kỳ vọng 0.
 - Sử dụng mô hình ARMA có tham số tự hồi quy không thoả mãn điều kiện dừng.



Mô hình hoá mức trung bình không dừng

- μ_t là một đa thức của thời gian:

$$y_t = \mu_t + \varepsilon_t = \sum_{j=0}^d \beta_j t^j + \psi(B)a_t$$
$$\Rightarrow E(y_t) = \sum_{j=0}^d \beta_j t^j$$

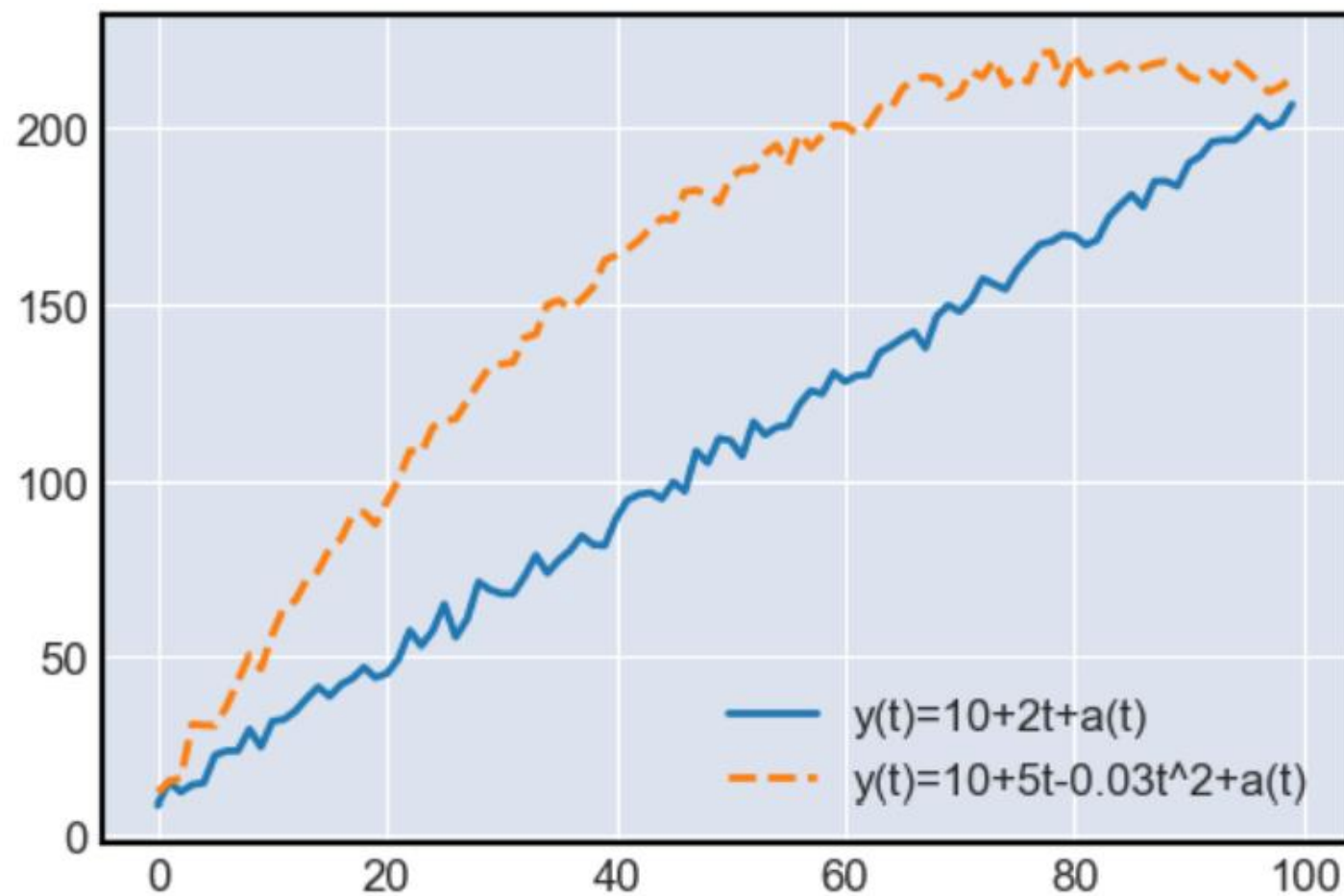
Do các hệ số β_j không đổi theo thời gian nên thành phần xu thế $E(y_t)$ được gọi là **tất định**.

- Xu thế này có thể loại bỏ bằng cách lấy sai phân:

$$\nabla^d y_t = d! \beta_d + \nabla^d \psi(B)a_t$$



Mô phỏng xu thế bất định





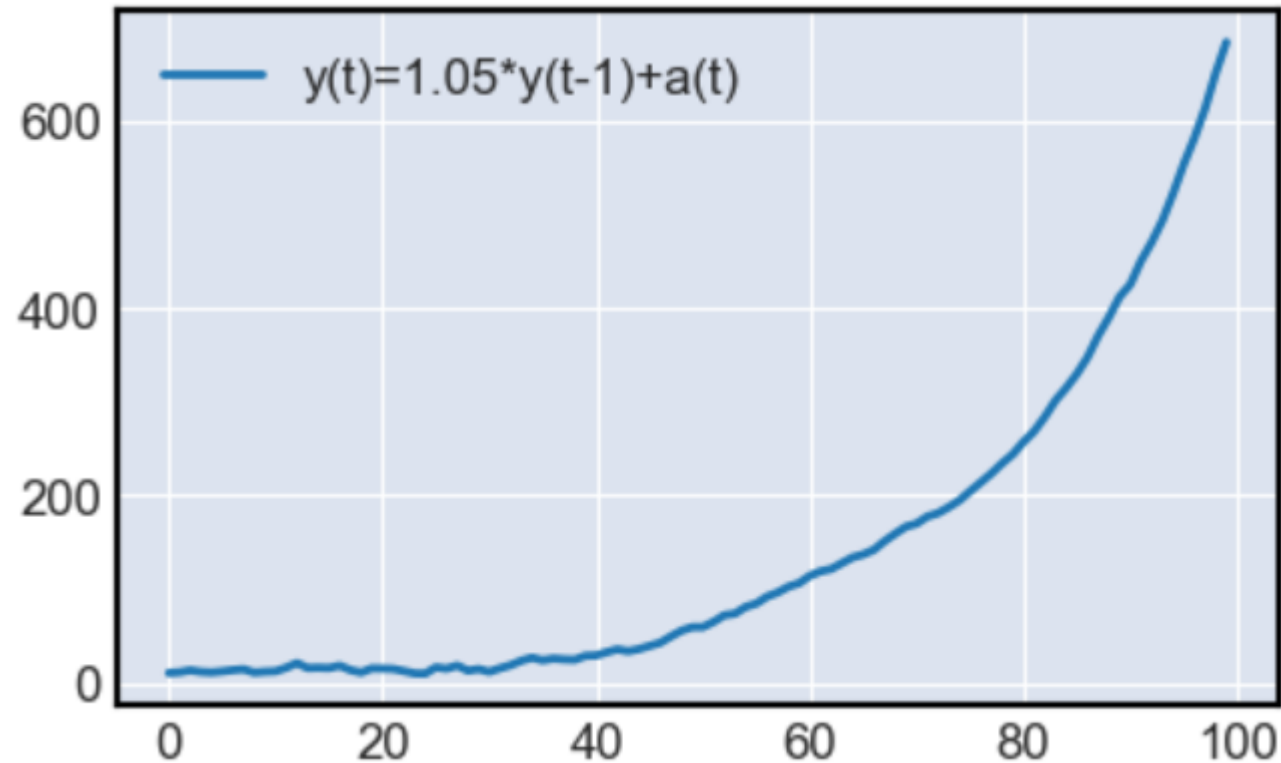
Mô hình hoá mức trung bình không dừng

- Sử dụng mô hình ARMA có tham số tự hồi quy không thoả mãn điều kiện dừng.
- Quá trình AR(1): $y_t = \phi y_{t-1} + a_t$, $\phi > 1$. Nếu bắt đầu tại thời điểm $t = 0$, phương trình sai phân có nghiệm:

$$y_t = y_0 \phi^t + \sum_{i=0}^t \phi^i a_{t-i}$$

Hàm xu thế của y_t tại thời điểm $t = 0$ là $y_0 \phi^t$. Hàm xu thế tại các thời điểm tiếp theo phụ thuộc vào dãy biến a_1, a_2, \dots . Do đó, xu thế thay đổi một cách **ngẫu nhiên**.

Mô phỏng xu thế ngẫu nhiên





2. Bước ngẫu nhiên

- Đề xuất bởi Karl Pearson và Lord Rayleigh trên Nature (1905).
- Mô tả "chiến lược tìm kiếm" tối ưu để tìm một người say rượu bị bỏ lại giữa cánh đồng trong đêm.
- Giới thiệu bởi Louis Bachelier trong luận án tiến sĩ "Theorie de Speculation" (1900) để mô tả diễn biến khó lường của giá cổ phiếu.
- Chuỗi thời gian y_t được gọi là bước ngẫu nhiên nếu:

$$y_t = y_{t-1} + a_t$$

Nếu có thêm hằng số θ_0 trong mô hình, y_t được gọi là bước ngẫu nhiên có dịch chuyển (random walk with drift):

$$y_t = y_{t-1} + \theta_0 + a_t$$



Bước ngẫu nhiên

- Nếu quá trình bắt đầu tại $t = 0$, thì y_t có dạng:

$$y_t = y_0 + t\theta_0 + \sum_{i=0}^t a_{t-i}$$

- Kỳ vọng, phương sai, hiệp phương sai:

$$E(y_t) = y_0 + t\theta_0 ; \gamma_{0,t} = \text{Var}(y_t) = t\sigma^2$$

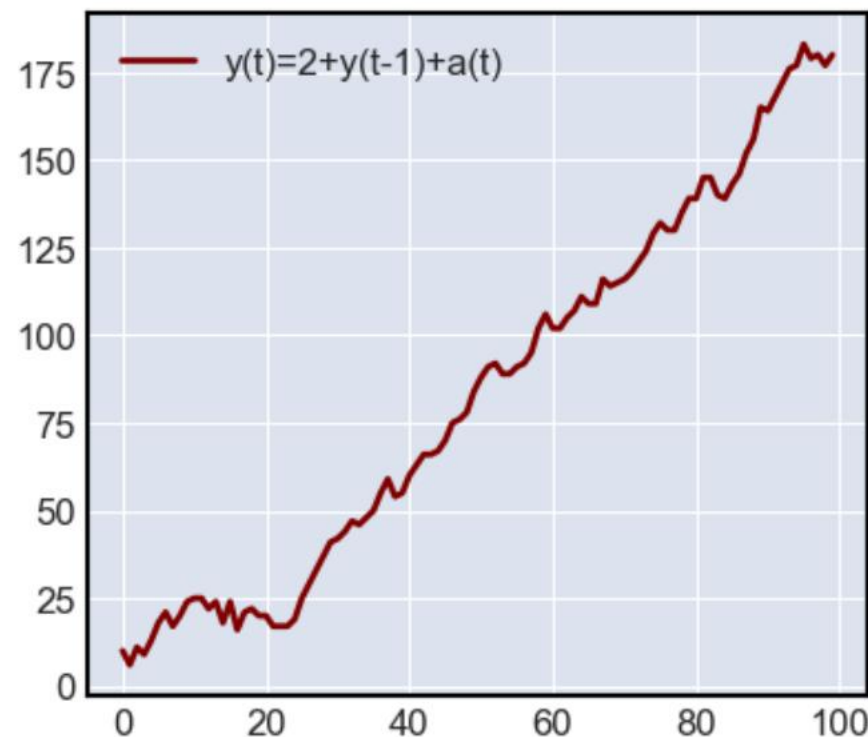
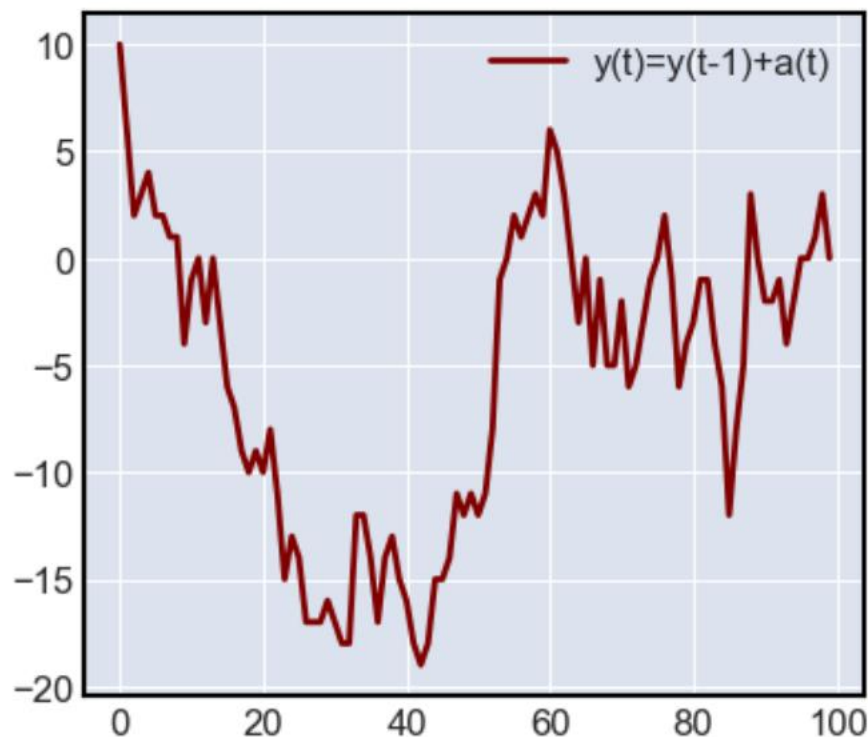
$$\gamma_{k,t} = \text{Cov}(y_t, y_{t-k}) = (t-k)\sigma^2, \quad k \geq 0$$

- Hệ số tự tương quan:

$$\rho_{k,t} = \frac{\gamma_{k,t}}{\sqrt{\gamma_{0,t} \cdot \gamma_{0,t-k}}} = \frac{t-k}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}} \xrightarrow{(t \rightarrow \infty)} 1$$



Mô phỏng bước ngẫu nhiên





3. Quá trình ARIMA

- Bước ngẫu nhiên là một ví dụ của lớp các quá trình không dừng, được gọi là các quá trình kết hợp:

$$y_t = y_{t-1} + \theta_0 + a_t \Leftrightarrow \nabla y_t = \theta_0 + a_t$$

- Giả sử có một chuỗi y_t cần lấy sai phân đến bậc d để thu được $w_t = \nabla^d y_t$ là một quá trình dừng tự tương quan. Hơn nữa, nếu $w_t \sim ARMA(p, q)$ thì mô hình cho chuỗi ban đầu có dạng:

$$\phi(B) \nabla^d y_t = \theta_0 + \theta(B) a_t$$

Khi đó, y_t được gọi là quá trình tự hồi quy kết hợp trung bình trượt bậc p, d , và q , viết ngắn gọn là **ARIMA**(p, d, q).

- Chuỗi y_t được gọi là kết hợp bậc d , ký hiệu $y_t \sim I(d)$.



Tính chất

- 1) Các hệ số tự tương quan của quá trình ARIMA sẽ gần 1 khi độ trễ k không quá lớn.

Ví dụ: xét quá trình ARMA(1,1)

$$y_t = \phi y_{t-1} + a_t - \theta a_{t-1}$$

Hàm ACF có dạng:

$$\rho_1 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}, \quad \rho_k = \phi\rho_{k-1} \quad (k > 1)$$

Khi $\phi \rightarrow 1$, ta thu được quá trình ARIMA(0,1,1):

$$\nabla y_t = a_t - \theta a_{t-1}$$

và mọi ρ_k hội tụ về 1.



Tính chất

2) Xét mô hình ARIMA với $\theta_0 = 0$:

$$\phi(B)\nabla^d y_t = \theta(B)a_t \Leftrightarrow \begin{cases} \phi(B)w_t = \theta(B)a_t \\ w_t = \nabla^d y_t \end{cases}$$

ở đó w_t là quá trình ARMA dừng và khả nghịch.

Với $d \geq 1$, ta có thể viết:

$$y_t = S^d w_t$$

với S là toán tử tổng vô hạn: $S = 1 + B + B^2 + \dots$

Điều đó có nghĩa là y_t thu được bằng cách kết hợp, d lần chuỗi dừng w_t , nên ta có thuật ngữ “**quá trình kết hợp**”.



Tính chất

Quá trình ARIMA là **không dừng đồng nhất**, tức là ứng xử cục bộ của chuỗi y_t độc lập với mức của y_t . Cụ thể:

$$\phi(B)\nabla^d(y_t + c) = \phi(B)\nabla^d y_t$$

3) Xét mô hình ARIMA với $\theta_0 \neq 0$:

$$\phi(B)w_t = \theta_0 + \theta(B)a_t$$

với

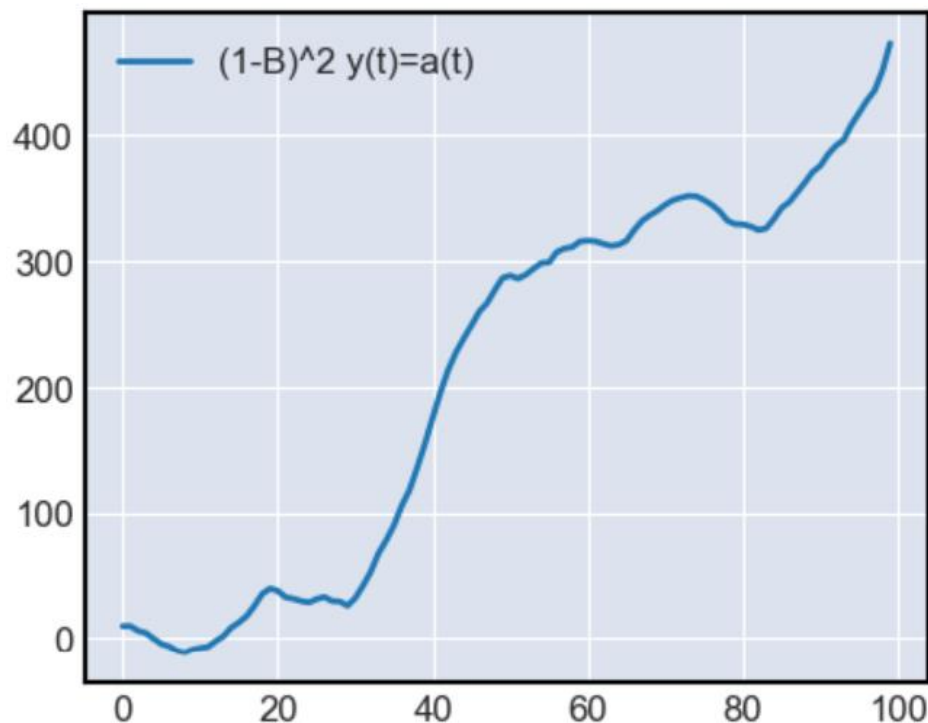
$$w_t = \nabla^d y_t$$

Khi đó,

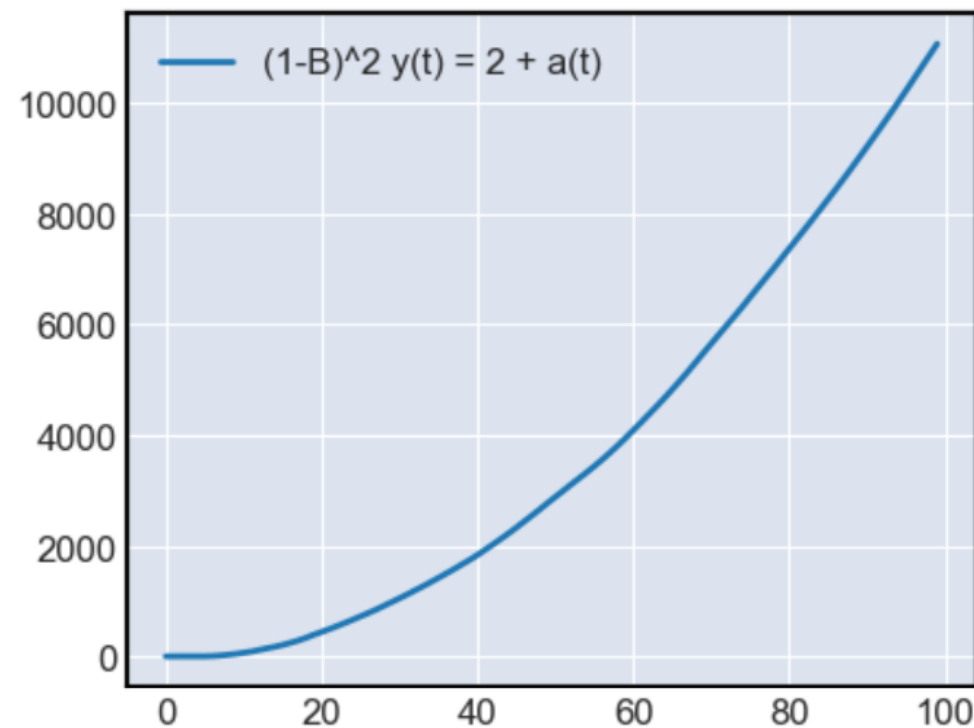
$$E(w_t) = \mu_w = \frac{\theta_0}{1 - \phi_1 - \dots - \phi_p} \neq 0$$



Mô phỏng: $\nabla^2 y_t = a_t$, với $y_0 = y_1 = 10$, $a_t \sim WN(0,9)$ cho thấy sự dịch chuyển **ngẫu nhiên** cả về **mức** và **độ dốc**.



Mô phỏng: $\nabla^2 y_t = 2 + a_t$, với $y_0 = y_1 = 10$ và $a_t \sim WN(0,9)$ cho thấy hàm **xu thế tất định** bậc 2 triệt tiêu ảnh hưởng của nhiễu.





Nhận xét

Mô hình ARIMA:

$$\phi(B)\nabla^d y_t = \theta_0 + \theta(B)a_t$$

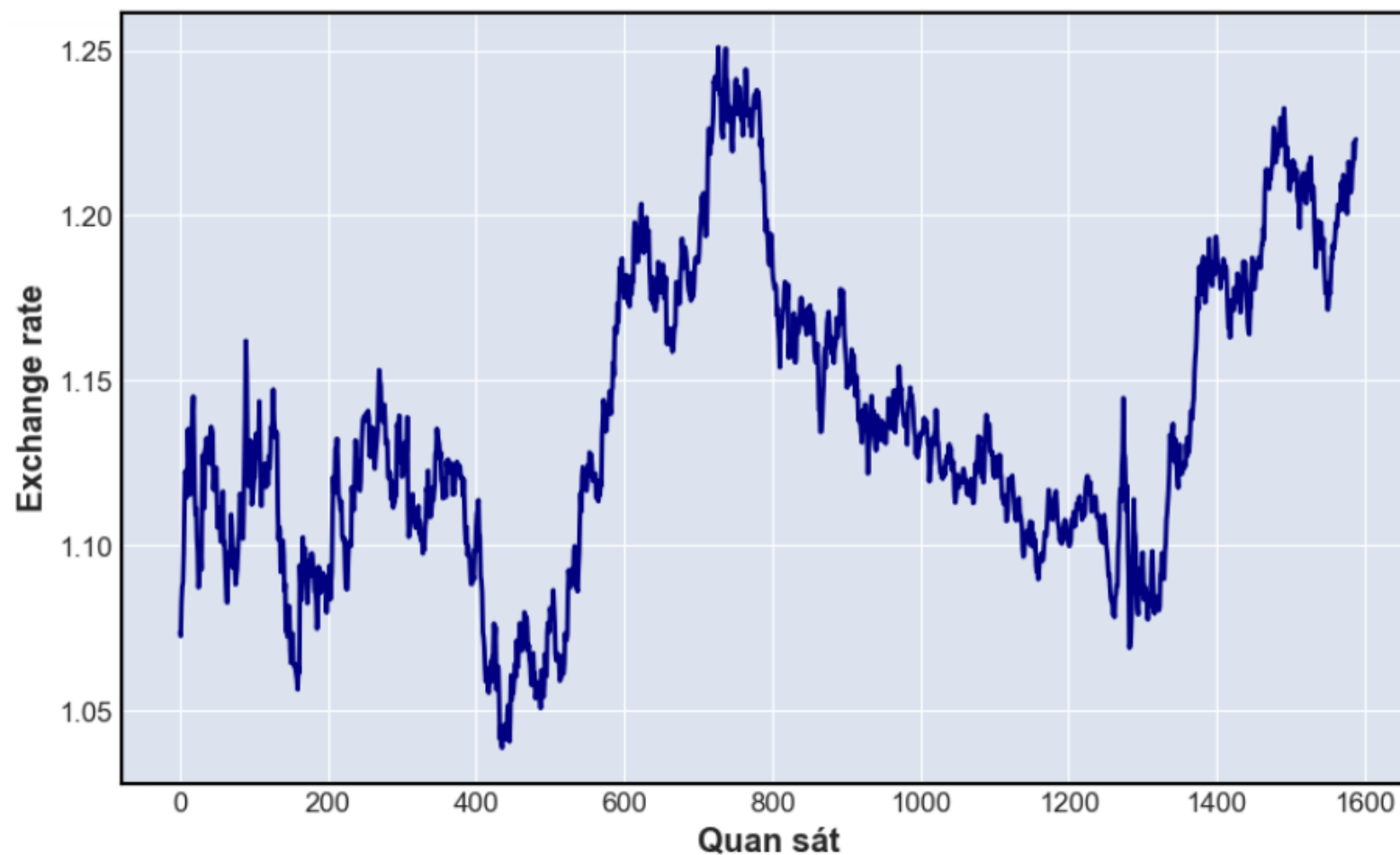
cho phép mô hình hoá cả xu thế ngẫu nhiên lẫn xu thế tất định.

- Khi $\theta_0 = 0$, mô hình chỉ có xu thế ngẫu nhiên.
- Khi $\theta_0 \neq 0$, mô hình chứa xu thế tất định, lẫn trong nhiễu không dừng, tự tương quan, chứa xu thế ngẫu nhiên.



Mô hình tỉ giá EUR/USD

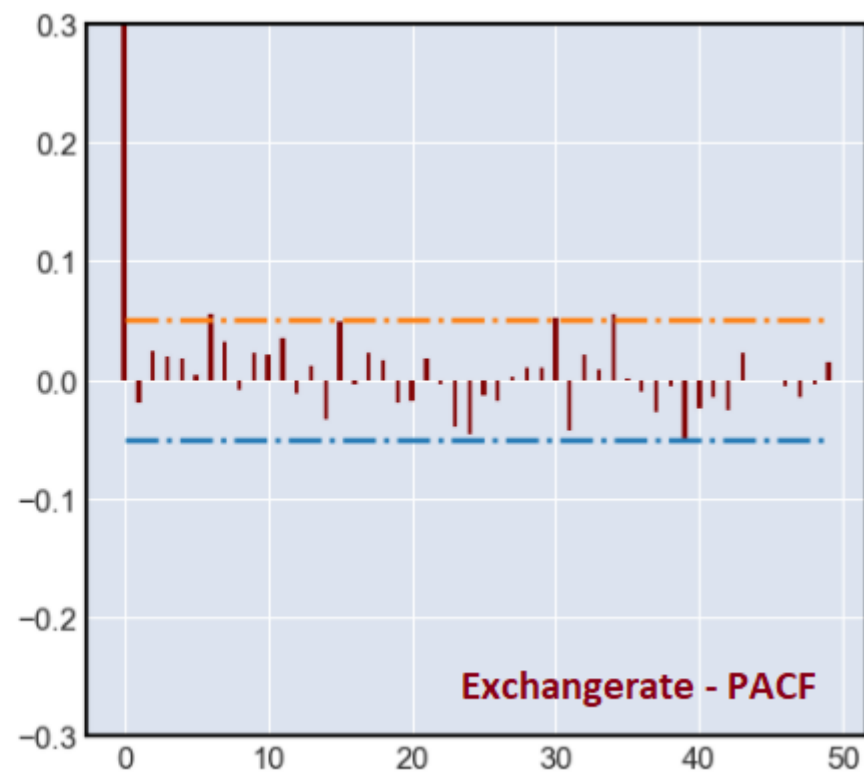
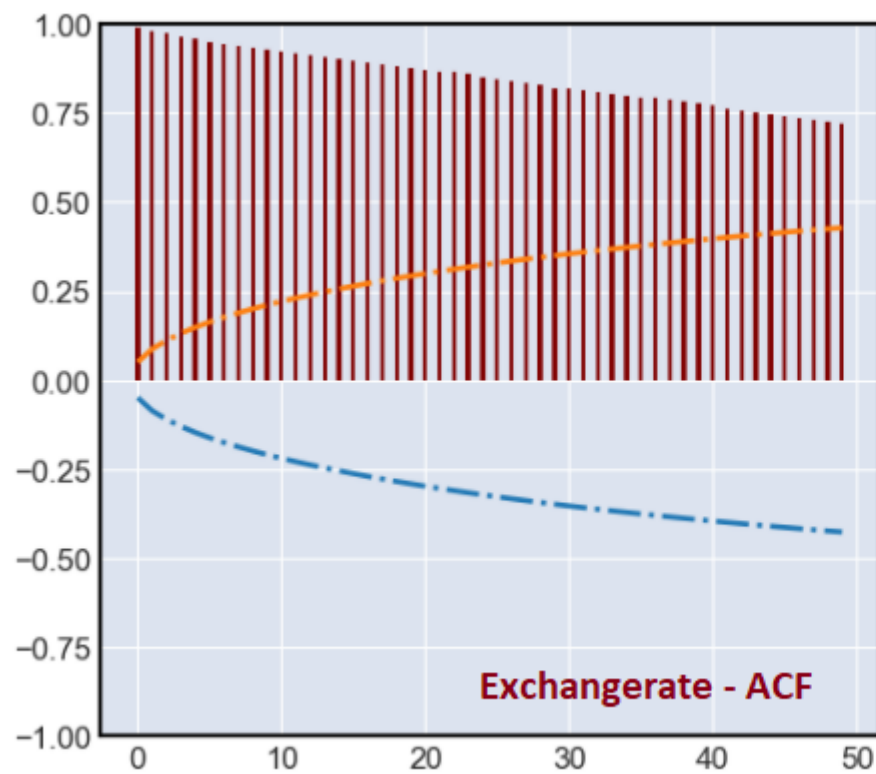
(Dữ liệu có thể lấy từ nguồn: <https://finance.yahoo.com/>)





Mô hình tỉ giá EUR/USD

Lược đồ ACF & PACF





Mô hình tỉ giá EUR/USD

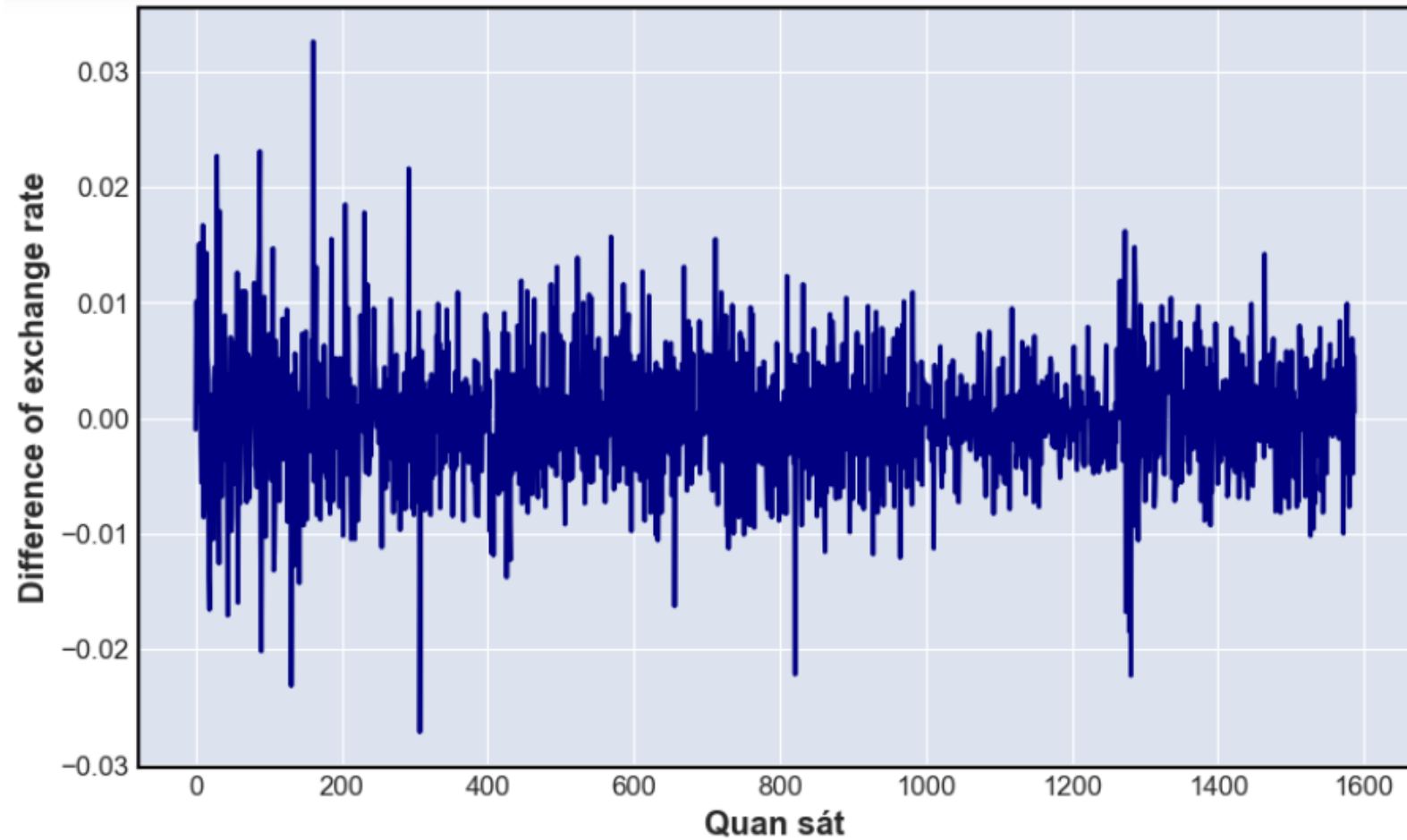
Lược đồ ACF & PACF

- Lược đồ tương quan cho thấy hàm ACF giảm dần về 0 chậm và tuyến tính. Đây là một đặc trưng của chuỗi không dừng.
- Để tìm mô hình ARIMA mô tả dữ liệu tỉ giá, ta lấy sai phân:

$$w_t = \nabla y_t = y_t - y_{t-1}$$

Mô hình tỉ giá EUR/USD

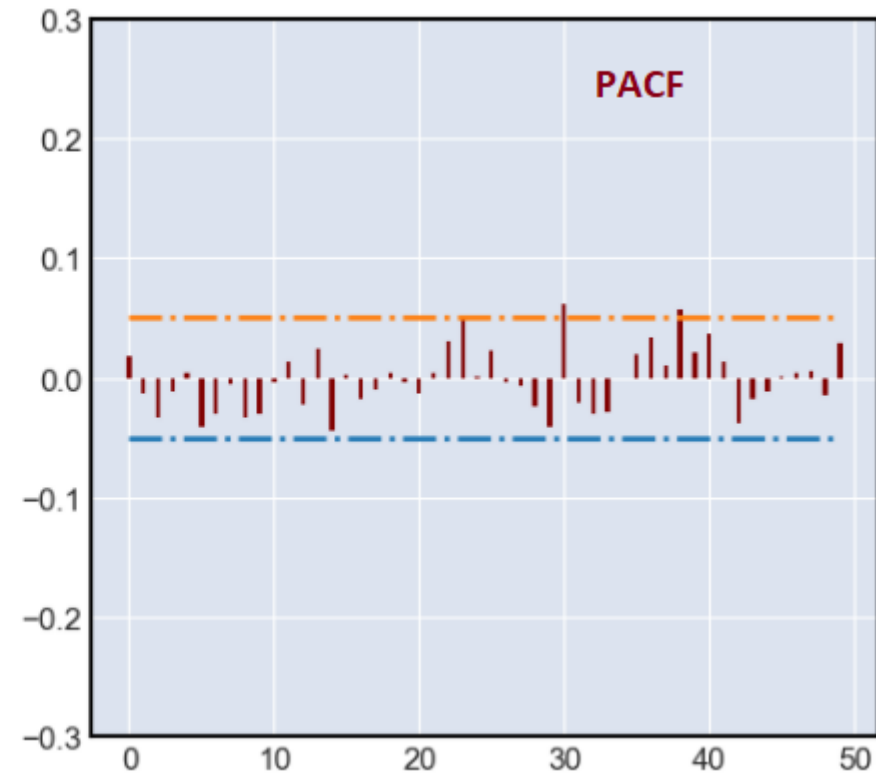
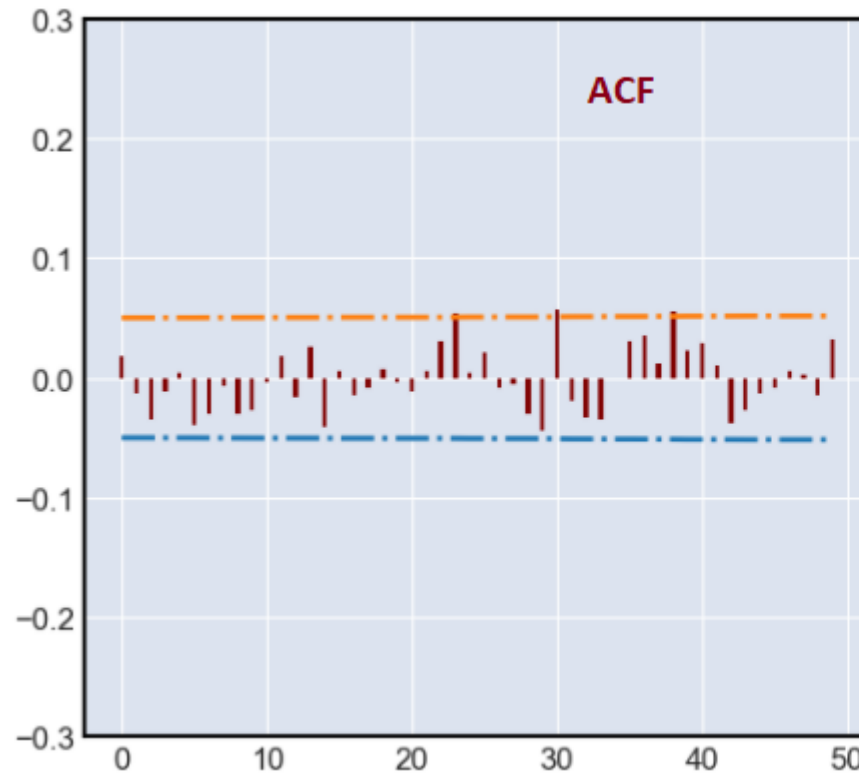
Biểu đồ chuỗi $w_t = \nabla y_t$





Mô hình tỉ giá EUR/USD

ACF & PACF của chuỗi $w_t = \nabla y_t$





Mô hình tỉ giá EUR/USD

Kết luận

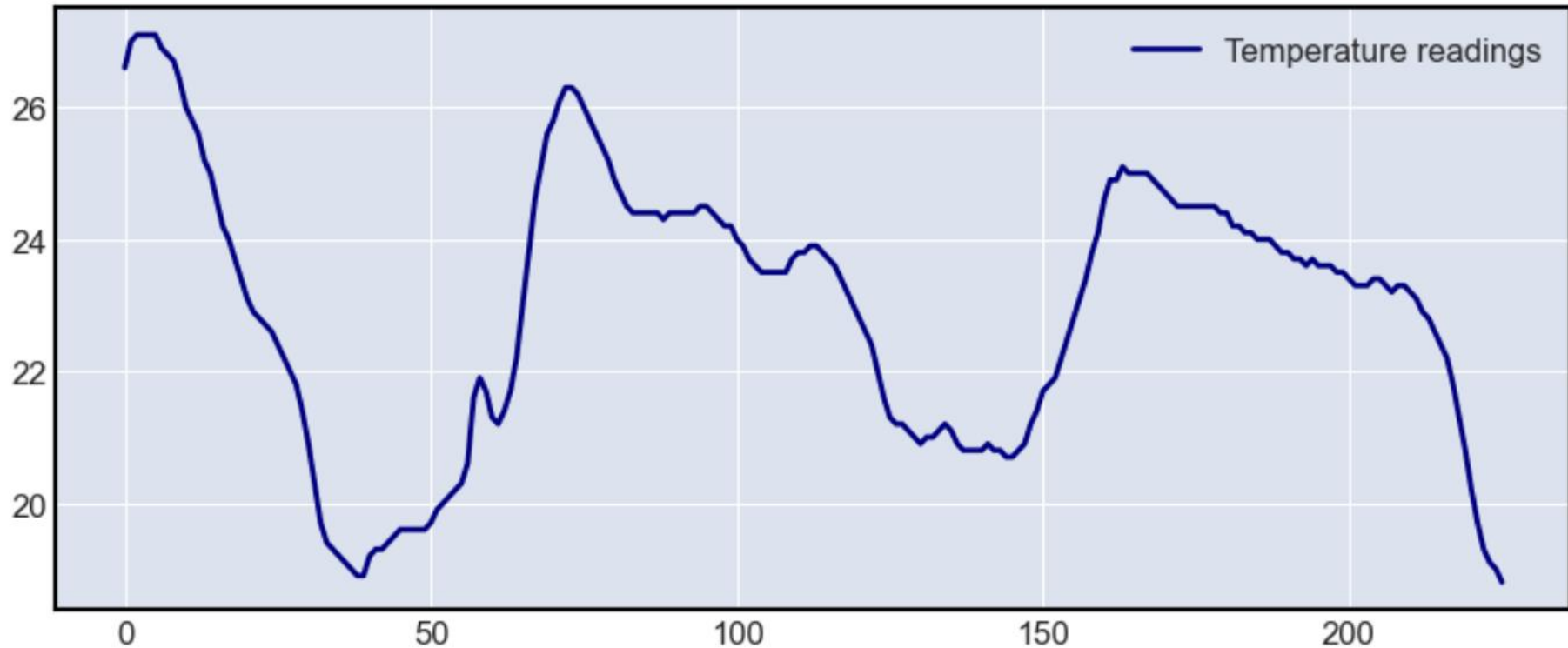
- Các biểu đồ chuỗi w_t cùng với lược đồ tương quan ACF và PACF của nó cho thấy w_t là nhiễu trắng.
- Do đó chuỗi tỉ giá y_t là bước ngẫu nhiên:

$$y_t = y_{t-1} + a_t$$



Mô hình nhiệt độ quá trình hoá học

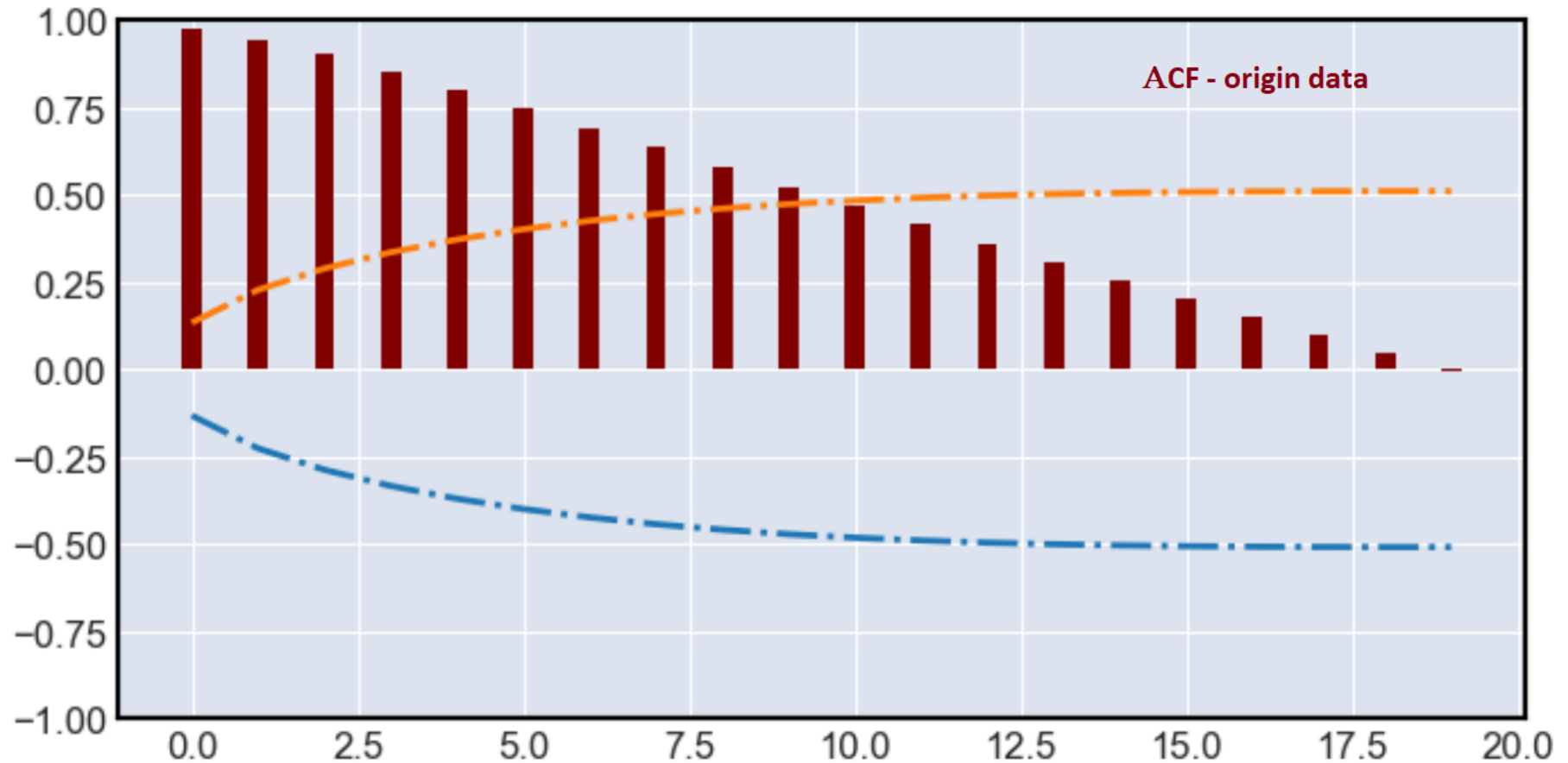
(Box et al., *Time Series Analysis: Forecasting and Control*, Series C)





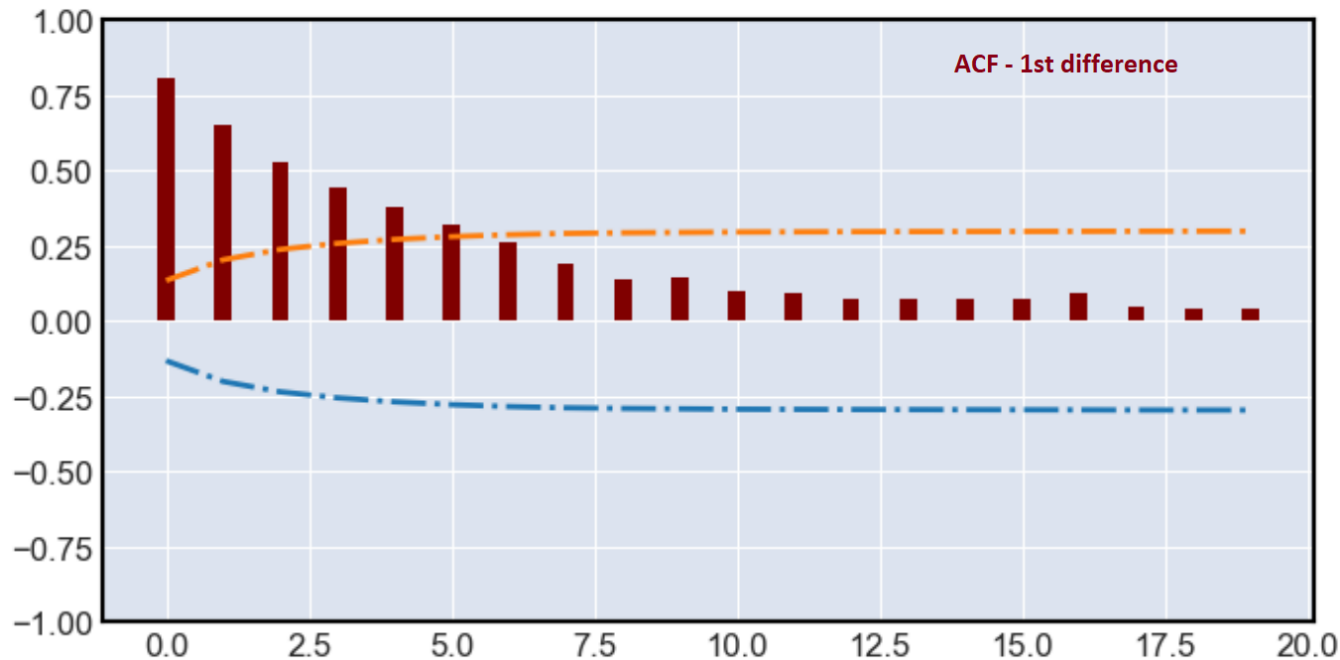
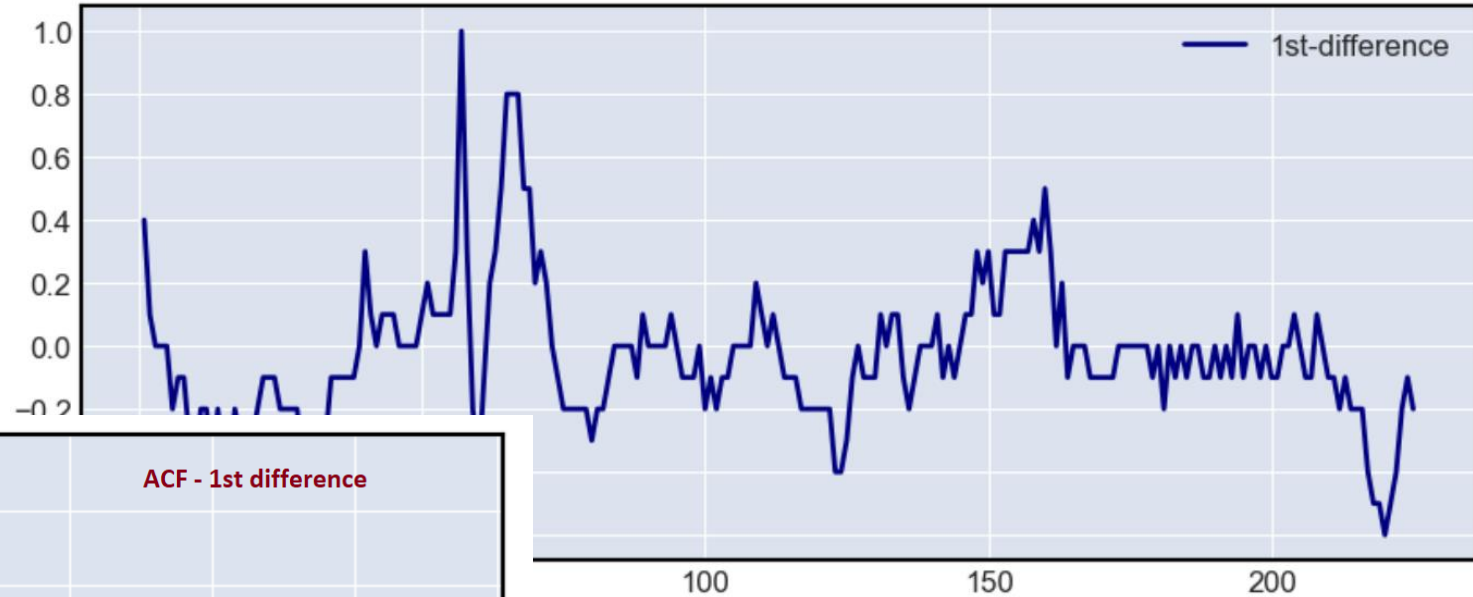
Mô hình nhiệt độ quá trình hoá học

Lược đồ tương quan ACF



Mô hình nhiệt độ quá trình hoá học

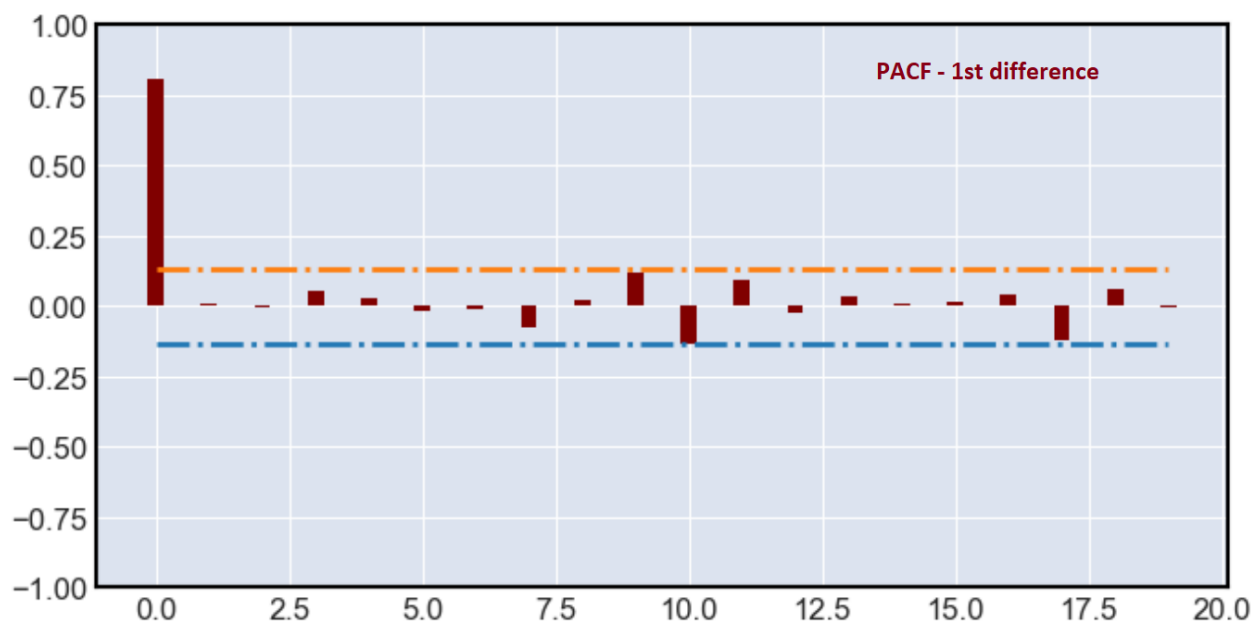
Sai phân và hàm ACF





Mô hình nhiệt độ quá trình hoá học

PACF và mô hình ARIMA



- Mô hình:

$$\nabla y_t = 0.8131 \nabla y_{t-1} + a_t$$

- Thống kê Q:

$$Q(20) = 22.9526$$

$$(p - value = 0.2394)$$

- Hệ số xác định:

$$R^2 = 0.6593$$



4. Kiểm định nghiệm đơn vị

Xác định bậc kết hợp d

Bậc kết hợp d là yếu tố quyết định đến tính chất của một chuỗi thời gian. Các giá trị phổ biến của d là 0 và 1.

- $y_t \sim I(0)$: giả sử $E(y_t) = 0$
 - Phương sai y_t hữu hạn và không phụ thuộc t ,
 - Nhiễu trắng a_t chỉ có ảnh hưởng tạm thời,
 - Kỳ vọng của khoảng thời gian giữa 2 lần cắt trục hoành là hữu hạn, sao cho y_t dao động quanh giá trị trung bình = 0,
 - Các hệ số tự tương quan ρ_k giảm đều về độ lớn khi k lớn, sao cho tổng các hệ số là hữu hạn.



Xác định bậc kết hợp d

- $y_t \sim I(1)$: giả sử $y_0 = 0$
 - Phương sai y_t tiến đến vô cùng khi t tiến đến vô cùng,
 - Nhiễu trắng a_t ảnh hưởng lâu dài đến y_t , vì y_t bằng tổng của các nhiễu trước đó, ví dụ

$$y_t = \nabla^{-1} a_t = S a_t = \sum_{i=0}^{t-1} a_{t-i}$$

- Kỳ vọng của khoảng thời gian giữa 2 lần cắt trục hoành là vô hạn,
- Các hệ số tự tương quan ρ_k tiến đến 1 khi k tiến đến vô cùng.



Xác định bậc kết hợp d

- Việc xác định “kiểu” không dừng là không dễ, nên ta phải kiểm tra (?) hàm ACF mẫu của chuỗi sai phân các cấp $\nabla^k y_t$.
- ACF của chuỗi không dừng:**

Xét phương trình đặc trưng của quá trình AR(p):

$$\phi(B) = (1 - g_1 B)(1 - g_2 B) \cdots (1 - g_p B)$$

Giả sử $g_1 = 1 - \delta$ (δ đủ nhỏ), hệ số tự tương quan có dạng:

$$\begin{aligned} \rho_k &= A_1 g_1^k + A_2 g_2^k + \cdots + A_p g_p^k \\ &\approx A_1 g_1^k = A_1 (1 - \delta)^k \approx A_1 (1 - \delta k) \end{aligned}$$

Kết luận, **ACF giảm chậm, tuyến tính về 0**. Khi đó ta xét tiếp lần lượt các chuỗi sai phân $\nabla y_t, \nabla^2 y_t, \dots$ cho đến khi có một chuỗi sai phân dừng.



Sai phân quá mức

- Dựa hoàn toàn vào ACF có thể dẫn đến hiện tượng lấy sai phân quá mức cần thiết.
- **Ví dụ:** Xét quá trình dừng MA(1) $y_t = (1 - \theta B)a_t$.

Lấy sai phân bậc nhất

$$\begin{aligned}\nabla y_t &= (1 - B)(1 - \theta B)a_t \\ &= (1 - \theta_1 B + \theta_2 B^2)a_t\end{aligned}$$

- Mô hình phức tạp hơn với 2 tham số.
- Mô hình không khả nghịch (vì đa thức đặc trưng có nghiệm 1), biểu diễn AR(∞) của ∇y_t không tồn tại, nên việc ước lượng mô hình khó hơn.



Kiểm định nghiệm đơn vị

- Bắt đầu với quá trình AR(1):

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

- Ta đã biết: $y_t \sim I(0)$ nếu $|\phi| < 1$, và

$$y_t \sim I(1) \text{ nếu } \phi = 1$$

Do đó phương pháp xác định bậc liên kết d là kiểm định giả thuyết (phương trình đặc trưng) “có nghiệm đơn vị”.

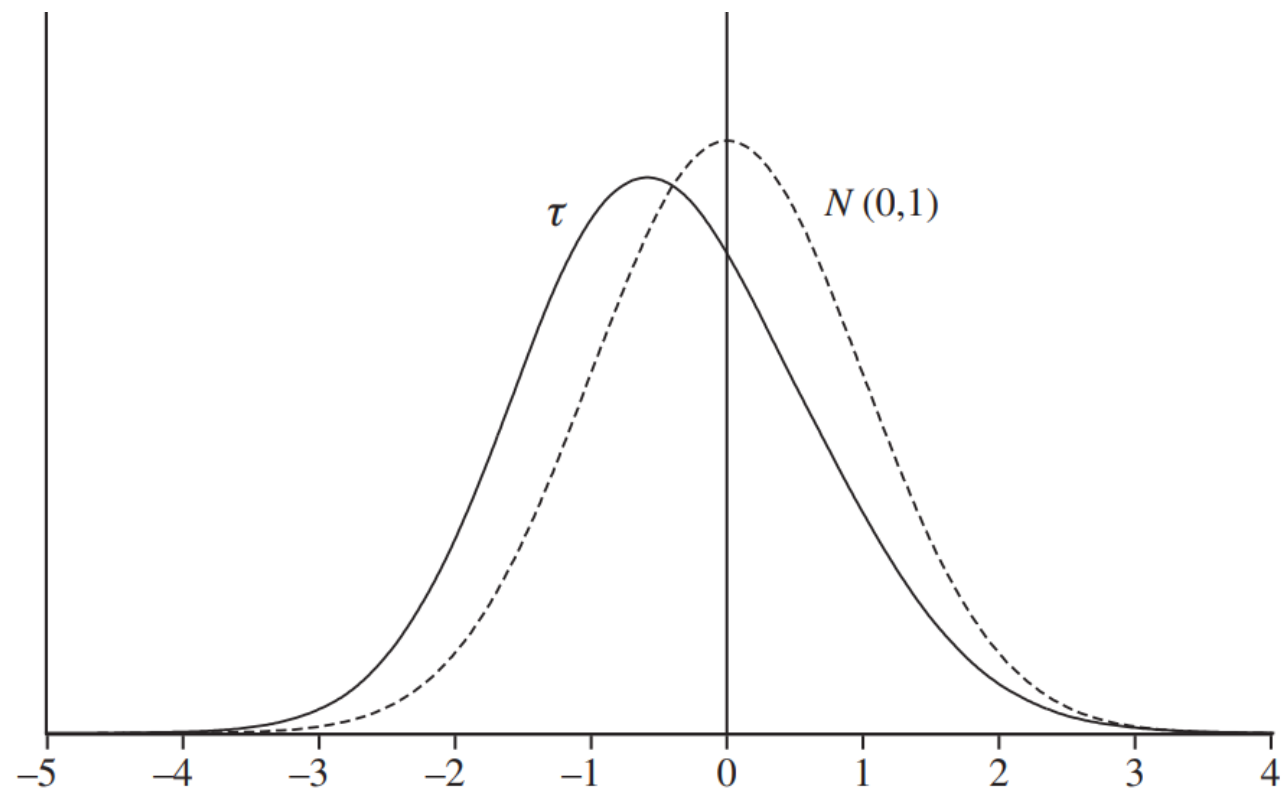
- Kiểm định **Dickey-Fuller**:

$$\tau = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$



Phân phối của thống kê τ

| Mức ý nghĩa | Giá trị tới hạn |
|-------------|-----------------|
| 5.0% | -1.95 |
| 2.5% | -2.23 |
| 1.0% | -2.58 |

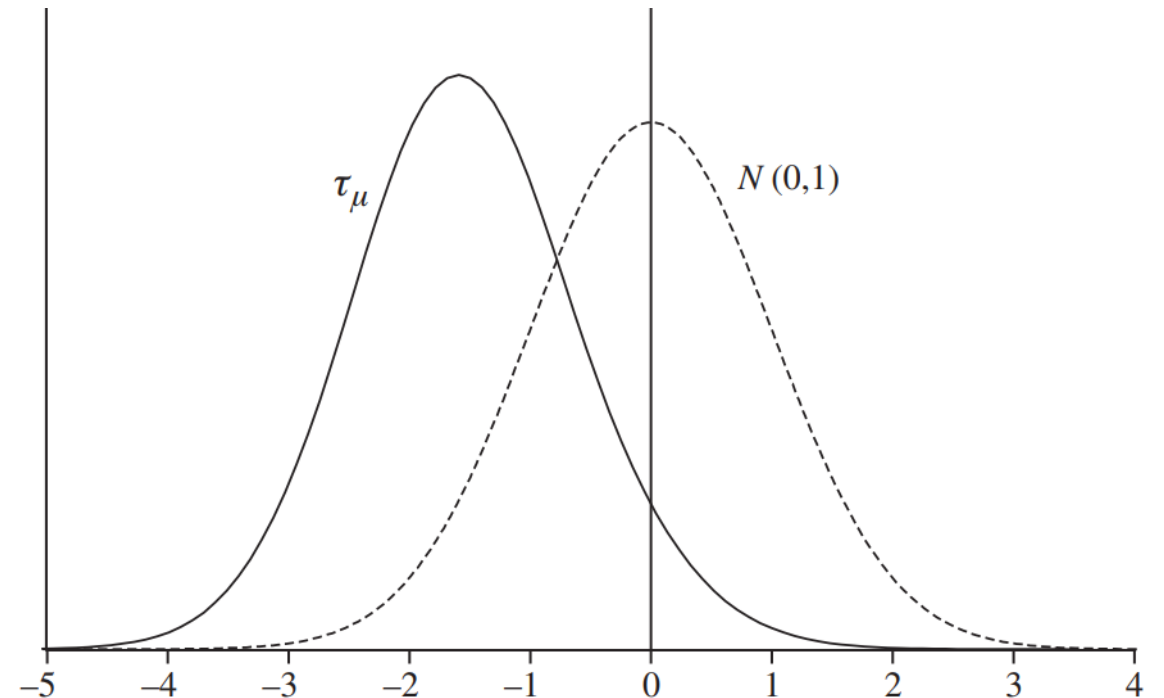


Phân phối của thống kê τ_μ

- Kiểm định Dickey-Fuller được mở rộng cho quá trình AR(1) có dịch chuyển:

$$y_t = \theta_0 + \phi y_{t-1} + a_t$$

| Mức ý nghĩa | Giá trị tới hạn |
|-------------|-----------------|
| 5.0% | -2.86 |
| 2.5% | -3.12 |
| 1.0% | -3.43 |



Mở rộng cho quá trình AR(p)

- Xét quá trình AR(p):

$$y_t = \theta_0 + \sum_{i=1}^p \phi_i y_{t-i} + a_t$$
$$\Leftrightarrow y_t = \theta_0 + \phi y_{t-1} + \sum_{i=1}^k \delta_i \nabla y_{t-i} + a_t$$

ở đó:

$$\phi = \sum_{i=1}^p \phi_i ; \quad \delta_i = - \sum_{j=i+1}^{p-1} \phi_j ; \quad k = p - 1$$

- Do đó, giả thuyết “có nghiệm đơn vị” tương đương với $\phi = 1$.
- Kiểm định ADF (**augmented Dickey-Fuller**):

$$\tau_\mu = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$





Ví dụ áp dụng

(Kiểm định mô hình tỉ giá EUR/USD)

So sánh hai mô hình: AR(1) và “bước ngẫu nhiên không dịch chuyển” cho dữ liệu tỉ giá EUR/USD

- Xây dựng mô hình hồi quy ADF:

$$y_t = \frac{0.00754}{(0.00339)} + \frac{0.99345}{(0.00298)} y_{t-1} + a_t$$

- Giá trị thống kê kiểm định:

$$\tau_\mu = \frac{\hat{\phi} - 1}{se(\hat{\phi})} = \frac{0.99345 - 1}{0.00298} = -2.19$$

- Giá trị tới hạn ở mức ý nghĩa 5% là -2.86 nên mô hình “bước ngẫu nhiên không dịch chuyển” phù hợp hơn.



Dừng xu thế (TS) - Dừng sai phân (DS)

- y_t là dừng sai phân (DS = difference stationary) nếu:

$$\nabla y_t = \theta_0 + \theta(B)a_t$$

- y_t là dừng xu thế (TS = trend stationary) nếu:

$$\nabla y_t = \beta_0 + \beta_1 t + \theta(B)a_t$$

- Để kiểm định chuỗi y_t là DS hay TS, ta bổ sung vào mô hình hồi quy ADF thành phần xu thế theo thời gian:

$$y_t = \beta_0 + \beta_1 t + \phi y_{t-1} + \sum_{i=1}^k \delta_i \nabla y_{t-i} + a_t$$

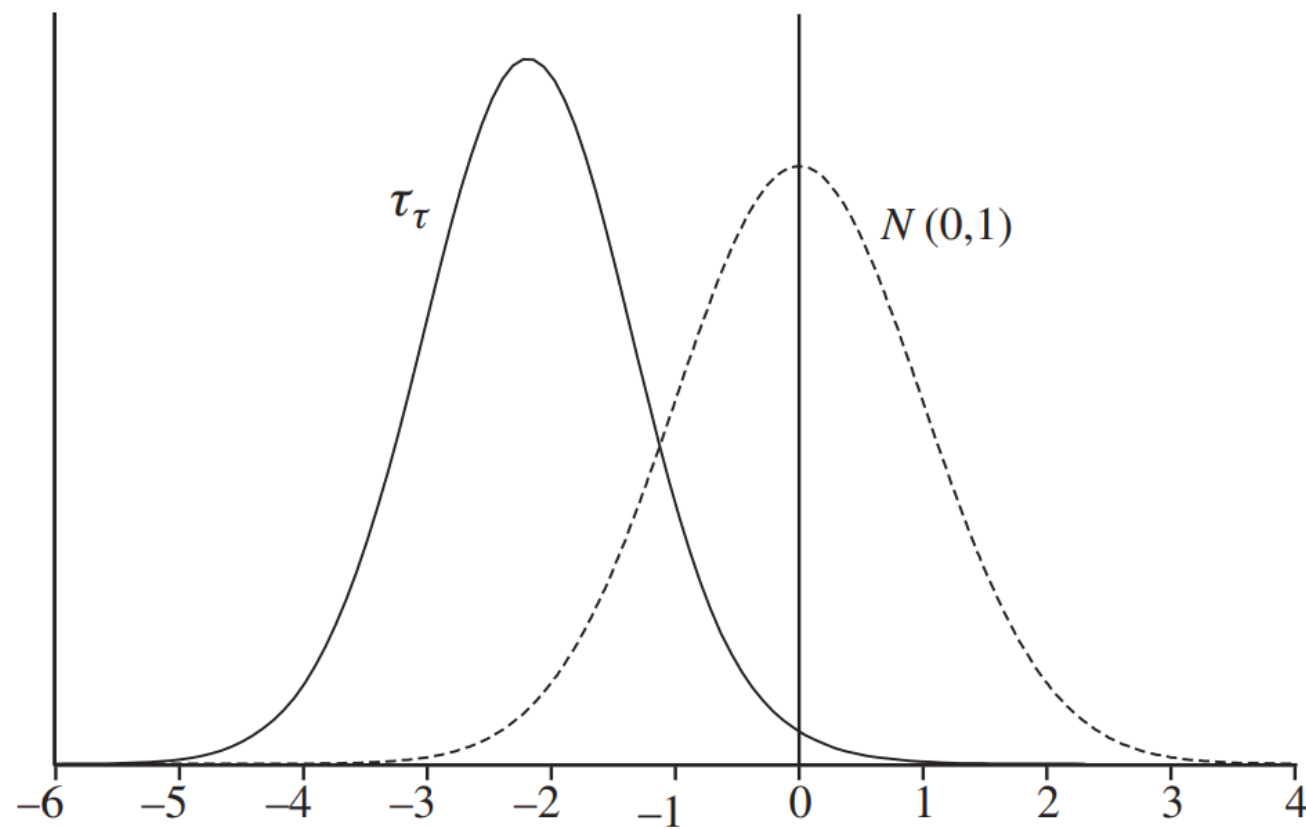
- Thống kê kiểm định τ_τ (có phân phối khác với τ_μ):

$$\tau_\tau = \frac{\hat{\phi} - 1}{se(\hat{\phi})}$$



Phân phối của thống kê τ_τ

| Mức ý nghĩa | Giá trị tới hạn |
|-------------|-----------------|
| 5.0% | -3.41 |
| 2.5% | -3.66 |
| 1.0% | -3.96 |



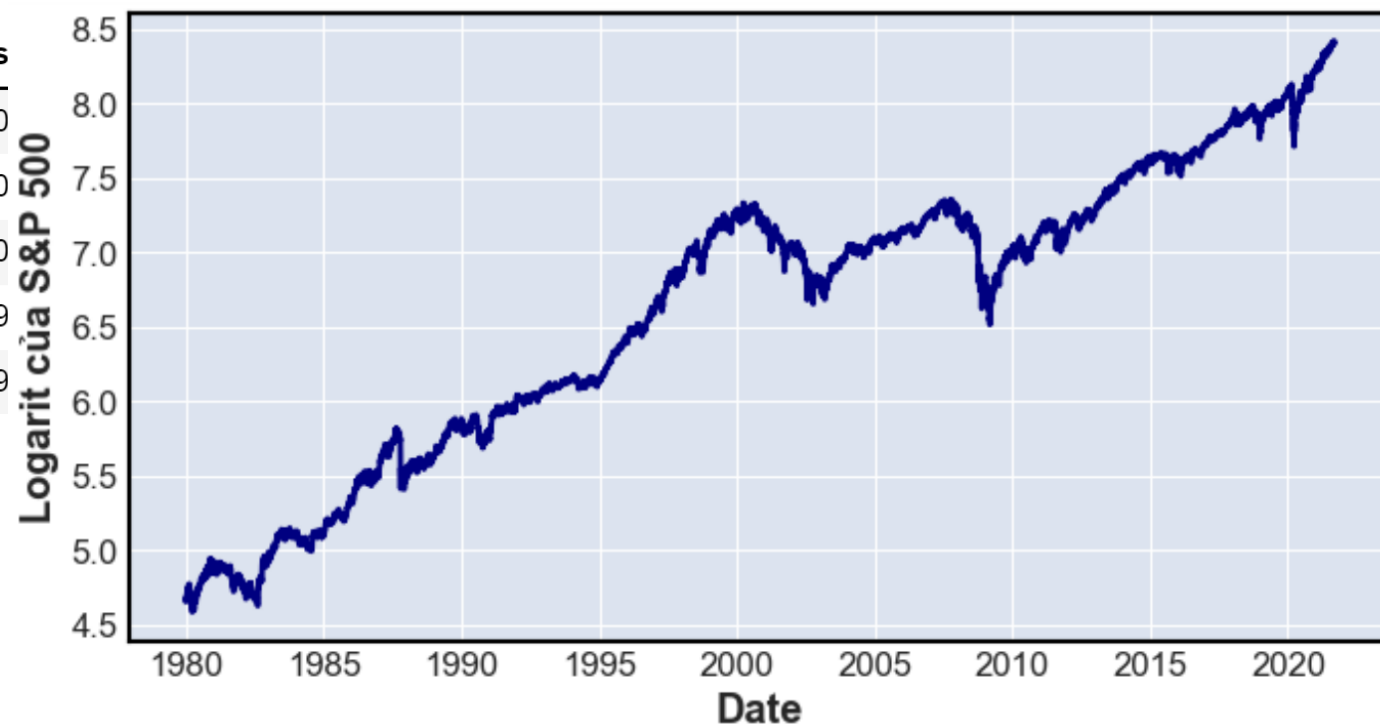


Ví dụ áp dụng

(Chỉ số S&P 500 là TS hay DS?)

```
# Tải dữ liệu
start = dt.datetime(1980,1,1)
end = dt.datetime.now()
df = web.DataReader('^GSPC', 'yahoo', start, end)
df.reset_index(inplace=True)
df.head()
```

| | Date | High | Low | Open | Clos |
|---|------------|------------|------------|------|-----------|
| 0 | 1979-12-31 | 108.529999 | 107.260002 | 0.0 | 107.94000 |
| 1 | 1980-01-02 | 108.430000 | 105.290001 | 0.0 | 105.76000 |
| 2 | 1980-01-03 | 106.080002 | 103.260002 | 0.0 | 105.22000 |
| 3 | 1980-01-04 | 107.080002 | 105.089996 | 0.0 | 106.51999 |
| 4 | 1980-01-07 | 107.800003 | 105.800003 | 0.0 | 106.80999 |





Ví dụ áp dụng

(Chỉ số S&P 500 là TS hay DS?)

- Ước lượng mô hình ADF với $k = 1$ (theo tiêu chuẩn BIC):

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\phi} y_{t-1} + \hat{\delta}_1 \nabla y_{t-1} + a_t$$

- Thống kê kiểm định:

$$\tau_{\tau} = \frac{\hat{\phi} - 1}{se(\hat{\phi})} = \frac{0.99909 - 1}{0.00041} = -2.2$$

- Giá trị tới hạn ở mức ý nghĩa 5% là -3.41 nên (logarit) chỉ số S&P 500 là dừng sai phân.

| | param | se(param) |
|---|---------------|--------------|
| 0 | 4.965058e-03 | 2.076282e-03 |
| 1 | 2.690535e-07 | 1.317321e-07 |
| 2 | 9.990923e-01 | 4.127564e-04 |
| 3 | -5.548825e-02 | 9.737955e-03 |

- Nếu chỉ số S&P 500 là TS, nó sẽ dao động xung quanh đường xu thế tuyến tính. Điều đó cho phép các nhà đầu tư đặt cược bất cứ khi nào chỉ số rời xa đường xu thế. Tuy nhiên, điều này là không chính xác.



5. Giới thiệu về dự báo

Dự báo quá trình ARIMA

- **Bài toán:** Giả sử dãy quan sát $y_{1-d}, y_{2-d}, \dots, y_T$ được sinh bởi một quá trình ARIMA(p,d,q):

$$\phi(B)\nabla^d y_t = \theta_0 + \theta(B)a_t$$

Hãy dự báo về giá trị tương lai y_{T+h} , h = thời hạn dự báo.

- Dự báo bao gồm
 - Giá trị dự báo cho y_{T+h} , ký hiệu $\hat{y}_{T,h}$
 - Sai số dự báo, ký hiệu $e_T(h) = y_{T+h} - \hat{y}_{T,h}$



Giá trị dự báo

- Đặt $\alpha(B) = \phi(B)\nabla^d = 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_{p+d} B^{p+d}$. Thời điểm $T + h$:
$$y_{T+h} = \alpha_1 y_{T+h-1} + \dots + \alpha_{p+d} y_{T+h-p-d} + \theta_0 + a_{T+h} + \theta_1 a_{T+h-1} + \dots + \theta_q a_{T+h-q}$$
- Dự báo “tốt nhất” cho y_{T+h} là: $\hat{y}_{T,h} = E[y_{T+h} | y_T, y_{T-1}, \dots]$, khi đó $E[e_T^2(h)]$ đạt cực tiểu.
- Ta biết rằng:

$$E[y_{T+j} | y_T, \dots] = \begin{cases} y_{T+j} & (j \leq 0) \\ \hat{y}_{T,j} & (j > 0) \end{cases}; \quad E[a_{T+j} | y_T, \dots] = \begin{cases} a_{T+j} & (j \leq 0) \\ 0 & (j > 0) \end{cases}$$

- Do đó, để ước lượng $\hat{y}_{T,h}$, ta phải làm như sau:
 - Thay thế kỳ vọng quá khứ ($j < 0$) bởi y_{T+j} , a_{T+j} đã biết
 - Thay thế kỳ vọng tương lai ($j \geq 0$) bởi $\hat{y}_{T,j}$ và 0.



Ví dụ

Tính giá trị dự báo

- Xét việc dự báo quá trình AR(2):

$$(1 - \phi_1 B - \phi_2 B^2)y_t = \theta_0 + a_t$$

- Ta có:

$$y_{T+h} = \phi_1 y_{T+h-1} + \phi_2 y_{T+h-2} + \theta_0 + a_{T+h}$$

- Các giá trị dự báo với $h = 1$, $h = 2$ và $h > 2$ là:

$$\hat{y}_{T,1} = \phi_1 y_T + \phi_2 y_{T-1} + \theta_0$$

$$\hat{y}_{T,2} = \phi_1 \hat{y}_{T,1} + \phi_2 y_T + \theta_0$$

$$\hat{y}_{T,h} = \phi_1 \hat{y}_{T,h-1} + \phi_2 \hat{y}_{T,h-2} + \theta_0$$

- Theo điều kiện dừng của AR(2), khi $h \rightarrow \infty$

$$\hat{y}_{T,h} \rightarrow \frac{\theta_0}{1 - \phi_1 - \phi_2} = E(y_t)$$



Sai số dự báo

- Biến đổi quá trình ARIMA thành $MA(\infty)$ với trọng số ψ :

$$\begin{aligned} y_{T+h} &= \mu + \sum_{j=0}^{\infty} \psi_j a_{T+h-j} & (\psi_0 = 1) \\ &= \mu + \sum_{j=0}^{h-1} \psi_j a_{T+h-j} + \sum_{j=h}^{\infty} \psi_j a_{T+h-j} \end{aligned}$$

- Suy ra

$$\hat{y}_{T,h} = \mu + \sum_{j=h}^{\infty} \psi_j a_{T+h-j}$$

- Sai số dự báo có dạng:

$$e_T(h) = y_{T+h} - \hat{y}_{T,h} = \sum_{j=0}^{h-1} \psi_j a_{T+h-j}$$



Sai số dự báo

- Kỳ vọng và phương sai của $e_T(h)$:

$$\begin{aligned} E[e_T(h)] &= 0 \\ \text{Var}[e_T(h)] &= \text{Var} \left[\sum_{j=0}^{h-1} \psi_j a_{T+h-j} \right] \\ &= \sigma^2 \sum_{j=0}^{h-1} \psi_j^2 = \sigma^2(h) \end{aligned}$$

- Nếu giả thiết $a_t \sim WN(0, \sigma^2)$ thì $e_T(h) \sim N(0, \sigma^2(h))$. Do đó, khoảng tin cậy $100(1 - \alpha)$ cho giá trị tương lai y_{T+h} là:

$$\hat{y}_{T,h} \pm z_{\alpha/2} \cdot \sigma(h)$$



Ví dụ

Tính trọng số ψ

Xét mô hình ARMA(1,1):

$$(1 - \phi B)y_t = (1 - \theta B)a_t$$

Các trọng số ψ trong biểu diễn MA(∞) được tìm từ phương trình:

$$\begin{aligned}(1 + \psi_1 B + \psi_2 B^2 + \dots)(1 - \phi B) &= 1 - \theta B \\ \Leftrightarrow 1 + (\psi_1 - \phi)B + (\psi_2 - \psi_1 \phi)B^2 + \dots &= 1 - \theta B\end{aligned}$$

Suy ra

$$\begin{aligned}\psi_1 - \phi &= -\theta & \Rightarrow \psi_1 &= \phi - \theta \\ \psi_2 - \psi_1 \phi &= 0 & \Rightarrow \psi_2 &= \phi(\phi - \theta) \\ \dots & & \dots & \end{aligned}$$

Tổng quát, với mô hình ARMA(1,1), $\psi_j = \phi^{j-1}(\phi - \theta), j = 1, 2, \dots$



Dự báo nhiệt độ quá trình hoá học

(Box et al., *Time Series Analysis: Forecasting and Control*, Series C)

- Mô hình ARIMA(1,1,0) cho nhiệt độ quá trình hoá học:

$$\begin{aligned}\nabla y_t &= 0.8131 \nabla y_{t-1} + a_t \\ \Leftrightarrow y_t &= 1.8131 y_{t-1} - 0.8131 y_{t-2} + a_t\end{aligned}$$

- Giá trị dự báo tại thời điểm T là:

$$\begin{aligned}\hat{y}_{T,1} &= 1.8131 y_T - 0.8131 y_{T-1} \\ \hat{y}_{T,2} &= 1.8131 \hat{y}_{T,1} - 0.8131 y_T \\ \hat{y}_{T,h} &= 1.8131 \hat{y}_{T,2} - 0.8131 \hat{y}_{T,1} \quad (h = 3, 4, \dots)\end{aligned}$$

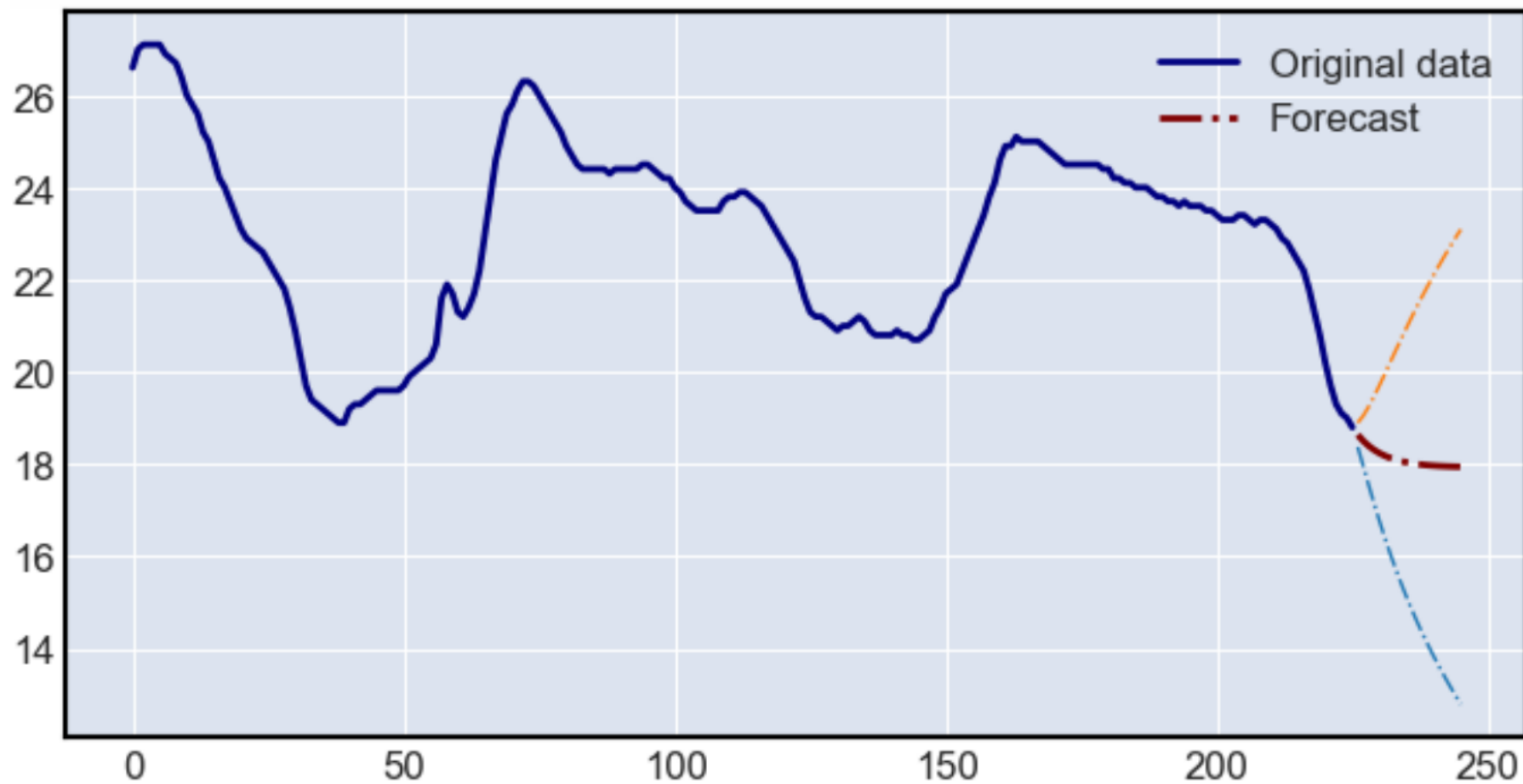
- Các trọng số ψ được tính bởi:

$$\begin{aligned}\psi_1 &= 1.8131 \\ \psi_j &= 1.8131 \psi_{j-1} - 0.8131 \psi_{j-2} \quad (j = 2, 3, \dots)\end{aligned}$$



Dự báo nhiệt độ quá trình hoá học

Giá trị và khoảng dự báo





Dự báo quá trình dừng xu thế

- Xét quá trình dừng xu thế (TS):

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

với ε_t là quá trình ARMA dừng

$$\phi(B)\varepsilon_t = \theta(B)a_t$$

- Giá trị dự báo của y_{T+h} tại thời điểm T là:

$$\hat{y}_{T,h} = \beta_0 + \beta_1 t + \hat{\varepsilon}_{T,h}$$

- Vì ε_t là quá trình dừng nên $\hat{\varepsilon}_{T,h} \rightarrow 0$ khi $h \rightarrow \infty$, do đó
 - Với h lớn, $\hat{y}_{T,h} = \beta_0 + \beta_1 t$ và dự báo được thực hiện đơn giản bằng ngoại suy xu thế tuyến tính.
 - Với h nhỏ, thành phần $\hat{\varepsilon}_{T,h}$ vẫn còn nhưng mất dần khi h tăng lên.
- Sai số dự báo:

$$e_T(h) = y_{T+h} - \hat{y}_{T,h} = \varepsilon_{T+h} - \hat{\varepsilon}_{T,h}$$



Ví dụ áp dụng

Dự báo chỉ số S&P 500

- Giả sử chỉ số S&P 500 được sinh bởi quá trình TS:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

- Ước lượng OLS cho ta mô hình:

$$y_t = 5.00155 + 0.00031t + \varepsilon_t$$

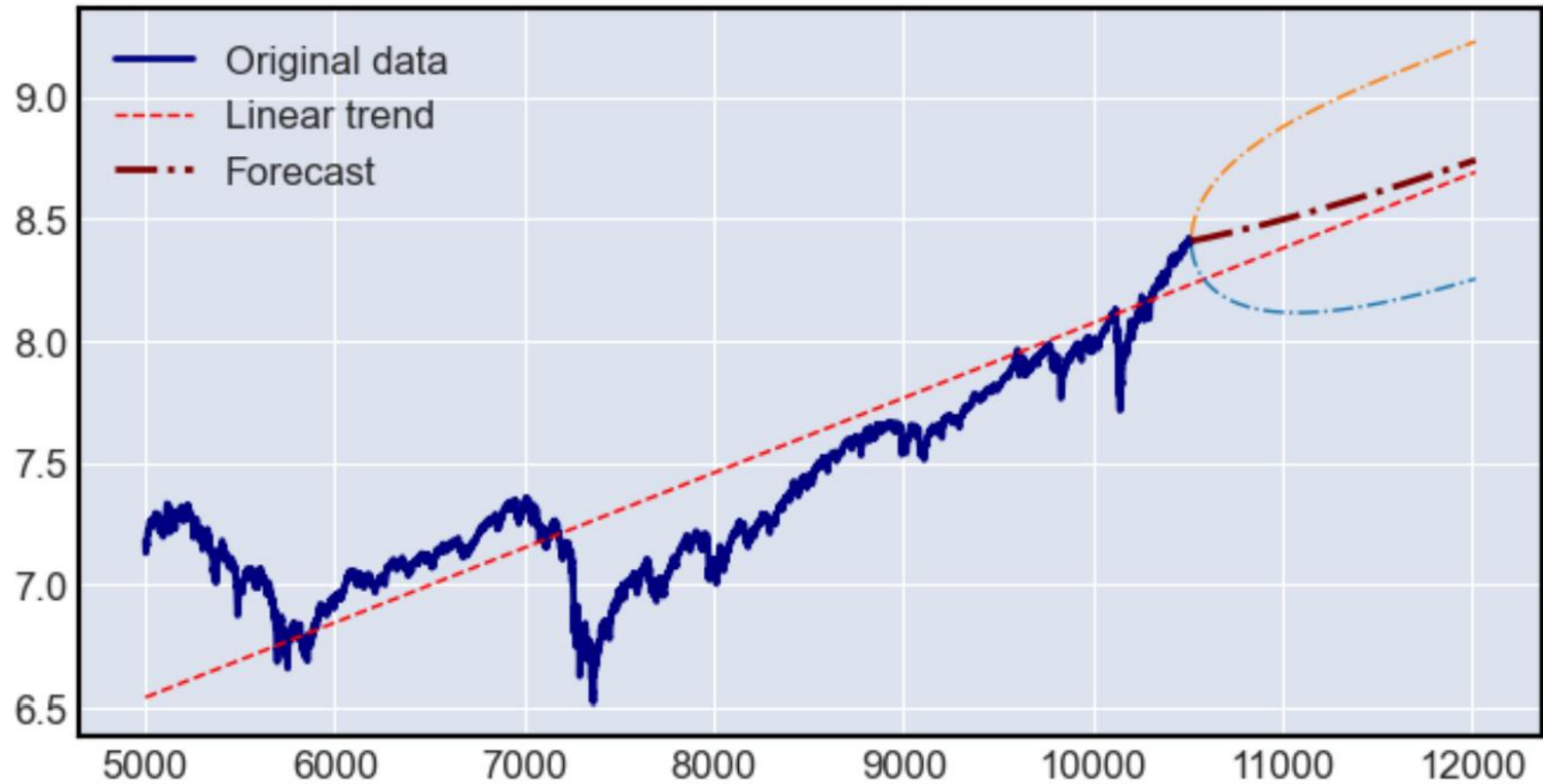
- ε_t là quá trình dừng được mô hình hoá bởi AR(2), theo tiêu chuẩn thông tin BIC:

$$\varepsilon_t = 0.94363 \varepsilon_{t-1} + 0.05546 \varepsilon_{t-2} + a_t$$



Ví dụ áp dụng

Dự báo chỉ số S&P 500





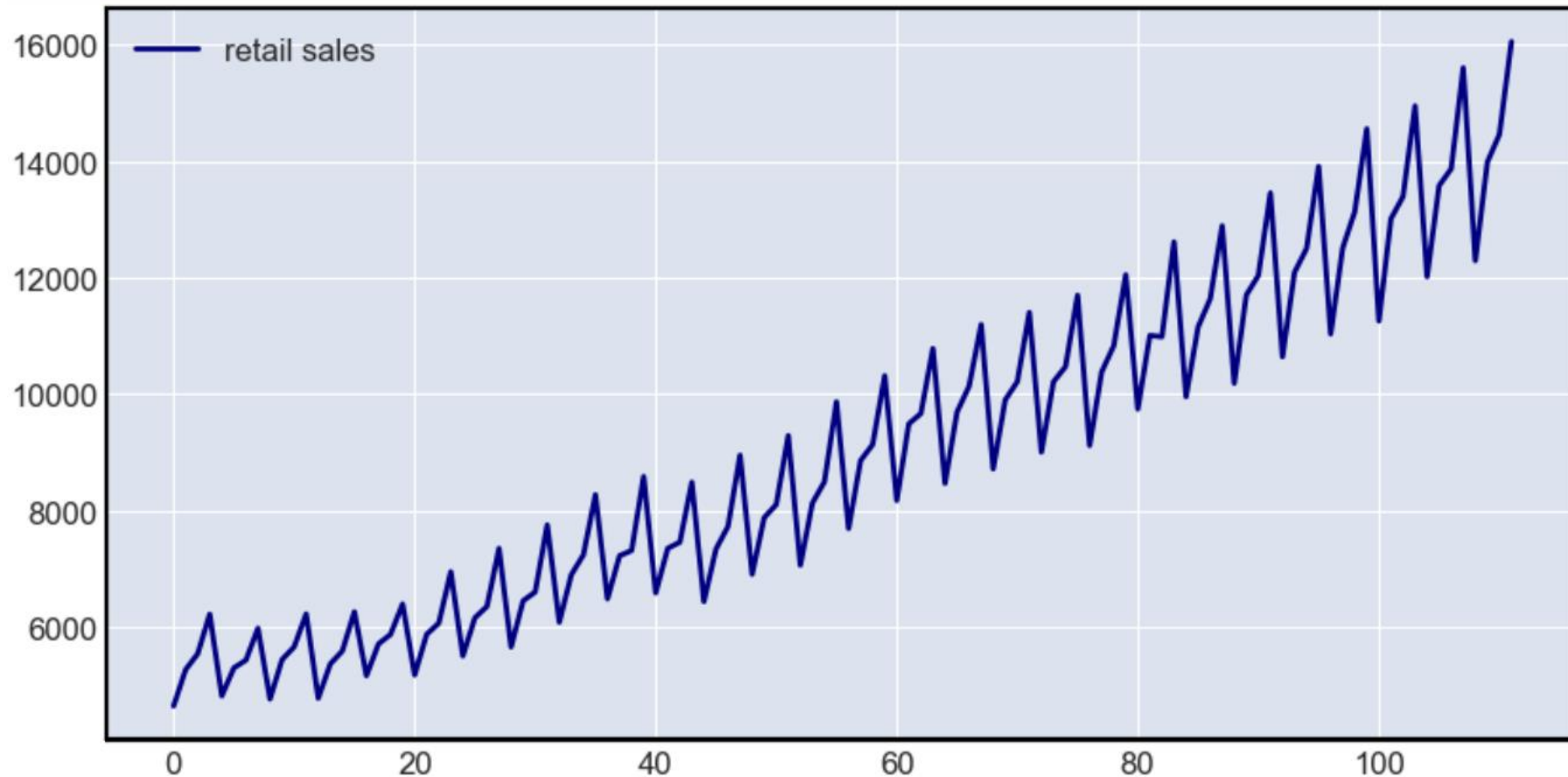
6. Các quá trình có thuộc tính mùa

- Thuộc tính mùa xuất hiện trong nhiều chuỗi thời gian, khi quan sát ở tần suất lớn hơn hàng năm, thường là hàng tháng, hay hàng quý.
- Sự xuất hiện của tính mùa có thể dễ dàng nhận thấy ở biểu đồ chuỗi thời gian, hay trên lược đồ tương quan ACF của dữ liệu được lấy sai phân tương thích.
- Rõ ràng, chuyển động mùa là một thuộc tính có thể dự báo được, và do đó có thể mô hình hoá, hay loại bỏ bằng các phương pháp hiệu chỉnh mùa phù hợp.

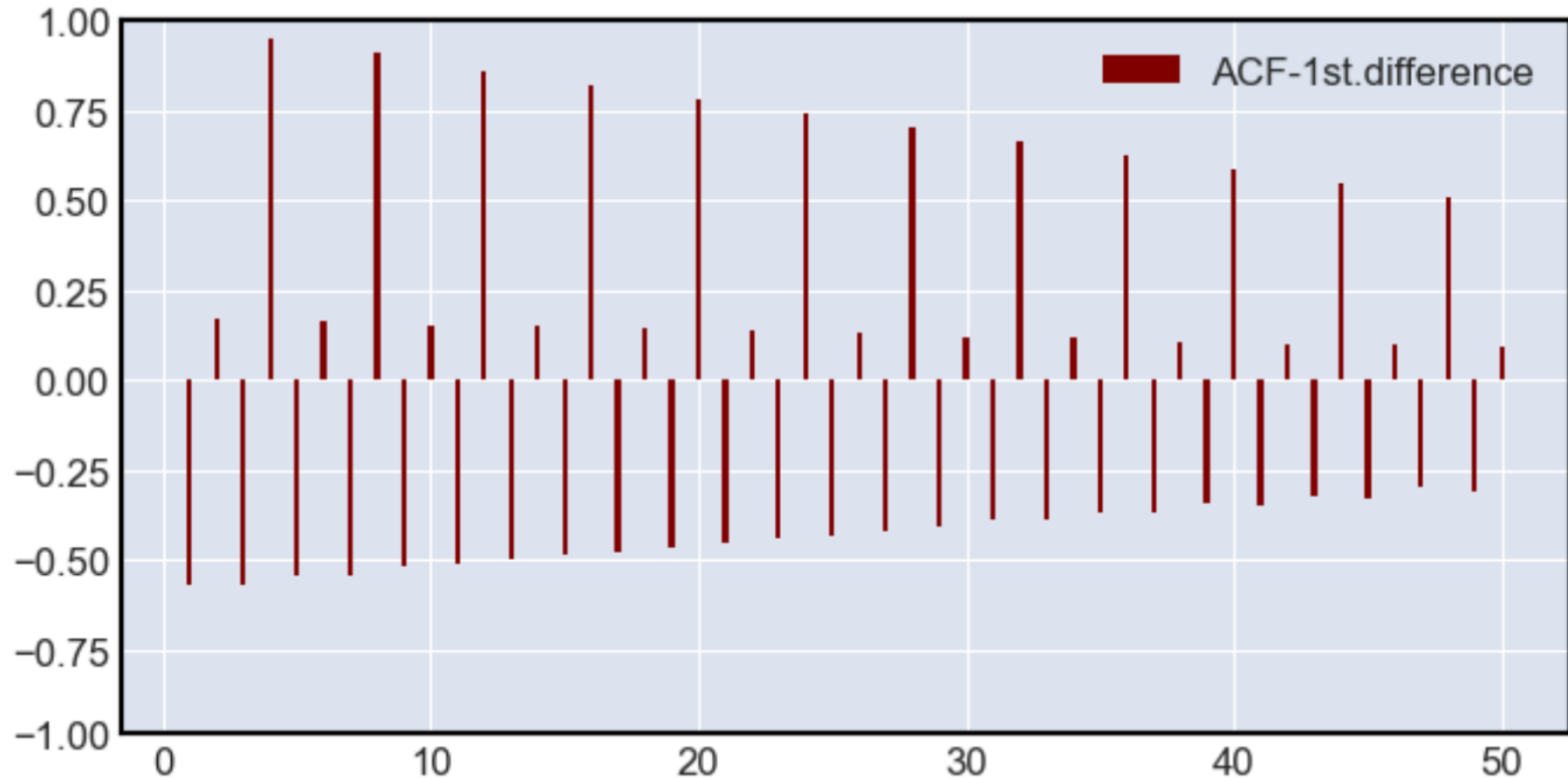


Doanh số bán lẻ đồ uống theo quý, 1992-2019

(nguồn: <https://fred.stlouisfed.org/series/MRTSSM4453USN>)



ACF của dữ liệu lấy sai phân cấp 1





Thành phần mùa tất định

- Một cách đơn giản để mô hình hoá thành phần là sử dụng “trung bình mùa” như sau:

$$y_t = \sum_{i=1}^m \alpha_i s_{i,t} + \varepsilon_t$$

trong đó:

- $s_{i,t}$ là biến mùa (giả), $s_{i,t} = \begin{cases} 1 & \text{mùa } i \\ 0 & \text{mùa } j \neq i \end{cases}$
- m là số mùa trong năm
- ε_t là quá trình ARIMA
- Mô hình dựa trên giả thiết rằng thành phần mùa là tất định, tức là các “trung bình mùa” α_i , $i = 1, 2, \dots, m$ không đổi theo thời gian.

Ví dụ

- Ta thử xây dựng mô hình mùa tất định cho dữ liệu doanh số đồ uống có cồn theo quý:

$$y_t = \sum_{i=1}^4 \alpha_i S_{i,t} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2)$$

- Ước lượng của “trung bình mùa” tính được:

| Q1 | Q2 | Q3 | Q4 |
|------|------|------|-------|
| 7801 | 8859 | 9143 | 10291 |

- Mô hình thực tế không chính xác vì ε_t không phải nhiễu trắng.

Thành phần mùa ngẫu nhiên

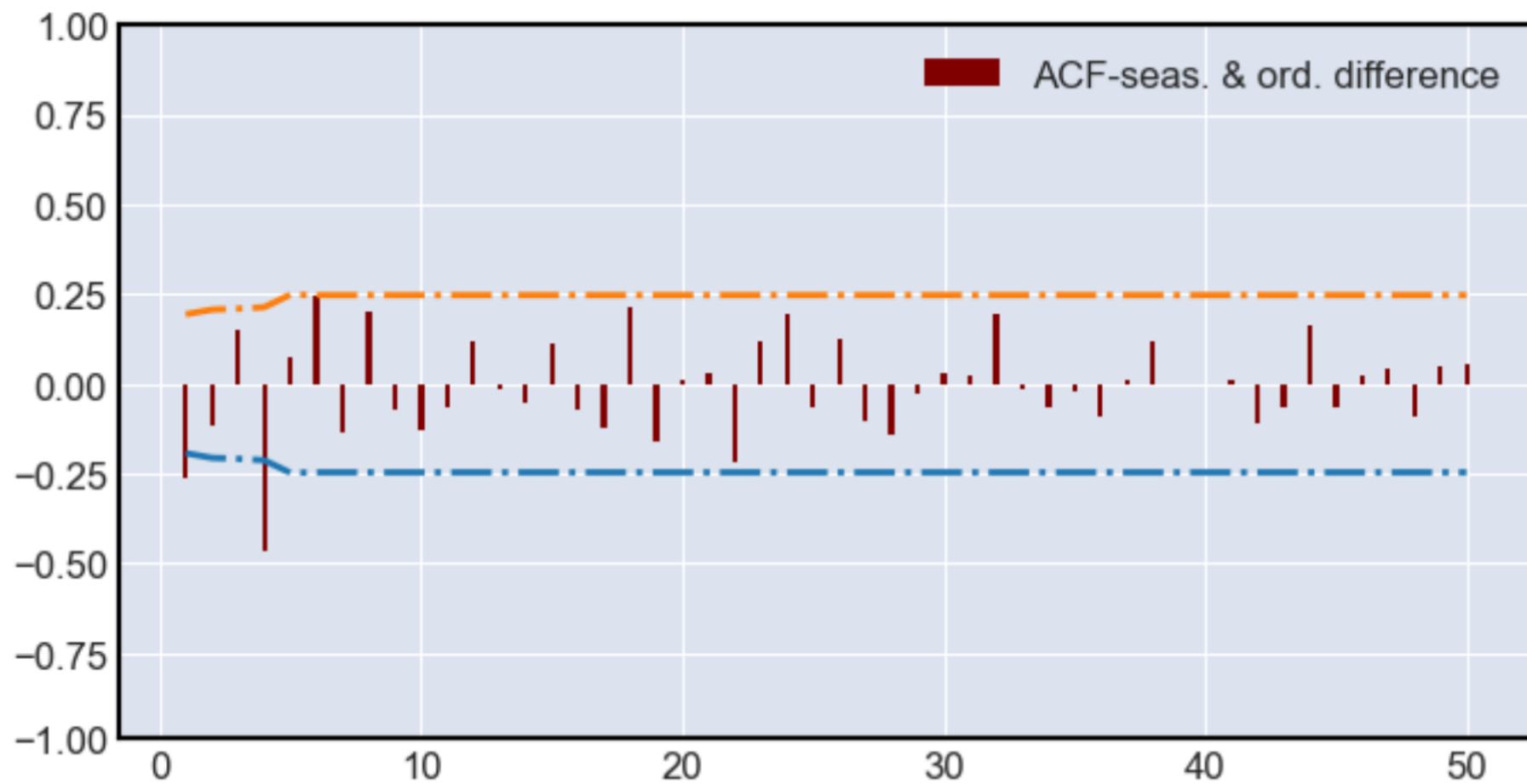
- Thành phần mùa ngẫu nhiên dễ bị bỏ qua.
- Có thể mô hình hoá bởi quá trình ARIMA.
- Để mô hình hoá, quan trọng là xác định kiểu quá trình khớp với các lược đồ ACF và PACF của dữ liệu.
- **Ví dụ:** từ lược đồ ACF của sai phân doanh số bán lẻ đồ uống
 - Tính mùa thể hiện ở các hệ số tự tương quan dương lớn tại các độ trễ mùa $4k$ ($k \geq 1$), bao quanh là các hệ số tương quan âm với độ trễ trong khoảng $[4(k - 1), 4(k + 1)]$,
 - Sự giảm chậm của các hệ số tương quan mùa thể hiện tính mùa không dừng.

Tính mùa loại này có thể loại bỏ bằng **sai phân mùa** kết hợp với sai phân thường, tức là toán tử $\nabla \nabla_4$.





ACF của $\nabla\nabla_4 y_t$





Mô hình ARIMA mùa

- Tổng quát, áp dụng toán tử ∇ và ∇_m tương ứng d và D lần:

$$\nabla^d \nabla_m^D \phi(B) y_t = \theta(B) a_t$$

Dạng thích hợp của các đa thức $\phi(B)$ và $\theta(B)$ có thể xác định bằng phương pháp lựa chọn (ví dụ tiêu chuẩn AIC, BIC).

- Có 2 khó khăn gặp phải:
 - PACF vừa khó xác định, vừa khó giải thích, nên chỉ dựa vào ACF.
 - Ít nhất 1 trong 2 đa thức có bậc tối thiểu là m , tức là số mô hình cần khảo sát rất lớn.

Phương pháp của Box-Jenkins

- Quan sát dữ liệu về doanh số đồ uống có cồn, có 2 khoảng thời gian quan trọng, tương ứng với quý và năm. Do đó, Box kì vọng có 2 mối quan hệ nảy sinh:
 - Doanh số giữa các quý liên tiếp trong năm,
 - Doanh số của cùng 1 quý trong các năm liên tiếp.

| Year | Q1 | Q2 | Q3 | Q4 |
|------|------|------|------|------|
| 1992 | 4647 | 5272 | 5544 | 6224 |
| 1993 | 4821 | 5301 | 5434 | 5982 |
| 1994 | 4767 | 5444 | 5662 | 6228 |
| 1995 | 4781 | 5369 | 5592 | 6265 |
| 1996 | 5168 | 5713 | 5877 | 6399 |





Phương pháp của Box-Jenkins

- Chẳng hạn: Q4 có liên hệ với Q4 của những năm trước đó, nên được mô hình hoá bởi:

$$\Phi(B^m)\nabla_m^D y_t = \Theta(B^m)\alpha_t$$

- Sai số tương ứng với 1 quý cố định trong những năm khác nhau là không tương quan.
- Tuy nhiên, các sai số tương ứng với các quý liên kề có thể tương quan, và được mô hình hoá bởi:

$$\phi(B)\nabla^d \alpha_t = \theta(B)a_t$$

- Kết hợp lại, ta thu được **mô hình mùa nhân tính**:

$$\phi_p(B)\Phi_P(B^m)\nabla^d\nabla_m^D y_t = \theta_q(B)\Theta_Q(B^m)a_t$$

Các chỉ số p, P, q, Q được thêm vào để xác định rõ bậc của các đa thức và quá trình được viết là $ARIMA(p, d, q)(P, D, Q)_m$.



Mô hình $ARIMA(0,1,1)(0,1,1)_m$

- Xét mô hình đơn giản: $\nabla \nabla_m y_t = (1 - \theta B)(1 - \Theta B^m) a_t$, với $|\theta|, |\Theta| < 1$
- Tự hiệp phương sai của $w_t = \nabla \nabla_m y_t$ là $\gamma_k = E(w_t, w_{t-k})$ có dạng:

$$\gamma_0 = (1 + \theta^2)(1 + \Theta^2)\sigma^2; \quad \gamma_1 = -\theta(1 + \Theta^2)\sigma^2$$

$$\gamma_{m-1} = \gamma_{m+1} = \theta\Theta\sigma^2; \quad \gamma_m = -\Theta(1 + \theta^2)\sigma^2; \quad \gamma_k = 0 \text{ với } k \text{ còn lại}$$

- Do đó, hàm tự tương quan ACF tính được là:

$$\rho_1 = -\frac{\theta}{1 + \theta^2}; \quad \rho_m = -\frac{\Theta}{1 + \Theta^2}$$
$$\rho_{m-1} = \rho_{m+1} = \rho_1 \rho_m = \frac{\theta\Theta}{(1 + \theta^2)(1 + \Theta^2)}$$



Mô hình $ARIMA(0,1,1)(0,1,1)_m$

- Với giả thiết mô hình $ARIMA(0,1,1)(0,1,1)_m$ là đúng, ước lượng của $\rho_k, k > m + 1$ có phương sai được tính bởi:

$$\text{Var}(r_k) = \frac{1 + 2(r_1^2 + r_{m-1}^2 + r_m^2 + r_{m+1}^2)}{T}$$

- Giá trị dự báo:

$$\hat{y}_{T,h} = E[y_{T+h-1} + y_{T+h-m} - y_{T+h-m-1} + a_{T+h} - \theta a_{T+h-1} - \Theta a_{T+h-m} + \theta \Theta a_{T+h-m-1} | y_T, y_{T-1}, \dots]$$

- Các trọng số ψ trong khai triển $MA(\infty)$ có dạng:

$$\begin{aligned}\psi_{rm+1} &= \psi_{rm+2} = \dots = \psi_{(r+1)m-1} = (1 - \theta)(r + 1 - r\Theta) \\ \psi_{(r+1)m} &= (1 - \theta)(r + 1 - r\Theta) + (1 - \Theta)\end{aligned}$$



Ví dụ áp dụng

Mô hình $ARIMA(0,1,1)(0,1,1)_4$ cho doanh số bán lẻ đồ uống

- Năm hệ số tự tương quan đầu tiên của $w_t = \nabla \nabla_4 y_t$ là:

$$r_1 = -0.2649, r_2 = -0.1184, r_3 = 0.1510, r_4 = -0.4646, r_5 = 0.0745$$

Ta thấy $r_2 \approx 0, r_1 r_4 = 0.1231$ nên $ARIMA(0,1,1)(0,1,1)_4$ có khả năng tương thích với dữ liệu.

- Ước lượng tham số của mô hình ta được:

$$\nabla \nabla_4 y_t = \left(1 - \frac{0.3938}{(0.1044)} B\right) \left(1 - \frac{0.4427}{(0.106)} B^4\right) a_t, \quad \hat{\sigma} = 164.87$$

(Gợi ý: sử dụng hàm **curve_fit** trong thư viện **SciPy**)



Ví dụ áp dụng

Mô hình $ARIMA(0,1,1)(0,1,1)_4$ cho doanh số bán lẻ đồ uống

- Giá trị dự báo, để đơn giản ta thay $\theta = 0.4$ và $\Theta = 0.45$:

$$\hat{y}_{T,1} = y_T + y_{T-3} - y_{T-4} - 0.4a_T - 0.45a_{T-3} + 0.33a_{T-4}$$

$$\hat{y}_{T,2} = \hat{y}_{T,1} + y_{T-2} - y_{T-3} - 0.45a_{T-2} + 0.33a_{T-3}$$

$$\hat{y}_{T,3} = \hat{y}_{T,2} + y_{T-1} - y_{T-2} - 0.45a_{T-1} + 0.33a_{T-2}$$

$$\hat{y}_{T,4} = \hat{y}_{T,3} + y_T - y_{T-1} - 0.45a_T + 0.33a_{T-1}$$

$$\hat{y}_{T,5} = \hat{y}_{T,4} + \hat{y}_{T,1} - y_T + 0.33a_T$$

$$\hat{y}_{T,h} = \hat{y}_{T,h-1} + \hat{y}_{T,h-4} - \hat{y}_{T,h-5}, \quad h > 5$$

- Các trọng số ψ :

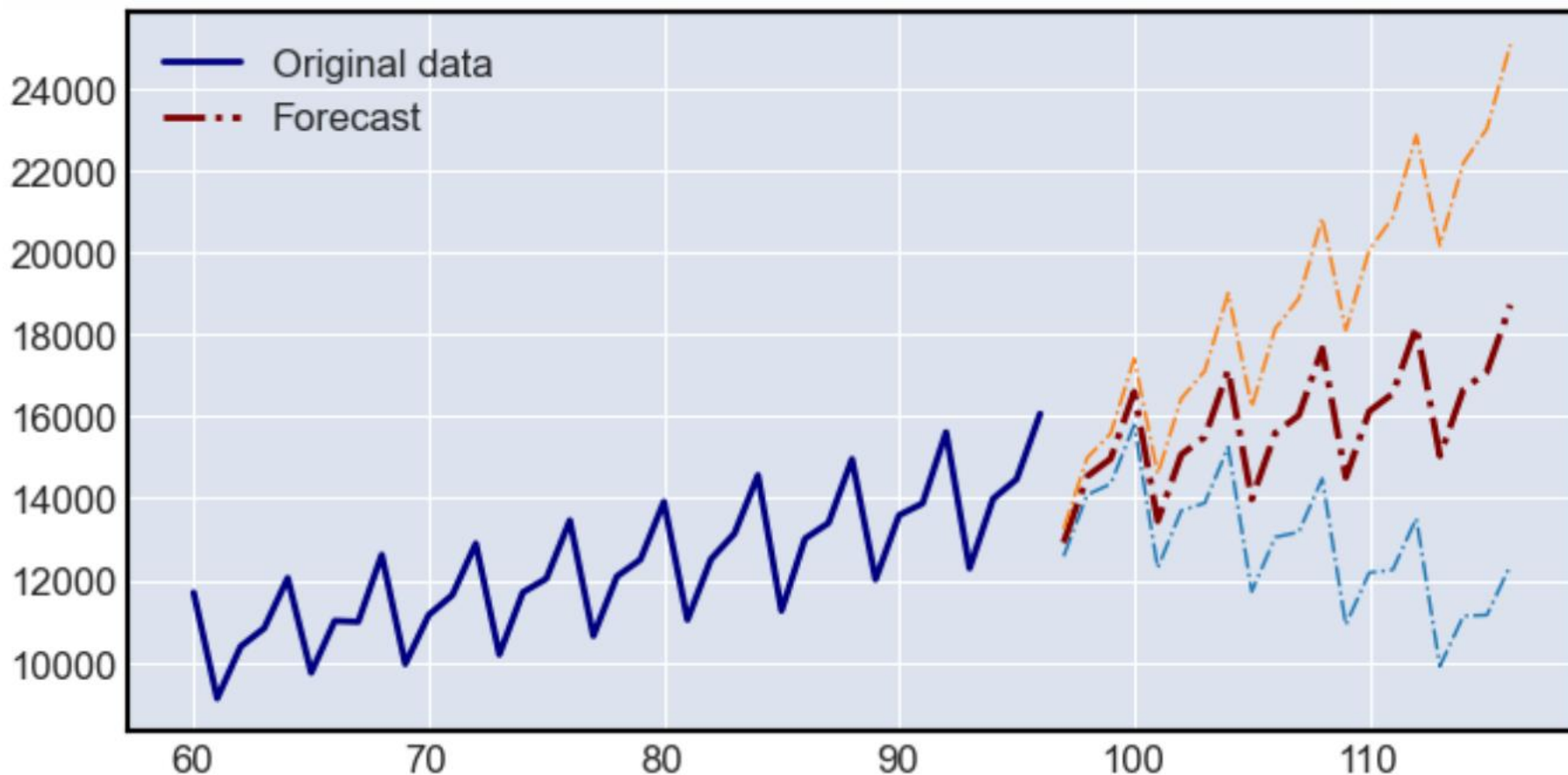
$$\psi_{4r+1} = \psi_{4r+2} = \dots = \psi_{4(r+1)-1} = 0.6 + 0.33r$$

$$\psi_{4(r+1)} = 1.15 + 0.33r$$



Ví dụ áp dụng

Mô hình $ARIMA(0,1,1)(0,1,1)_4$ cho doanh số bán lẻ đồ uống





Bài tập