

TRƯỜNG ĐẠI HỌC THỦY LỢI

PHÂN TÍCH CHUỖI THỜI GIAN VÀ CÁC KỸ THUẬT DỰ BÁO

[Tài liệu giảng dạy ở bậc đại học]

Nguyễn Thị Vinh

HÀ NỘI 2010

MỤC LỤC

1	CHƯƠNG 1: CÁC KHÁI NIỆM CHUNG VỀ DỰ BÁO	1
1.1	Bài toán dự báo	1
1.1.1	Các bài toán	1
1.1.2	Dự báo hỗ trợ quá trình ra quyết định trong các tình huống	1
1.1.3	Tiến trình dự báo chung	2
1.2	Một số khái niệm cơ bản trong dự báo	2
1.2.1	Chuỗi thời gian (Time Series)	2
1.2.2	Các phương pháp hiển thị chuỗi thời gian	3
1.2.3	Các định dạng dữ liệu	4
1.3	Tiêu chuẩn dự báo	6
1.3.1	Các đặc tính thống kê:	6
1.3.2	Các đặc tính định dạng	6
1.4	Liên hệ giữa tính toán hồi qui và dự báo chuỗi thời gian	6
1.5	BÀI TẬP CHƯƠNG 1	7
2	CHƯƠNG 2: CÁC MÔ HÌNH TRON	8
2.1	Khái niệm chung về các mô hình tron	8
2.2	Phương pháp ngây thơ (naive) - phương pháp đơn giản nhất	8
2.3	Các mô hình tron không có tính mùa (thời vụ)	9
2.3.1	Mô hình trung bình trượt đơn (Moving Average)	9
2.3.2	Mô hình trung bình trượt với trọng số dạng hàm mũ	9
2.3.3	Các mô hình xu thế	11
2.4	Các mô hình tron có yếu tố thời vụ (mùa) của Winters	17
2.4.1	Các khái niệm chung	17
2.4.2	Mô hình Winters cho dạng xu thế tuyến tính, thời vụ cộng tính	18
2.4.3	Mô hình Winters cho dạng xu thế mũ, thời vụ nhân tính	18
2.4.4	Mô hình Winters cho dạng xu thế tuyến tính, thời vụ nhân tính (dạng phổ biến nhất)	18
2.4.5	Mô hình Winters cho dạng xu thế mũ, thời vụ cộng tính	19
2.4.6	Các nhận xét chung về các mô hình Winters:	19
2.5	Các phương pháp phân ly (Decomposition)	22
2.5.1	Các công thức chung	22
2.5.2	Phương pháp phân ly cổ điển (Classical Decomposition)	23
2.5.3	Các ví dụ	23
2.6	BÀI TẬP CHƯƠNG 2	26
3	CHƯƠNG 3 : PHÂN TÍCH CHUỖI THỜI GIAN VÀ CÁC MÔ HÌNH CỦA BOX-JENKINS	28
3.1	Các mô hình chuỗi thời gian ARMA (AutoRegressive-Moving Average)	28
3.1.1	Mô hình tự hồi quy bậc p - AR(p)	28
3.1.2	Mô hình trung bình trượt bậc q - MA(q)	29
3.1.3	Mô hình hỗn hợp tự hồi quy-trung bình trượt bậc (p,q) - ARMA(p,q)	29

3.2	Các điều kiện cần về tính dừng và tính khả nghịch.....	29
3.2.1	Điều kiện dừng	29
3.2.2	Điều kiện khả nghịch.....	30
3.3	Các trợ giúp cho việc phân tích chuỗi thời gian.....	31
3.3.1	Biểu diễn đồ họa chuỗi thời gian.....	31
3.3.2	Hệ số tự tương quan ACF (Auto Correlation Function)	31
3.3.3	Hàm tự tương quan riêng phần PACF	33
3.3.4	Thông kê Q của Box-Pierce.....	36
3.4	Các ứng dụng của các hệ số tự tương quan.....	37
3.4.1	Kiểm tra tính ngẫu nhiên của dữ liệu và phần dư.....	37
3.4.2	Xác định tính dừng của chuỗi thời gian	37
3.4.3	Loại bỏ tính không dừng của chuỗi thời gian.....	39
3.4.4	Nhận biết tính thời vụ trong chuỗi thời gian	40
3.5	Các mô hình ARIMA	43
3.5.1	Các mô hình ARIMA không có tính thời vụ.....	43
3.5.2	Các mô hình ARIMA có tính thời vụ.....	46
3.6	BÀI TẬP CHƯƠNG 3	53
4	CHƯƠNG 4: CÁC PHƯƠNG PHÁP DỰ BÁO CỦA BOX-JENKINS..	55
4.1	Các khâu chính trong phương pháp Box-Jenkins	55
4.2	Các nguyên tắc lựa chọn mô hình ARIMA(p,d,q) phù hợp	56
4.3	Các hàm dự báo của các mô hình ARMA(p,q)	58
4.3.1	Một số mô hình ARMA thường gặp:	59
4.3.2	Giới hạn cho phép của các dự báo.....	60
4.4	Các ví dụ minh họa.....	60
4.5	BÀI TẬP CHƯƠNG 4	64
5	PHỤ LỤC: GIỚI THIỆU PHẦN MỀM DỰ BÁO SIBYL.....	65
5.1	Môi trường làm việc của Sibyl.....	65
5.2	Một số phương pháp dự báo trong Sibyl.....	66
5.2.1	Các phương pháp trung bình trượt	66
5.2.2	Các phương pháp hồi quy tìm đường cong phù hợp với chuỗi dữ liệu (Trend-Cycle Regression Curve-Fitting Methods)	66
5.2.3	Các phương pháp làm trơn dạng mũ	67
5.2.4	Các phương pháp phân ly	68
5.2.5	Phương pháp Box-Jenkins.....	69

1 CHƯƠNG 1: CÁC KHÁI NIỆM CHUNG VỀ DỰ BÁO

Dự báo là quá trình tạo ra các nhận định về các hiện tượng mà thông thường các đầu ra của chúng còn chưa quan sát được.

<http://en.wikipedia.org/wiki/Forecast>

1.1 Bài toán dự báo

1.1.1 Các bài toán

Dự báo là một trong những yếu tố quan trọng nhất trong việc ra các quyết định quản lý bởi vì ảnh hưởng sau cùng của một quyết định thường phụ thuộc vào sự tác động của các nhân tố không thể nhìn thấy tại thời điểm ra quyết định. Vai trò của dự báo là nhạy cảm trong các lĩnh vực như tài chính, nghiên cứu thị trường, lập kế hoạch sản xuất, hành chính công, điều khiển quá trình sản xuất hay nghiên cứu, ...

Trong giới doanh nhân, các câu hỏi thường xuyên được đưa ra là:

Lượng hàng sẽ bán trong tháng tới là bao nhiêu?

Tháng này nên đặt mua bao nhiêu hàng?

Nên giữ bao nhiêu cổ phiếu ?

Nên mua bao nhiêu nguyên liệu ?

Mục tiêu bán hàng sắp tới là gì?

Có nên tăng nhân công không?

1.1.2 Dự báo hỗ trợ quá trình ra quyết định trong các tình huống

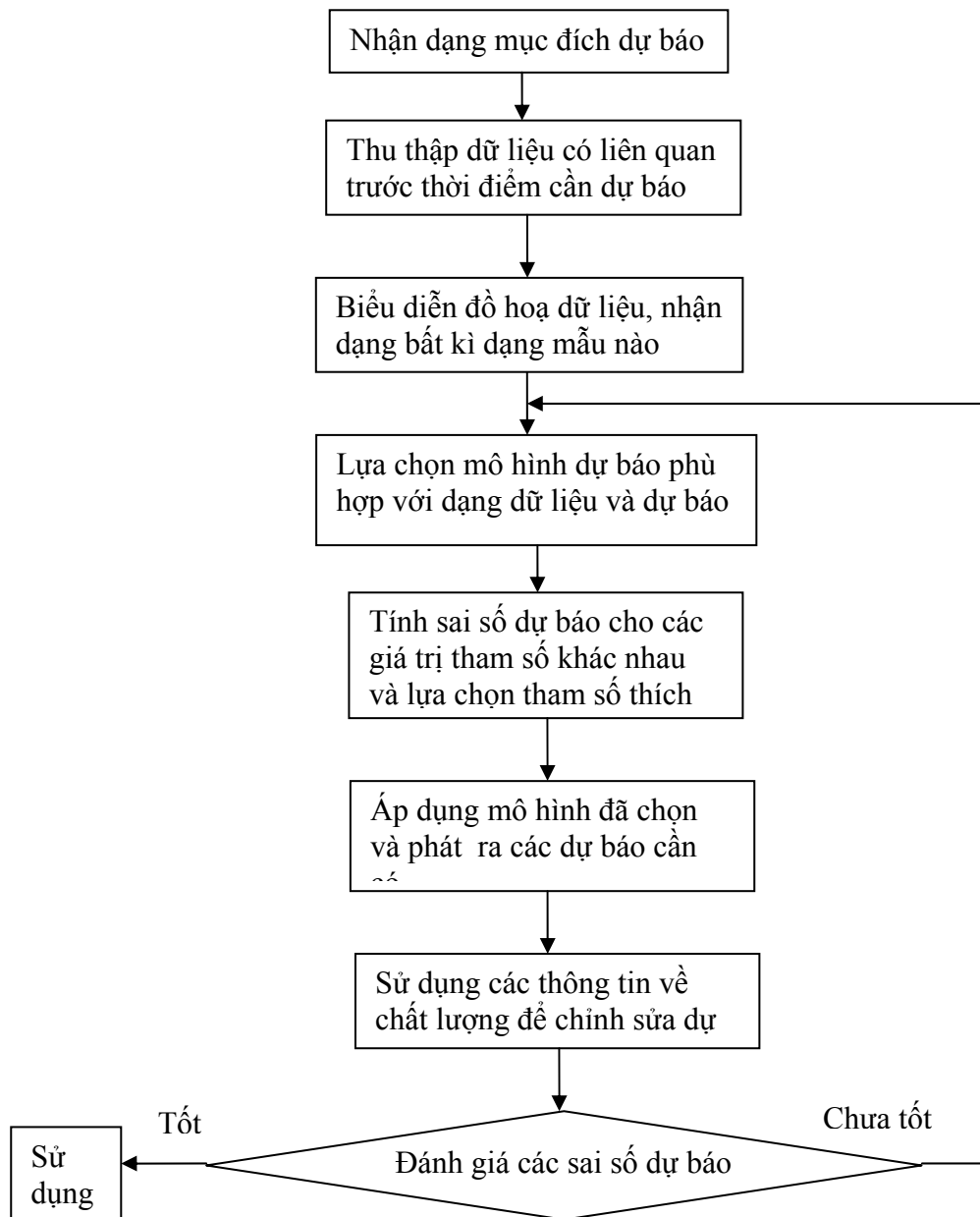
i> Điều tiết nguồn tài nguyên sẵn có: Dự báo nhu cầu cho sản phẩm, nguyên liệu, nhân công, tài chính hay dịch vụ như là một đầu vào thiết yếu để điều tiết kế hoạch sản xuất, vận tải, tiền vốn và nhân lực.

ii> Yêu cầu thêm tài nguyên: Dự báo giúp xác định tài nguyên cần có trong tương lai (như nhân lực, máy móc thiết bị, vốn ...)

iii> Thiết kế, lập quy hoạch: Dự báo các hiện tượng thiên nhiên như lũ lụt, hạn hán để thiết kế các công trình như đê, đập, hồ chứa và quy hoạch vùng sản xuất.

Nhược điểm của dự báo là không thể tránh khỏi sai số. Trên quan điểm thực tiễn, cần hiểu rõ cả mặt mạnh lẫn mặt hạn chế của các phương pháp dự báo và tính đến chúng trong khi sử dụng dự báo.

1.1.3 Tiến trình dự báo chung



1.2 Một số khái niệm cơ bản trong dự báo

1.2.1 Chuỗi thời gian (Time Series)

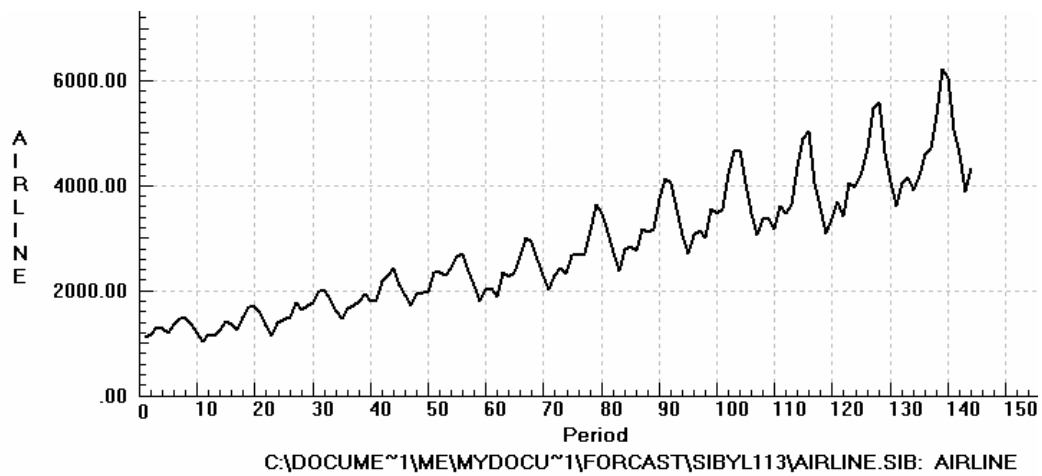
Chuỗi thời gian là một dãy dữ liệu được quan sát ở các thời điểm kế tiếp nhau với cùng một đơn vị đo mẫu.

Trong chuỗi thời gian, trình tự thời gian đóng một vai trò thực sự quan trọng, vì vậy các tính toán thống kê thông thường như trung bình mẫu, độ lệch quân phương mẫu, khoảng tin cậy, kiểm định các giả thuyết, ... không còn thích hợp

Một chuỗi thời gian thường bao gồm những thành phần sau đây

- i>. Thành phần ổn định
- ii>. Thành phần xu thế

- iii> Thành phần mùa (thời vụ)
- iv> Thành phần ngẫu nhiên
- v> Thành phần chu kì (dài hạn)



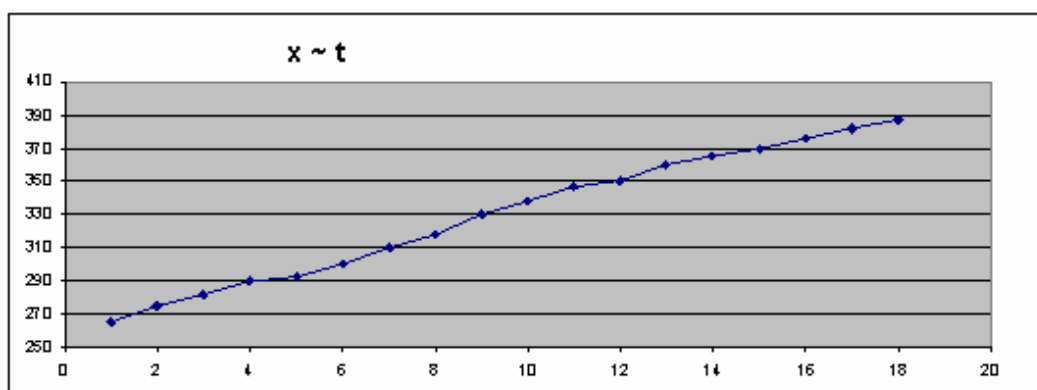
1.2.2 Các phương pháp hiển thị chuỗi thời gian

Phân tích chuỗi thời gian bao gồm việc nghiên cứu dạng dữ liệu trong quá khứ và giải thích các đặc điểm chính của nó. Một trong các phương pháp đơn giản và hiệu quả nhất là hiển thị trực quan chuỗi đó. Các đặc điểm không dễ thấy trong bảng dữ liệu thường nổi lên qua các minh họa đồ thị.

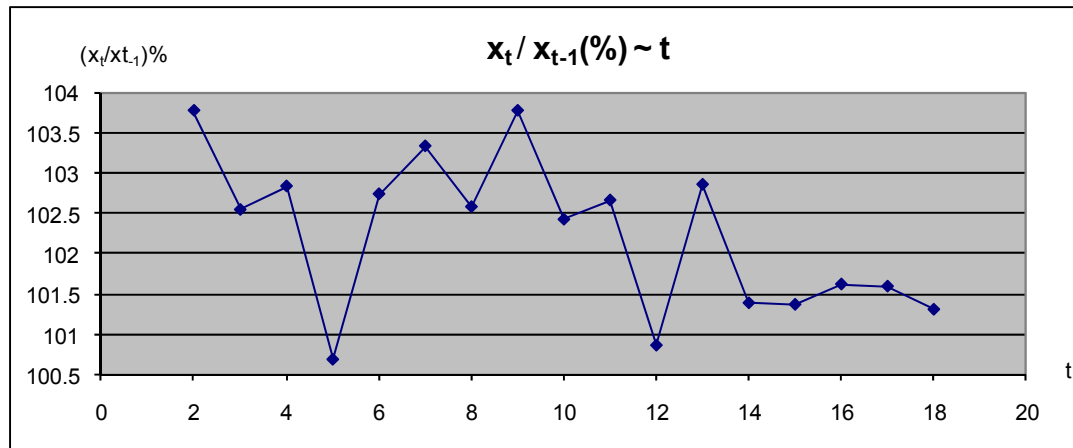
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
x_t	265	275	282	290	292	300	310	318	330	338	347	350	360	365	370	376	382	387
x_t/x_{t-1}		104	103	103	101	103	103	103	104	102	103	101	103	101	101	102	102	101
$x_t - x_{t-1}$		10	7	8	2	8	10	8	12	8	9	3	10	5	5	6	6	5

Ba loại đồ thị minh họa chuỗi thời gian là

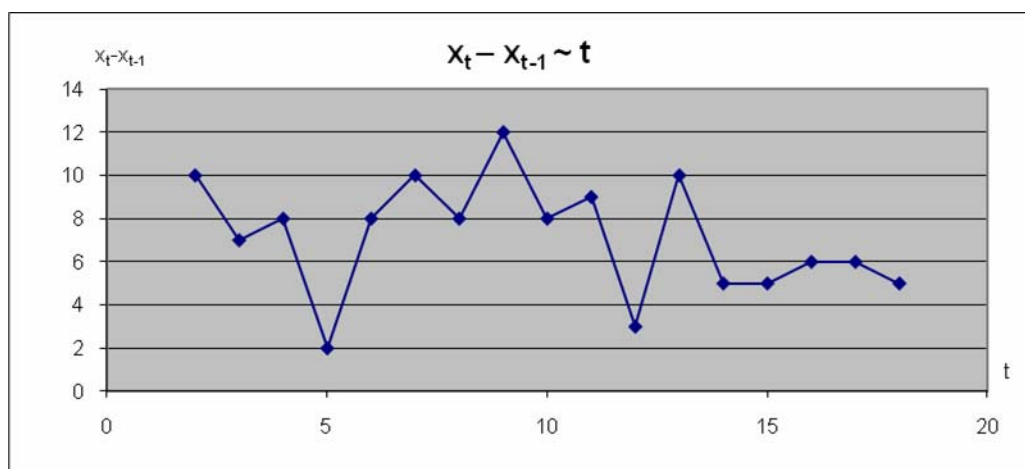
- i> Đồ thị của x_t theo t: cung cấp lịch sử dữ liệu gốc chưa bị chuyển đổi qua bất cứ phép biến đổi nào, giúp cho việc nghiên cứu xu thế và nhận dạng.



ii> Đồ thị của $x_t / x_{t-1} \times 100$ theo t : mỗi điểm trên đồ thị này cho biết giá trị hiện thời của chuỗi tăng hay giảm so với giá trị trước đó. Ví dụ giá trị tại thời điểm $t = 2$ là 102,9% chỉ ra rằng chuỗi đã tăng 2,9% từ thời điểm $t = 2$ sang thời điểm $t = 3$. Nếu mọi giá trị đều lớn hơn 100% nhưng theo xu thế giảm dần thì đồ thị đó chứng tỏ rằng chuỗi này có xu thế tăng nhưng tỉ lệ tăng lại giảm dần.



iii> Đồ thị của $x_t - x_{t-1}$ theo t : Đồ thị này biểu diễn sự thay đổi giữa các bước thời gian kế tiếp nhau. Nhìn vào đồ thị ta thấy được khoảng các giá trị biến đổi giữa các bước kế nhau.



Ví dụ, từ bảng các giá trị $x_t \sim t$ ở trang trước, người ta vẽ được 3 đồ thị tương ứng ở các phần i>, ii>, iii>.

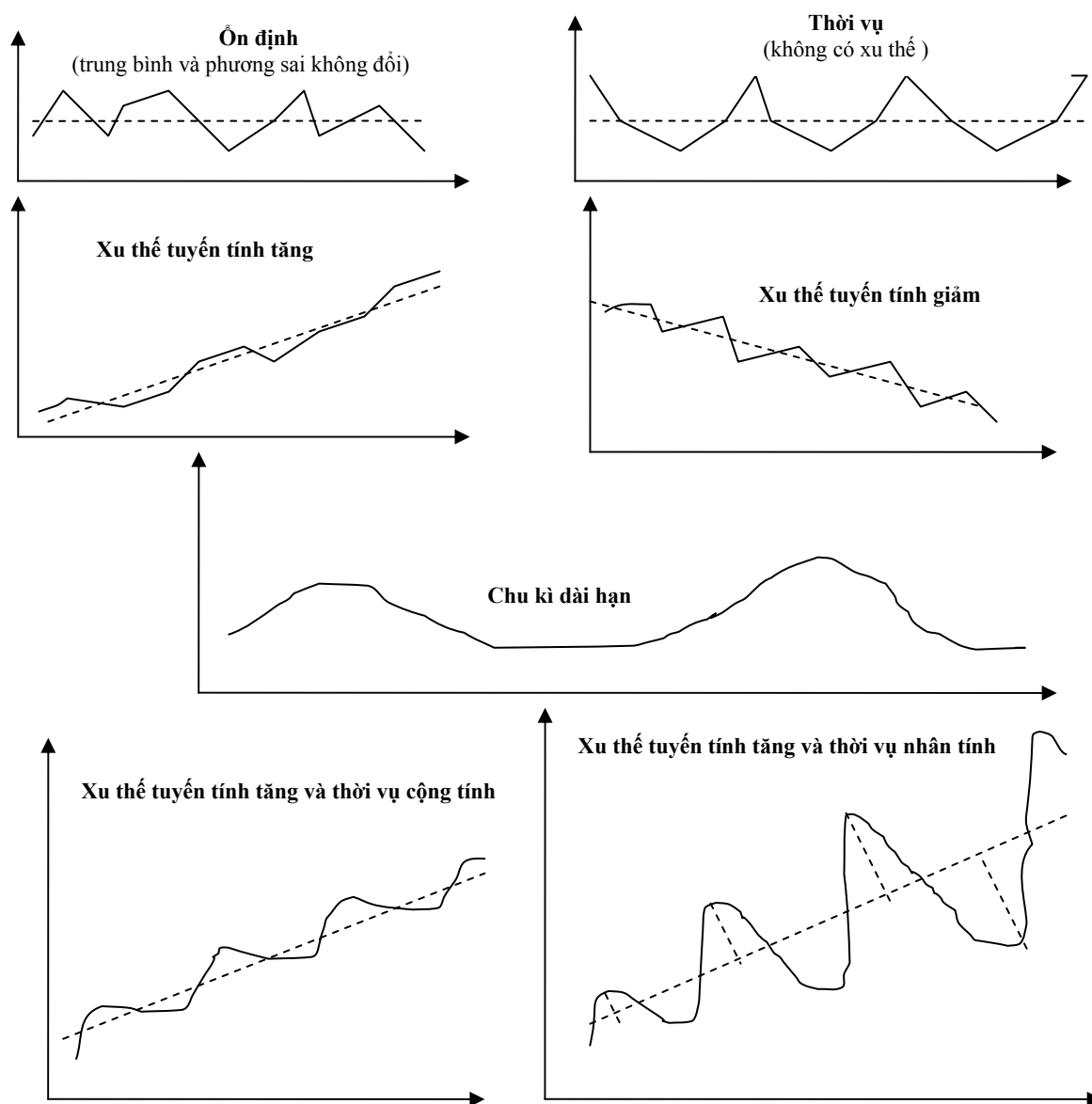
1.2.3 Các định dạng dữ liệu

Trước khi áp dụng bất cứ một phương pháp dự báo khoa học cho một tình huống nào, cần phải ghép nối các thông tin (dữ liệu có liên quan) về tình huống đó càng nhiều càng tốt. Những dữ liệu đó được phân thành 2 loại:

- i> Các dữ liệu bên trong, ví dụ số liệu sản phẩm bán ra trong quá khứ, ...
- ii> Các dữ liệu bên ngoài, ví dụ như các thống kê của ngân hàng về tình hình tài chính của công ty (phản ánh thông tin bên trong).

Từ các thông tin này, người làm dự báo phải chọn ra thông tin liên quan nhiều nhất đến tình huống cần dự báo. Chẳng hạn, trong dự báo bán hàng, báo cáo hàng bán được trong quá khứ của công ty sẽ cung cấp những thông tin tối thiểu cho việc dự báo. Thông tin tối thiểu cần thỏa mãn các yêu cầu về:

- Tính liên quan: Nó có phải là thông tin liên quan trực tiếp nhất không?
- Độ tin cậy: Dữ liệu được thu thập như thế nào? Có đáng tin cậy không?
- Tính thời sự: Liệu các thông tin mới nhất đã được cập nhật chưa? Chúng có sẵn khi cần không?



Một số định dạng dữ liệu

Khi đã có những thông tin tối thiểu cần thiết, ta cần phải nghiên cứu đặc điểm của nó bằng cách minh họa đồ thị. Dạng dữ liệu quá khứ là rất quan trọng vì nó quyết định

việc lựa chọn mô hình dự báo. Mô hình dự báo được chọn phải tương thích với dạng dữ liệu mẫu trong quá khứ.

1.3 Tiêu chuẩn dự báo

Các tiêu chuẩn chung đánh giá sự thành công của một mô hình dự báo khi áp dụng vào một tập dữ liệu là:

- i> Trùng càng nhiều với các thay đổi ngẫu nhiên trong dữ liệu càng tốt.
- ii> Không vượt quá xa bất kì một đặc tính nào của dữ liệu

Xét về mặt sai số, hai loại đặc tính cần quan tâm khi thử nghiệm một công thức dự báo trên dữ liệu là

1.3.1 Các đặc tính thống kê:

Một phương pháp dự báo tốt thường cho sai số trung bình nhỏ. Trong các mô hình dự báo, người ta thường sử dụng các loại sai số như

$$MAE = \frac{1}{n} \sum |e_i| \quad (\text{Mean Absolute Error})$$

$$MSE = \frac{1}{n} \sum e_i^2 \quad (\text{Mean Square Error})$$

$$RMSE = \sqrt{MSE} \quad (\text{squareRoot Mean Square Error})$$

ở đây sai số $e_i = x_i - f_i$ với f_i là dự báo của x_i

1.3.2 Các đặc tính định dạng

Trong các mô hình dự báo, sự có mặt của các *dạng sai số* (như tính lệch, tính chu kì, tính kiên định, ...) đều bị xem là dấu hiệu không tốt. Sự xuất hiện của bất cứ xu thế nào trong sai số cũng nên khử càng nhanh càng tốt. Có thể sai phân hóa chuỗi các giá trị ban đầu để đối phó với các tác động này

Tóm lại có hai tiêu chuẩn dự báo về định lượng và định tính là: sai số nhỏ và không tuân theo một định dạng nào.

1.4 Liên hệ giữa tính toán hồi qui và dự báo chuỗi thời gian

Tính toán hồi qui dựa trên quan hệ nhân – quả của hệ thống và cực tiểu sai số bằng phương pháp bình phương bé nhất

Dự báo chuỗi thời gian dựa trên quan hệ nội tại của dữ liệu để phát ra các dự báo cho các bước thời gian tiếp theo.

1.5 BÀI TẬP CHƯƠNG 1

1. Trong các định dạng có thể có của chuỗi thời gian, những định dạng nào có tính loại trừ nhau?
2. Giải thích tại sao một kết quả dự báo có sai số không ngẫu nhiên, tức là tuân theo một định dạng nào đó, là một dự báo không tốt?

2 CHƯƠNG 2: CÁC MÔ HÌNH TRON

2.1 Khái niệm chung về các mô hình tron

Cơ sở của các phương pháp này là làm tron (lấy trung bình hoặc trung bình có trọng số) các quan sát trong quá khứ của chuỗi thời gian để nhận được dự báo cho tương lai. Trong việc làm tron các giá trị quá khứ, các sai số ngẫu nhiên được tính trung bình. Các mô hình tron dùng trong dự báo thích hợp cho một số tình huống.

Các ưu điểm chính của các phương pháp làm tron là:

- i> Chi phí thấp
- ii> Dễ dùng (ở những nơi có thể áp dụng được)
- iii> Tốc độ tính nhanh (ở những nơi chấp nhận được)

Những phương pháp làm tron rất hấp dẫn khi cần phải dự báo ở rất nhiều bước thời gian tương lai, chẳng hạn trong công tác kiểm kê.

2.2 Phương pháp ngây thơ (naive) - phương pháp đơn giản nhất:

Giả sử người quản lý siêu thị muốn biết một khách hàng điển hình tiêu bao nhiêu tiền cho một lần mua sắm. Lấy ngẫu nhiên một mẫu 12 khách hàng và nhận được kết quả sau:

Khách hàng	1	2	3	4	5	6	7	8	9	10	11	12
Số tiền đã tiêu (\$)	19	18	19	22	21	17	23	19	19	22	21	20

Có thể lấy giá trị làm cực tiểu sai số MSE, trong trường hợp này là trung bình mẫu $\bar{X} = \$20$.

➤ Các ưu điểm của phương pháp trung bình

- i> Cực tiểu sai số
- ii> Ước lượng không chệch
- iii> Cho dự báo tốt nếu dữ liệu có tính *ổn định* (trung bình không đổi) và tính ngẫu nhiên (không có xu thế tăng /giảm, không có tính thời vụ hay chu kỳ).

➤ Các nhược điểm của phương trung bình

- i> Cho kết quả tồi nếu dữ liệu có tính xu thế hoặc có định dạng xác định
- ii> Cần mẫu có dung lượng lớn
- iii> Dự báo tồi nếu có đột biến

Kết luận: chỉ sử dụng phương pháp ngây thơ khi chuỗi thời gian có tính ổn định, ngẫu nhiên, và khi không biết phương pháp dự báo nào khác.

2.3 Các mô hình trơn không có tính mùa (thời vụ)

2.3.1 Mô hình trung bình trượt đơn (Moving Average)

- **Phương pháp:** Lấy trung bình N giá trị liên tiếp của các quan sát gần nhất làm dự báo cho thời điểm thứ N+1. Thuật ngữ trung bình trượt có nghĩa là quan sát cũ nhất sẽ bị loại đi mỗi khi có quan sát mới. Nói cách khác, số quan sát trong khi tính là không đổi và chỉ bao gồm các quan sát gần với hiện tại nhất.

- **Lập công thức:**

$$f_{t+1} = (x_t + x_{t-1} + \dots + x_{t-N+1}) / N \quad (2.1)$$

$$= (x_{t-1} + x_{t-2} + \dots + x_{t-N}) / N + x_t / N - x_{t-N} / N$$

$$\text{hay } f_{t+1} = f_t + x_t / N - x_{t-N} / N \quad (2.2)$$

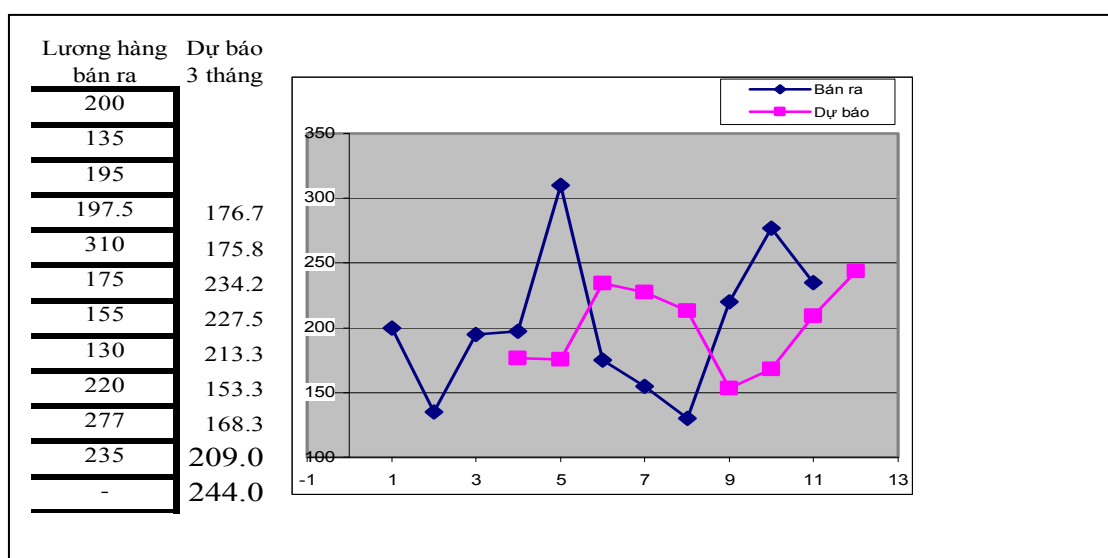
- **Nhận xét:**

i> Dự báo ở thời điểm t+1 chỉ là điều chỉnh của dự báo ở thời điểm t trước đó. Khi N tăng đủ lớn thì lượng điều chỉnh $x_t / N - x_{t-N} / N \rightarrow 0$ và trung bình trượt trở thành trung bình mẫu như phương pháp ngây thơ, độ chính xác thấp.

ii> Chỉ nên áp dụng phương pháp này khi số giá trị quan sát được là ít và tập dữ liệu có tính *ổn định* theo thời gian

- **Ví dụ:**

Bảng dưới đây cho biết lượng hàng bán ra của các tháng 1, 2, ..., 11. Nếu sử dụng mô hình trung bình trượt MA với N = 1 ta coi lượng hàng bán ra của tháng trước là dự báo cho tháng sau; với N = 11 ta sử dụng trung bình mẫu cho dự báo của tháng 12; với N = 3 ta sử dụng trung bình của 3 tháng gần nhất làm dự báo cho tháng tới



2.3.2 Mô hình trung bình trượt với trọng số dạng hàm mũ (Exponentially Weighted Moving Averages) hay mô hình trơn dạng mũ đơn

- **Phương pháp:**

Hai hạn chế của mô hình MA là:

i> N giá trị quá khứ bắt buộc phải có đủ

ii> Trọng số trung bình cho các quan sát là như nhau ($1/N$)

Trên thực tế, các quan sát càng gần càng chứa nhiều thông tin cho các giá trị sắp xảy ra, do đó cần cho chúng các trọng số lớn hơn so với các quan sát ở xa

➤ **Lập công thức:**

Giả sử chuỗi dữ liệu quan sát được là *ổn định (có trung bình không đổi)* và không có quan sát thứ $N-t$. Khi đó từ công thức (2,2) lấy f_t thay cho x_{N-t} ta được

$$f_{t+1} = f_t + x_t / N - f_t / N = (1-1/N) f_t + x_t / N, \text{ vì } N > 0 \text{ nên } 0 < 1/N < 1.$$

$$\text{Đặt } w = 1/N \text{ ta có } f_{t+1} = (1-w) f_t + w x_t \quad (2.3)$$

Thuật ngữ *dạng hàm mũ* xuất phát từ việc biến đổi công thức (2.3):

$$\begin{aligned} F_{t+1} &= w x_t + (1-w) f_t = w x_t + (1-w) [w x_{t-1} + (1-w) f_{t-1}] = \dots \\ &= w x_t + w (1-w) x_{t-1} + w (1-w)^2 x_{t-2} + \dots \end{aligned}$$

→ các trọng số áp dụng cho mỗi giá trị quá khứ giảm dần theo luật hàm mũ

➤ **Nhận xét:**

i> Từ công thức dự báo (2.3)

$$\rightarrow f_{t+1} = w x_t + (1-w) f_t = f_t + w (x_t - f_t)$$

hay

$$f_{t+1} = f_t + w e_t$$

(dự báo mới bằng tổng của dự báo cũ và điều chỉnh sai số). Đây chính là nguyên tắc phản hồi hay phương pháp thích ứng của dự báo.

ii> Một số trường hợp riêng:

$$w = 0: f_{t+1} = f_t$$

$$w = 1: f_{t+1} = x_t$$

$w \approx 1$: cho các dự báo phản ánh các thay đổi gần đây nhất

$$w = 0,1: f_{t+1} = 0,1x_t + 0,09x_{t-1} + 0,081x_{t-2} + \dots \text{ cho các dự báo xấp xỉ nhau}$$

$$w = 0,9: f_{t+1} = 0,9x_t + 0,09x_{t-1} + 0,009x_{t-2} + \dots \text{ dự báo bám theo mẫu 1 bước}$$

iii> Chú ý rằng việc phản hồi sự biến đổi của mẫu được cải thiện khi w gần 1. Tuy nhiên việc phản hồi được thực hiện nhanh hay chậm còn tùy vào khả năng làm trơn các dao động ngẫu nhiên.

iv> Các ưu điểm của phương pháp EWMA là không cần biết nhiều số liệu quá khứ và tính toán đơn giản; dữ liệu càng gần càng có trọng số lớn; *thích hợp khi phải dự báo cho nhiều bước thời gian* (khi đó w thường là 0,2 hoặc 0,3)

➤ **Một số vấn đề nảy sinh và cách khắc phục:**

i> Thời điểm đầu tiên $t=1$: không có dự báo cho thời điểm trước đó để tính dự báo f_1 theo công thức (2.3). Các giải pháp là lấy $f_1 = x_1$ hoặc $f_0 = \bar{x}$ hoặc sử dụng trung bình cộng của vài giá trị đầu làm giá trị f_0 ;

ii> Chọn giá trị w theo một trong ba tiêu chí sau

- w là tốt nhất cho mô hình theo nghĩa sai số MSE là nhỏ nhất. Giá trị này phải được tính thử cho các giá trị w khác nhau để lựa chọn. Trong ví dụ trên

$$w = 0,1 \quad \text{MSE} = 3438,3$$

$$w = 0,5 \quad \text{MSE} = 4347,2$$

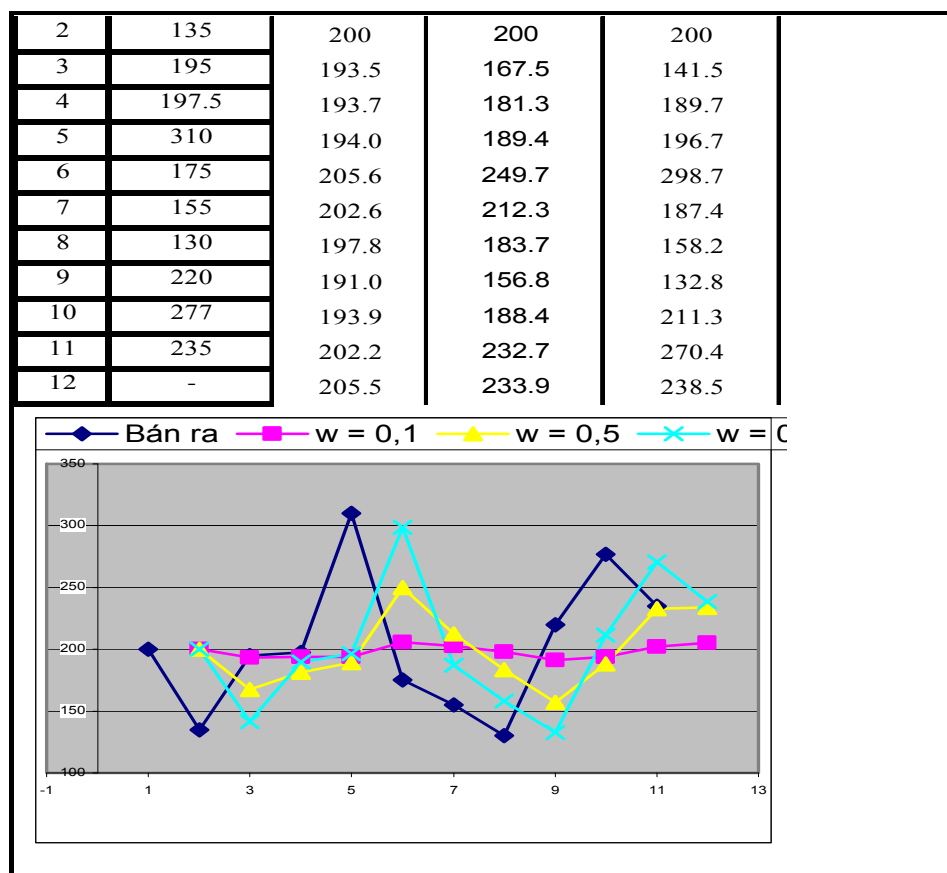
$$w = 0,9 \quad \text{MSE} = 5039,4$$

Trong ví dụ này, MSE giảm khi w giảm, chứng tỏ dữ liệu là ngẫu nhiên.

- Ở một số bước đầu nên chọn w gần 1 vì không có f_0 để tính toán. Có thể tiến hành chọn các w_t lớn hơn giá trị tối ưu. Ví dụ khi $w = 0,2$ là tối ưu thì nên chọn $w_t = 1/t$ cho đến khi $w_t < 0,2$. Vậy

$$w_1 = 1,0 \quad w_2 = 0,5 \quad w_3 = 0,33 \quad w_4 = 0,25 \quad w = 0,2 \text{ với } t \geq 5$$

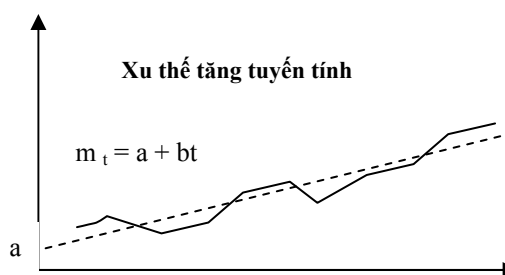
- Định ra các giá trị dường như là tốt nhất cho mỗi tình huống cụ thể, chẳng hạn $w=1$ khi $t = 1$; $w = 0,3$ khi $t = 2, 3, 4$ và $w = 0,3$ khi $t \geq 5$



2.3.3 Các mô hình xu thế

- **Đặt vấn đề:** Việc áp dụng các mô hình trung bình trượt cho tập dữ liệu chứa xu thế (tăng hoặc giảm) sẽ cho những dự báo thiên nhỏ hoặc thiên lớn so với giá trị thực. Giả sử có N quan sát x_t , $t = 1, \dots, N$ theo *xu thế tăng tuyến tính* như hình vẽ. Ta gọi mức tăng của mẫu tại thời điểm t là

$m_t = a + bt$
 trong đó
 $a = \text{mức tăng tại } t = 0$
 $b = \text{độ dốc}$



Các dự báo được tạo ra tại gốc $t = N$ sẽ là

$$f_{N+\tau} = a + b(N + \tau), \tau = 1, 2, \dots$$

hay

$$f_{N+\tau} = m_N + b \tau$$

Vai trò của các mô hình xu thế là **ước lượng m_N và b** từ các dữ liệu quá khứ. Kí hiệu các ước lượng đó là m_N^* và b^* ta có

$$f_{N+\tau} = m_N^* + b^* \tau \quad (2.4)$$

Các cách ước lượng khác nhau cho ta các mô hình tuyến tính khác nhau

➤ **Mô hình bình phương bé nhất (Least Mean Square)**

Việc cực tiểu bình phương tổng các sai số

$$S = \sum_{t=1}^N [x_t - (a + bt)]^2$$

dẫn đến các giá trị

$$a^* = \bar{x} - b^* \bar{t}$$

$$b^* = \frac{N \sum_{t=1}^N tx_t - \sum_{t=1}^N t \sum_{t=1}^N x_t}{N \sum_{t=1}^N t^2 - (\sum_{t=1}^N t)^2} \quad (2.5)$$

ta có công thức (2.4) với $m_N^* = a^* + b^* N$ cho bởi (2.5)

➤ **Ví dụ:** Cho chuỗi quan sát

t	1	2	3	4	5	6	7	8	9	10
x_t	60	70	85	60	99	68	106	75	66	124

$$\rightarrow f_{10+\tau} = 102.81 + 4.58 \tau \rightarrow f_{11} = 107,4; f_{12} = 112$$

Nhận xét: Công thức (2.5) sử dụng trọng số bình quân để tính m_N và b

➤ **Mô hình trung bình trượt kép (DMA)**

Lập công thức

Mô hình này là sự mở rộng của mô hình MA bằng cách sử dụng các số hạng bám theo xu thế của mẫu. Hai giá trị trung bình trượt được tính tại thời điểm T là

$$\begin{cases} M_T = \frac{x_T + x_{T-1} + \dots + x_{T-N+1}}{N} = \frac{1}{N} \sum_{i=0}^{N-1} x_{T-i} \\ M_T^{(2)} = \frac{M_T + M_{T-1} + \dots + M_{T-N+1}}{N} = \frac{1}{N} \sum_{i=0}^{N-1} M_{T-i} \end{cases}$$

N là số bước thời gian được chọn để lấy trung bình trượt M_T

Ta có công thức tương đương

$$\begin{cases} M_T = M_{T-1} + \frac{x_T - x_{T-N}}{N} \\ M_T^{(2)} = M_{T-1}^{(2)} + \frac{M_T - M_{T-N}}{N} \end{cases}$$

Công thức tính các dự báo tại thời điểm $t = n$ cho τ bước phía trước:

$$\boxed{f_{N+\tau} = m_N^* + b^* \tau} \text{ với } \begin{cases} m_N^* = 2M_N - M_N^{(2)}, \\ b^* = \frac{2}{N-1}(M_N - M_N^{(2)}) \end{cases} \quad (2.6)$$

Chứng minh công thức (2.6)

Ở thời điểm T, mức tăng $m_T = a + bT$

Ở thời điểm T-1, mức tăng $m_{T-1} = a + b(T-1) = m_T - b$

Ở thời điểm T-2, mức tăng $m_{T-2} = a + b(T-2) = m_T - 2b$

.....

Vậy kì vọng của M_T là

$$\begin{aligned} E(M_T) &= \frac{1}{N} \sum_{i=0}^{N-1} E(x_{T-i}) = \frac{1}{N} \sum_{i=0}^{N-1} m_{T-i} = \frac{1}{N} \sum_{i=0}^{N-1} (m_T - ib) = \frac{1}{N} (Nm_T - b \sum_{i=0}^{N-1} i) \\ &= \frac{1}{N} (Nm_T - b \frac{N(N-1)}{2}) = m_T - \frac{N-1}{2} b \end{aligned} \quad (2.7)$$

$$\begin{aligned} \rightarrow E(M_T^{(2)}) &= \frac{1}{N} E(M_T + M_{T-1} + \dots + M_{T-N+1}) \\ &= \left[m_T - \frac{N-1}{2} b + m_{T-1} - \frac{N-1}{2} b + \dots + m_{T-N+1} - \frac{N-1}{2} b \right] \\ &= \frac{1}{N} [m_T + m_{T-1} + \dots + m_{T-N+1}] - \frac{N-1}{2} b \\ &= \frac{1}{N} [m_T + (m_T - b) + (m_T - 2b) + \dots + (m_T - (N-1)b)] - \frac{N-1}{2} b \\ &= m_T - \frac{N-1}{2} b - \frac{N-1}{2} b = m_T - (N-1)b \end{aligned} \quad (2.8)$$

Sử dụng phương pháp ước lượng các moment ta nhận được

$$\begin{cases} M_T = m_T^* - \frac{N-1}{2} b^* \\ M_T^{(2)} = m_T^* - (N-1)b^* \end{cases}$$

Giải hệ 2 phương trình đại số tuyến tính 2 ẩn m_T^* và b^* ta nhận được (2.6)

Ví dụ: Dùng trung bình trượt kép với

$N = 6$ tính các dự báo với $\tau = 1$ và $\tau = 2$

$$\begin{aligned} m_{14}^* &= 2M_{14} - M_{14}^{(2)} \\ &= 209,34 - 94,89 = 114,45 \end{aligned}$$

$$\begin{aligned} b^* &= \frac{2}{5}(M_{14} - M_{14}^{(2)}) \\ &= 2(104,67 - 94,89) / 5 = 3,91 \end{aligned}$$

Vậy $f_{14+\tau} = 114,45 + 3,91 \tau$

$\tau = 1$ $f_{15} = 118,36$; $\tau = 2$ $f_{16} = 122,27$

➤ Mô hình trơn dạng mũ kép (DEWMA)

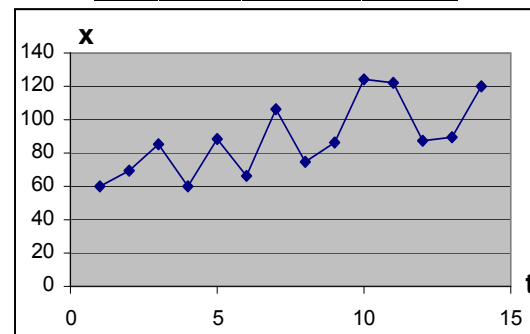
Hạn chế của các mô hình trung bình trượt đơn hay kép là

i> Đòi hỏi N dữ liệu cuối

ii> Trọng số như nhau ở N điểm này, trọng số 0 cho các điểm khác

Phương pháp làm trơn dạng mũ kép sẽ khắc phục được các hạn chế trên và trong đa số các trường hợp là thích hợp hơn trung bình trượt kép

t	x_t	M_T	$M_T^{(2)}$
1	60		
2	70		
3	85		
4	60		
5	88		
6	66	71.50	
7	106	79.17	
8	75	80.00	
9	86	80.17	
10	124	90.83	
11	122	96.50	83.03
12	87	100.00	87.78
13	89	97.17	90.78
14	120	104.67	94.89



Công thức:

Gọi x_i là dữ liệu gốc ở thời điểm thứ i

S_i là giá trị làm trơn dạng mũ đơn ở thời điểm thứ i

S_i' là giá trị làm trơn dạng mũ kép ở thời điểm thứ i

a_i là ước lượng của a ở thời điểm thứ i

b_i là ước lượng của b ở thời điểm thứ i

Ta có các quan hệ giữa chúng

$$\begin{cases} S_i = \alpha X_i + (1 - \alpha) S_{i-1} \end{cases} \quad (2.9)$$

$$\begin{cases} S_i' = \alpha S_i + (1 - \alpha) S_{i-1}' \end{cases} \quad (2.10)$$

Từ đó người ta suy ra được

$$\begin{cases} a_i = 2S_i - S_i' & \text{(theo công thức trung bình trượt kép)} \end{cases} \quad (2.11)$$

$$\begin{cases} b_i = \alpha(S_i - S_i') / (1 - \alpha) \end{cases} \quad (2.12)$$

và công thức dự báo DEWMA là

$$f_{N+\tau} = a_N + b_N \tau \quad (2.13)$$

Tham số trơn α :

Về mặt lý thuyết, α có thể nhận bất cứ giá trị nào giữa 0 và 1. Thực nghiệm cho thấy rằng giá trị tối ưu của α nằm giữa 0,1 và 0,2.

$\alpha = 0,1$ cho các dự báo bảo thủ

$\alpha = 0,2$ cho các dự báo phản hồi hệ thống tốt hơn.

Các giá trị ban đầu a_0 , b_0 , S_0 và S_0' :

$$b_0 = x_2 - x_1 \text{ và } a_0 = x_1 - b_0 = 2x_1 - x_2$$

- Có thể lấy trung bình N quan sát sau cùng làm ước lượng của S_0

$$S_0 = \frac{x_1 + x_2 + \dots + x_N}{N} \rightarrow S_0' = S_0 - \frac{1-\alpha}{\alpha} b_0$$

- Có thể sử dụng các ước lượng thống kê cho a_0 , b_0 : chẳng hạn để sử dụng phương pháp làm trơn dạng mũ kép từ chuỗi 11 quan sát, ta có thể lấy hồi qui tuyến tính các giá trị này làm ước lượng mức tăng và độ dốc a_0 , b_0

Khuyến nghị:

Phương pháp này thích hợp cho dữ liệu không có yếu tố mùa và không ổn định (có xu thế tăng hoặc giảm)

Ví dụ:

Cho chuỗi 24 số liệu một mặt hàng bán ra của 24 tháng. Hãy dự báo mức bán ra của tháng tiếp theo với tham số trơn $\alpha = 0,2$

Bước 1: Sử dụng phương pháp hồi quy tuyến tính cho các dữ liệu quan sát được ta tính được mức tăng và độ dốc cho xu thế chung của mô hình

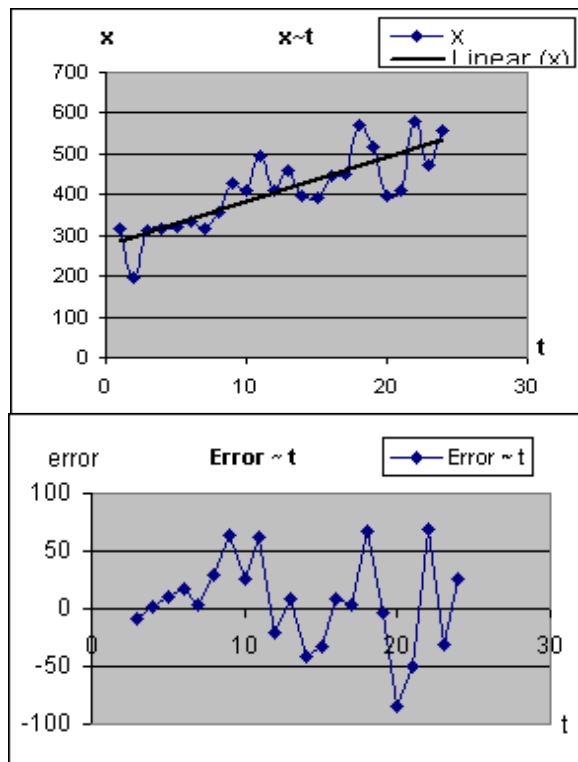
$$m_t = 275 + 10,88 t, t = 1, 2, \dots, 24 \rightarrow \text{chọn } a_0 = 275 \text{ và } b_0 = 10,88$$

$$S_0 = \frac{x_1 + x_2 + \dots + x_{24}}{24} = 411$$

$$\rightarrow S_0' = S_0 - \frac{1-\alpha}{\alpha} b_0 = 411 - \frac{0,8}{0,2} 10,88 = 367,5$$

t	x_t	S_t	S_t'	a_t	b_t	e_t	$\alpha = 0.2$
0		411	367.5	275	10.88		
1	317	392	372.4	412	4.944	-100	
2	194	353	368.5	337	-3.97	-139	
3	312	344	363.7	325	-4.8	-8.4	
4	316	339	358.7	319	-4.98	2.13	

5	322	335	354	317	-4.65	9.86
6	334	335	350.2	320	-3.78	17.8
7	317	332	346.5	317	-3.75	4.24
8	356	336	344.5	328	-2.02	29.7
9	428	355	346.5	363	2.049	63
10	411	366	350.4	382	3.891	25.6
11	494	392	358.6	425	8.233	61.3
12	412	396	366.1	425	7.404	-21
13	460	409	374.5	443	8.496	8.99
14	395	406	380.8	431	6.256	-42
15	392	403	385.3	421	4.452	-33
16	447	412	390.6	433	5.319	8.56
17	452	420	396.4	443	5.861	2.82
18	571	450	407.2	493	10.73	67.2
19	517	463	418.4	509	11.26	-2.8
20	397	450	424.8	476	6.351	-85
21	410	442	428.3	456	3.473	-50
22	579	470	436.5	503	8.253	68.2
23	473	470	443.2	497	6.741	-31
24	558	488	452.2	523	8.905	25.7
					$f_{25} =$	532.29



Bước 2: Tính các S_i và S_i' theo công thức (2.9) và (2.10), $i = 1, 2, \dots, 24$ rồi áp dụng công thức (2.11) và (2.12) ta tính được a_i , b_i và dự báo được

$$f_{24+1} = a_{24} + b_{24} \cdot (1) = 523,4 + 8,9 = 532,3 \approx 532$$

Nhận xét: Sai số là đại lượng ngẫu nhiên.

➤ Mô hình Holt

Mô hình Holt tương tự như mô hình trơn dạng mũ kép ngoại trừ việc nó không áp dụng công thức trơn kép mà *tách riêng việc làm trơn các giá trị xu thế*. Điều này làm tăng tính mềm dẻo, vì nó cho phép phần xu thế được làm trơn với tham số khác tham số được sử dụng trong chuỗi quan sát ban đầu. Cụ thể là:

$$\begin{cases} a_i = \alpha x_i + (1 - \alpha) (a_{i-1} + b_{i-1}) & \text{là mức tăng ở thời điểm } i \\ b_i = \beta (a_i - a_{i-1}) + (1 - \beta) b_{i-1} & \text{là xu thế (gradient) ở thời điểm } i \end{cases}$$

$$\text{Công thức dự báo: } f_{n+\tau} = a_n + b_n \tau \quad (2.10)$$

Các giá trị ban đầu của a và b là $a_0 = 2x_1 - x_2$; $b_0 = x_2 - x_1$

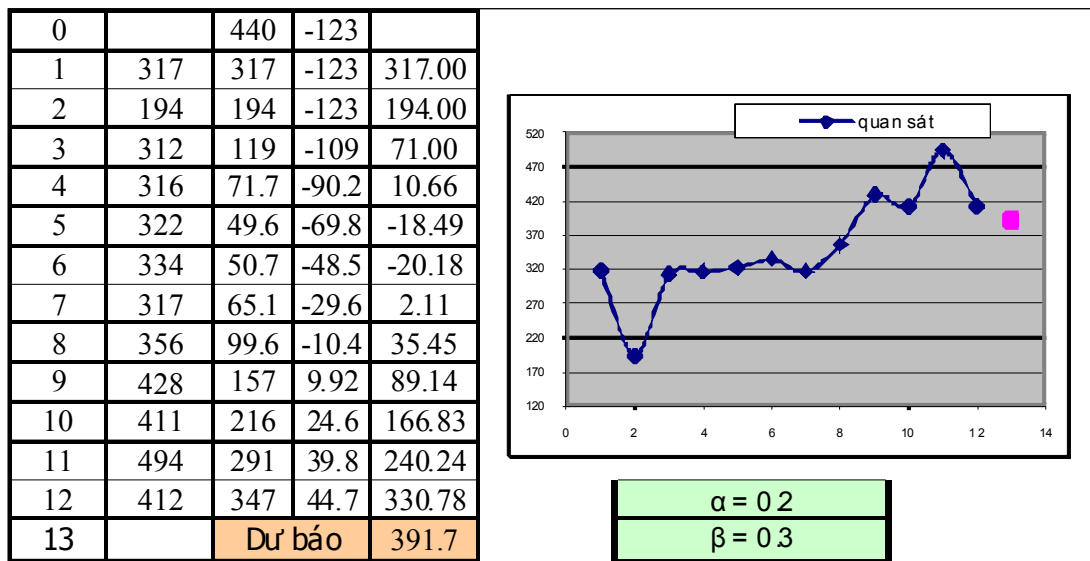
Các giá trị của α , β :

Nếu có sẵn một tập các giá trị ban đầu của dữ liệu thì nên sử dụng nó để tìm ra các giá trị α , β tốt nhất. Nếu ta lấy sai số trung bình bình phương (MSE) làm tiêu chuẩn ước lượng, ta có thể ước lượng một khoảng các giá trị khác nhau của α , β .

Ví dụ: Cho chuỗi dữ liệu hàng bán ra của 12 tháng năm ngoái. Hãy dự báo mức bán ra của tháng Giêng năm nay với $\alpha = 0,2$ và $\beta = 0,3$

Nhận xét: Nếu số lượng quan sát ít thì các phương pháp dự báo đều cho kết quả nghèo nàn, vì vậy các dự báo nhận được qua vài quan sát ban đầu nên bỏ qua khi tính sai số

MSE. Các dạng mô hình tron bậc cao hơn có thể sử dụng khi xu thế của mẫu có dạng bậc hai, dạng mũ, ...



2.4 Các mô hình tron có yếu tố thời vụ (mùa) của Winters

2.4.1 Các khái niệm chung

Các mô hình này có dạng tron bậc cao hơn, ưu điểm nổi trội của chúng là sự kết hợp chặt chẽ giữa tính xu thế và yếu tố thời vụ.

Các bước phân tích chung

Bước 1: Vẽ đồ thị biểu diễn chuỗi thời gian $x_t \sim t$

Bước 2: Phân tích ban đầu

a) Dữ liệu có thể hiện

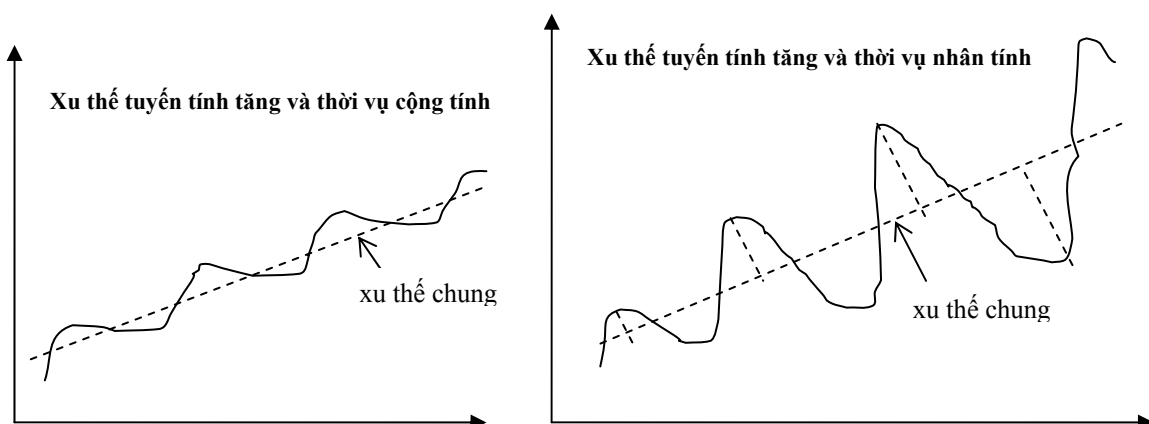
i> Yếu tố thời vụ?

ii> Tính xu thế?

b) Nếu có xu thế thì đó là xu thế tuyến tính hay xu thế mũ, có tắt dần không?

c) Nếu có yếu tố thời vụ thì đó là tác động cộng tính hay nhân tính, với bước thời vụ là bao nhiêu?

Việc nhận dạng dữ liệu sẽ dẫn đến sự lựa chọn mô hình dự báo phù hợp



Đối với dạng thời vụ cộng tính, các biến đổi theo thời vụ (loại trừ các nhiễu động) là không đổi về mức độ trung bình hoặc xu thế. Đối với dạng này, yếu tố thời vụ thường được ước lượng bởi sự khác biệt giữa giá trị quan sát được với xu thế chung. Ta có mô hình $X_t = T_t + I_t + a_t$ trong đó T_t là giá trị xu thế tại thời điểm t , I_t là yếu tố thời vụ tại thời điểm t và a_t là nhiễu động tại thời điểm t .

Đối với dạng thời vụ nhân tính, do tác động thời vụ tăng / giảm so với mức độ trung bình nên yếu tố thời vụ thường được ước lượng bởi tỉ lệ tăng / giảm so với xu thế chung. Ta có mô hình

$$X_t = T_t \cdot I_t + a_t \text{ hoặc } X_t = T_t \cdot I_t \cdot a_t$$

Các mô hình Winters dưới đây đều bao gồm các phương trình trơn dạng mũ tách biệt cho phần xu thế và phần thời vụ

2.4.2 Mô hình Winters cho dạng xu thế tuyến tính, thời vụ cộng tính

Các phương trình tính toán các thành phần bao gồm:

$$\begin{aligned} S_t &= \alpha (x_t - I_{t-L}) + (1 - \alpha) (S_{t-1} + b_{t-1}) \\ b_t &= \beta (S_t - S_{t-1}) + (1 - \beta) b_{t-1} \quad (\text{tương tự như mô hình Holt}) \\ I_t &= \gamma (X_t - S_t) + (1 - \gamma) I_{t-L} \end{aligned} \quad (2.11)$$

trong đó S_t là mức trơn tại thời điểm t

b_t là xu thế tại thời điểm t

I_t là yếu tố thời vụ tại thời điểm t

L là độ dài của thời vụ

Dự báo tại thời điểm $t = n$ cho các bước tiếp theo $\tau = 1, 2, 3, \dots, L$ là

$$f_{n+\tau} = S_n + b_n \cdot \tau + I_{n+\tau-L}$$

2.4.3 Mô hình Winters cho dạng xu thế mũ, thời vụ nhân tính

Các phương trình tính toán các thành phần bao gồm:

$$\begin{aligned} S_t &= \alpha \frac{x_t}{I_{t-L}} + (1 - \alpha) S_{t-1} b_{t-1} \\ b_t &= \beta \frac{S_t}{S_{t-1}} + (1 - \beta) b_{t-1} \\ I_t &= \gamma \frac{x_t}{S_t} + (1 - \gamma) I_{t-L} \end{aligned} \quad (2.12)$$

Dự báo tại thời điểm $t = n$ cho các bước tiếp theo $\tau = 1, 2, 3, \dots$ là

$$f_{n+\tau} = S_n \cdot b_n^\tau \cdot I_{n+\tau-L}$$

2.4.4 Mô hình Winters cho dạng xu thế tuyến tính, thời vụ nhân tính (dạng phổ biến nhất)

Các phương trình tính toán các thành phần bao gồm:

$$\begin{aligned}
S_t &= \alpha \frac{x_t}{I_{t-L}} + (1 - \alpha) (S_{t-1} + b_{t-1}) \\
b_t &= \beta (S_t - S_{t-1}) + (1 - \beta) b_{t-1} \\
I_t &= \gamma \frac{x_t}{S_t} + (1 - \gamma) I_{t-L}
\end{aligned}
\tag{2.13}$$

Dự báo tại thời điểm $t = n$ cho các bước tiếp theo $\tau = 1, 2, 3, \dots$ là

$$f_{n+\tau} = (S_n + b_n \tau) I_{n+\tau-L}$$

2.4.5 Mô hình Winters cho dạng xu thế mũ, thời vụ cộng tính

Các phương trình tính toán các thành phần bao gồm:

$$\begin{aligned}
S_t &= \alpha (x_t - I_{t-L}) + (1 - \alpha) S_{t-1} \quad b_{t-1} \\
b_t &= \beta \frac{S}{S_{t-1}} + (1 - \beta) b_{t-1} \\
I_t &= \gamma (x_t - S_t) + (1 - \gamma) I_{t-L}
\end{aligned}
\tag{2.14}$$

Dự báo tại thời điểm $t = n$ cho các bước tiếp theo $\tau = 1, 2, 3, \dots$ là

$$f_{n+\tau} = S_n b_n \tau + I_{n+\tau-L}$$

2.4.6 Các nhận xét chung về các mô hình Winters:

- **Ưu điểm:** Dễ hiểu, sử dụng nhiều trong thực tế, rất phù hợp cho dạng dữ liệu có tính xu thế và yếu tố thời vụ biến đổi.
- **Nhược điểm:** Đòi hỏi 3 tham số, một khi đã được tính toán tối ưu về sai số thì khó điều chỉnh khi nhập thêm quan sát mới.

Chú ý: Để tính toán tối ưu các tham số α, β, γ cần tính các giá trị ban đầu S_0, b_0 , và I_1, I_2, \dots, I_L có một số cách sau đây:

Cách 1: Dự báo lùi: dùng chuỗi x_t, x_{t-1}, \dots, x_1 dự báo các giá trị quá khứ x_0, x_{-1}, \dots phục vụ cho việc ước lượng S_0, b_0 , và I_1, I_2, \dots, I_L

Cách 2: Tách dữ liệu làm 2 phần:

- Phần 1: dùng để ước lượng S_0, b_0 , và I_1, I_2, \dots, I_L . Giả sử có các quan sát cho m thời vụ đầu và \bar{x}_j là trị trung bình của các quan sát ở thời vụ thứ j , với $j = 1, 2, \dots, m$. Ta có các ước lượng

$$b_0 = \frac{\bar{x}_m - \bar{x}_1}{(m-1)L};$$

$$S_0 = \bar{x}_1 - \frac{L}{2} b_0;$$

Yếu tố thời vụ tại các thời điểm $t = 1, 2, \dots, mL$ được tính theo công thức

$$I_t = \frac{x_t}{\bar{x}_1 - [(L+1)/2 - j] b_0}$$

với \bar{x}_i là trị trung bình của thời vụ thứ i tương ứng với thời điểm t , j là vị trí của thời điểm t trong thời vụ thứ i (ví dụ với $L+1 \leq t \leq 2L$ thì $i = 2$, nếu $t = L+1$ thì $j = 1$). Lấy trung bình các I_t trong m thời vụ ta được L giá trị

$$\bar{I}_t = \frac{1}{m} \sum_{k=0}^{m-1} I_{t+kL} \quad \forall t = 1, 2, \dots, L$$

Cuối cùng, các giá trị ban đầu I_1, I_2, \dots, I_L được chọn là chuẩn hóa của các đại lượng \bar{I}_t tương ứng

$$I_t = \frac{\bar{I}_t}{1/L \sum_{k=1}^L \bar{I}_k} \quad \forall t = 1, 2, \dots, L$$

- Phần 2 : dùng để tối ưu hóa α, β, γ theo các mục tiêu làm cực tiểu MSE, RMSE hay MAE. Các kỹ thuật dò tìm có thể là phương pháp thử sai, phương pháp đường dốc nhất, ...
- **Ví dụ:** Cho dãy 48 số liệu một loại nước giải khát đóng chai bán ra hàng tháng (tính theo kiện) của một hãng trong 4 năm liền. Với các tham số trơn $\alpha = 0,2$, $\beta = 0,1$ và $\gamma = 0,1$ hãy sử dụng bảng tính Excel dự báo lượng hàng sẽ bán trong 4 tháng tới.

Giải:

1. Đồ thị biểu diễn lượng hàng bán ra theo tháng cho thấy biên độ thời vụ ($L=12$) tăng theo lượng hàng bình quân bán ra (có xu thế tăng tuyến tính), do đó mô hình Winters với xu thế tuyến tính, thời vụ nhân tính là lựa chọn phù hợp.

2. Số liệu của 2 năm đầu được dùng để tính các giá trị ban đầu, ta có

$$\bar{x}_1 = 359,42; \bar{x}_2 = 493,58;$$

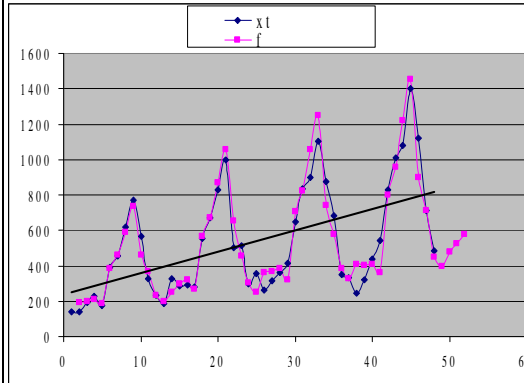
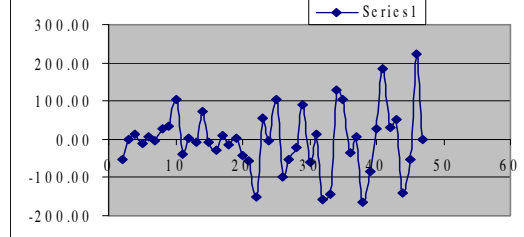
$$b_0 = \frac{493,58 - 352,75}{(2-1).12} = 12,01$$

$$S_0 = 352,75 - \frac{12}{2} \cdot 10,49 = 289,83$$

Cuối cùng các giá trị ban đầu I_1, I_2, \dots, I_L được chọn là chuẩn hóa của các đại lượng \bar{I}_t tương ứng

$$I_t = \frac{\bar{I}_t}{1/L \sum_{k=1}^L \bar{I}_k} \quad \forall t = 1, 2, \dots, L$$

3. Phần còn lại dùng để tối ưu hóa α, β, γ theo các mục tiêu làm cực tiểu MSE, RMSE hay MAE. Các kỹ thuật dò tìm có thể là phương pháp thử sai, phương pháp đường dốc nhất, ...

t	xt	S _t	b _t	l _t	f _t	e _t					
1	143	300.31	10.49	0.48	143.02	-0.02	L =	12	x _{1TB} =	352.75	
2	138	293.46	8.75	0.60	191.39	-53.39	α =	0.2	x _{2TB} =	478.58	
3	195	301.92	8.72	0.65	195.92	-0.92	β =	0.1	b ₀ =	10.49	
4	225	314.52	9.11	0.69	211.81	13.19	γ =	0.1	S ₀ =	289.83	
5	175	320.06	8.75	0.57	185.21	-10.21	m =	2			
6	389	329.78	8.85	1.17	383.32	5.68					
7	454	337.80	8.77	1.36	459.66	-5.66	t	x _t		l _t	l _t TB
8	618	349.58	9.07	1.71	592.22	25.78	1	143		0.48	0.47
9	770	362.16	9.42	2.05	734.09	35.91	2	138		0.45	0.60
10	564	388.56	11.12	1.26	459.12	104.88	3	195		0.62	0.64
11	327	391.31	10.28	0.91	365.20	-38.20	4	225		0.69	0.67
12	235	402.68	10.39	0.58	231.88	3.12	5	175		0.52	0.56
13	189	409.83	10.07	0.47	196.71	-7.71	6	389		1.12	1.14
14	326	444.35	12.51	0.61	252.46	73.54	7	454		1.27	1.33
15	289	454.68	12.29	0.65	296.07	-7.07	8	618		1.68	1.68
16	293	459.11	11.51	0.68	319.97	-26.97	9	770		2.03	2.01
17	279	474.43	11.89	0.57	268.13	10.87	10	564		1.45	1.21
18	552	483.64	11.62	1.16	567.61	-15.61	11	327		0.82	0.90
19	674	495.61	11.66	1.36	671.60	2.40	12	235		0.57	0.57
20	827	502.27	11.16	1.71	869.82	-42.82	13	189		0.45	0.98
21	1000	508.08	10.62	2.05	1054.96	-54.96	14	326		0.76	
22	502	494.82	8.23	1.23	652.10	-150.10	15	289		0.65	
23	512	515.48	9.48	0.91	455.74	56.26	16	293		0.65	
24	300	523.76	9.36	0.58	303.43	-3.43	17	279		0.60	t
25	359	577.75	13.82	0.49	253.08	105.92	18	552		1.17	49
26	264	559.18	10.58	0.60	363.51	-99.51	19	674		1.39	50
27	315	553.21	8.93	0.64	368.52	-53.52	20	827		1.67	51
28	361	555.81	8.29	0.68	382.52	-21.52	21	1000		1.98	52
29	414	596.14	11.50	0.58	322.43	91.57	22	502		0.97	
30	647	597.22	10.46	1.16	707.64	-60.64	23	512		0.97	
31	836	609.40	10.63	1.36	824.28	11.72	24	300		0.56	
32	901	601.54	8.78	1.69	1058.95	-157.95					
33	1104	596.16	7.36	2.03	1248.75	-144.75					
34	874	624.60	9.47	1.25	744.10	129.90					
35	683	656.60	11.72	0.93	579.97	103.03					
36	352	656.57	10.55	0.57	385.95	-33.95					
37	332	669.37	10.77	0.49	326.47	5.53					
38	244	625.42	5.30	0.58	408.26	-164.26					
39	320	604.72	2.70	0.63	403.07	-83.07					
40	437	614.96	3.45	0.68	411.46	25.54					
41	544	681.07	9.72	0.61	361.08	182.92					
42	830	696.18	10.26	1.16	798.88	31.12					
43	1011	714.05	11.02	1.36	959.33	51.67					
44	1081	708.22	9.34	1.67	1223.10	-142.10					
45	1400	712.20	8.80	2.02	1454.23	-54.23					
46	1123	756.55	12.35	1.27	900.92	222.08					
47	713	768.91	12.36	0.93	712.94	0.06					
48	487	794.89	13.72	0.58	447.95						

Nhận xét: Sai số là các đại lượng ngẫu nhiên, có biên độ tăng dần. Nguyên nhân là do số quan sát dùng để tối ưu các tham số α , β , γ là quá ít (chỉ có 2 thời vụ)

2.5 Các phương pháp phân ly (Decomposition)

2.5.1 Các công thức chung

Các mô hình làm trơn đã xét trước đây đều dựa trên ý tưởng là nếu chuỗi thời gian có một định dạng (mẫu) thì mẫu này có thể được tách khỏi tính ngẫu nhiên bằng cách làm trơn các giá trị quá khứ. Tác dụng của việc làm trơn là loại bỏ thành phần ngẫu nhiên trong chuỗi rồi sử dụng mẫu cho việc dự báo. *Các phương pháp làm trơn đều chưa nhận dạng được từng thành phần riêng biệt của mẫu.*

Trên thực tế, mẫu có thể được tách (phân ly) thành hai hoặc nhiều nhân tố, đặc biệt là khi xuất hiện các kiểu thời vụ trong dữ liệu. Trong nhiều tình huống, sẽ là rất tốt nếu người dự báo biết được tỉ lệ nào của dữ liệu tại thời điểm đã biết phản ánh mức tăng / giảm chung và tỉ lệ nào của dữ liệu chỉ đơn giản thể hiện sự dao động của thời vụ.

Các phương pháp phân ly là một trong các cách dự báo cổ điển nhất. Các phương pháp này thường cố gắng nhận dạng 3 thành phần tách biệt của chuỗi thời gian là xu thế, chu kì và thời vụ.

Xu thế là tính xuyên suốt của chuỗi như tăng, giảm, ổn định.

Chu kì là thời kì tăng trưởng hay suy thoái của nền kinh tế, của một ngành công nghiệp; giai đoạn ElNino hay LaNina của khí hậu.

Thời vụ là các dao động của các quan sát theo một chiều dài thời gian cố định (mùa, năm, ...)

Dựa trên giả thiết dữ liệu được cấu thành từ một mẫu cùng với sai số (ngẫu nhiên)

Dữ liệu = mẫu + sai số = hàm(xu thế, chu kì, thời vụ) + sai số

Mô hình chung của các phương pháp phân ly là

$$x_t = f(T_t, C_t, S_t, E_t) \quad (2.15)$$

Nhận xét: để nhận diện được thành phần chu kì, ta cần có ít nhất 10 năm số liệu. Trong dự báo ngắn hạn, thành phần xu thế T_t thường bao gồm luôn thành phần chu kì C_t .

Dạng hàm chính xác của quan hệ (2.15) phụ thuộc vào phương pháp phân ly cụ thể được sử dụng. Ta có các mô hình sau đây

i> $x_t = T_t + S_t + E_t$ mô hình cộng tính

ii> $x_t = T_t \cdot S_t \cdot E_t$ mô hình nhân tính

iii> $x_t = T_t \cdot S_t + E_t$ mô hình nhân tính với sai số cộng tính

Các mô hình nhân tính thường xuất hiện nhiều trong lĩnh vực kinh tế. Đối với mỗi loại mô hình trên, phải vẽ đồ thị để kiểm tra xem yếu tố thời vụ là cộng tính hay nhân tính.

2.5.2 Phương pháp phân ly cổ điển (Classical Decomposition).

Phương pháp này phân ly chuỗi thời gian ra các thành phần như thời vụ, xu thế. Dự báo được coi như ngoại suy tuyến tính của chuỗi thời gian trong quá khứ. Các bước tiến hành như sau:

- **Tính các trung bình trượt trung tâm M_t của chuỗi x_t với độ dài $N = L$** (độ dài của thời vụ) nhằm mục đích loại bỏ thành phần thời vụ và thành phần ngẫu nhiên, chỉ giữ lại thành phần xu thế (và chu kỳ).

- **Nhận lại thành phần thời vụ và ngẫu nhiên:**

$$S_t + E_t = x_t - M_t \text{ đối với mô hình cộng tính thời vụ}$$

$$S_t + E_t = x_t / M_t \text{ đối với mô hình nhân tính thời vụ}$$

- **Cô lập các yếu tố thời vụ** bằng cách lấy trung bình các yếu tố này tại các thời điểm cách nhau một khoảng thời gian L . Sau đó chuẩn hóa các yếu tố trung bình này để **loại bỏ thành phần ngẫu nhiên có ở bước 2** (chỉ còn lại S_t , $t=1, 2, \dots, L$).

- **Tách thành phần thời vụ ra khỏi dữ liệu**

$$d_t = x_t - S_t \text{ nếu yếu tố thời vụ có dạng cộng tính}$$

$$d_t = x_t / S_t \text{ nếu yếu tố thời vụ có dạng nhân tính}$$

- **Tính thành phần xu thế $T_t = a + bt$** bằng cách sử dụng phương pháp bình phương bé nhất tìm b và a từ hệ phương trình

$$\begin{cases} Na + (\sum t) b = \sum d_t \\ (\sum t) a + (\sum t^2) b = \sum td_t \end{cases}$$

$$\rightarrow b = \frac{N \sum t \cdot d_t - \sum t \cdot \sum d_t}{N \sum t^2 - (\sum t)^2}; \quad a = \frac{\sum d_t}{N} - b \frac{\sum t}{N}$$

- **Tách thành phần xu thế T_t khỏi M_t để nhận được thành phần chu kỳ (nếu cần)**

$$C_t = M_t - T_t$$

- **Tính sai số**

$$E_t = X_t - T_t - S_t \text{ nếu yếu tố thời vụ có dạng cộng tính}$$

$$E_t = X_t - T_t / S_t \text{ nếu yếu tố thời vụ có dạng nhân tính}$$

- **Dự báo**

$$f_t = T_t + S_t \text{ nếu yếu tố thời vụ có dạng cộng tính}$$

$$f_t = T_t / S_t \text{ nếu yếu tố thời vụ có dạng nhân tính}$$

2.5.3 Các ví dụ

Ví dụ 1:

Sau khi phân tích chuỗi thời gian dưới đây, người ta nhận thấy chuỗi này có xu thế tuyến tính và yếu tố thời vụ cộng tính, $L=3$. Vậy mô hình phân ly thích hợp là $x_t = T_t + S_t + E_t$

- Tính các trung bình trượt trung tâm M_t của x_{t-1} , x_t và x_{t+1} , $t = 2, 3, \dots, 11$

- Tính các yếu tố thời vụ S_t

$S_1 = -2$ là trung bình cộng của $x_1 - M_1$, $x_4 - M_4$, $x_7 - M_7$ và $x_{10} - M_{10}$

$S_2 = 0$ là trung bình cộng của $x_2 - M_2$, $x_5 - M_5$, $x_8 - M_8$ và $x_{11} - M_{11}$

$S_3 = 2$ là trung bình cộng của $x_3 - M_3$, $x_6 - M_6$, $x_9 - M_9$ và $x_{12} - M_{12}$

$\rightarrow S_1 + S_2 + S_3 = 0$ có nghĩa là các S_t đã được chuẩn hóa sẵn!

- Tách thành phần thời vụ ra khỏi dữ liệu bằng cách tính

$$d_t = x_t - S_t$$

- Để xác định thành phần xu thế $T_t = a + bt$ ta sử dụng phương pháp bình phương bé nhất tính a và b :

$$\sum t = 78, \sum t^2 = 650, \sum td_t = 1704.1, \sum d_t = 217.5$$

$$\rightarrow b = 2.02, a = 5.04$$

Từ đó ta lập được bảng tính dưới đây và tính được

$$F_{13} = T_{13} + S_{13} = 5.04 + 2.02 * 13 + (-2) = 29.26$$

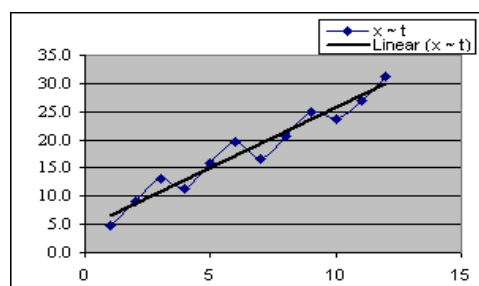
$$F_{14} = T_{14} + S_{14} = 5.04 + 2.02 * 14 - 0.03 = 33.24$$

.....

t	x_t	M_t	$x_t - M_t$	S_t	d_t	T_t	E_t
1	4.7			-2.00	6.7	7.06	-0.36
2	9.0	8.9	0.1	0.00	9.0	9.07	-0.07
3	13.0	11.1	1.9	2.00	11.0	11.09	-0.09
4	11.3	13.4	-2.1	-2.00	13.3	13.11	0.19
5	15.9	15.6	0.3	0.00	15.9	15.12	0.78
6	19.6	17.4	2.2	2.00	17.6	17.14	0.46
7	16.7	19.0	-2.3	-2.00	18.7	19.16	-0.46
8	20.7	20.8	-0.1	0.00	20.7	21.18	-0.48
9	25.0	23.1	1.9	2.00	23.0	23.19	-0.19
10	23.6	25.2	-1.6	-2.00	25.6	25.21	0.39
11	27.0	27.3	-0.3	0.00	27.0	27.23	-0.23
12	31.3			2.00	29.3	29.24	0.06
13				-2.00		31.26	
14				-0.03		33.28	

Dự báo	29.26	a =	5.04
	33.24	b =	2.02

Chú ý: Khi tính trung bình trượt trung tâm M_t của chuỗi thời gian có độ dài thời vụ L là một số chẵn, xuất hiện vấn đề là đặt M_t vào đâu? Để khắc phục tình huống này, người ta lấy trung bình cộng của 2 trung bình trượt kề nhau làm giá trị của M_t .



Ví dụ 2:

Chuỗi thời gian dưới đây có yếu tố thời vụ với độ dài $L=4$. Hãy phân ly nó thành các thành phần, biết rằng nó có xu thế tuyến tính và thời vụ nhân tính. Ta áp dụng dạng mô hình phân ly:

$$x_t = T_t \cdot S_t + E_t$$

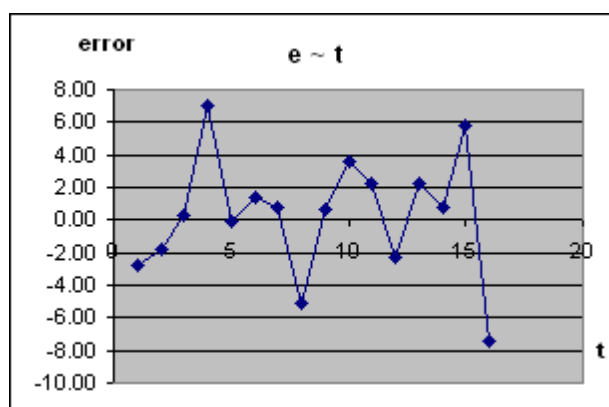
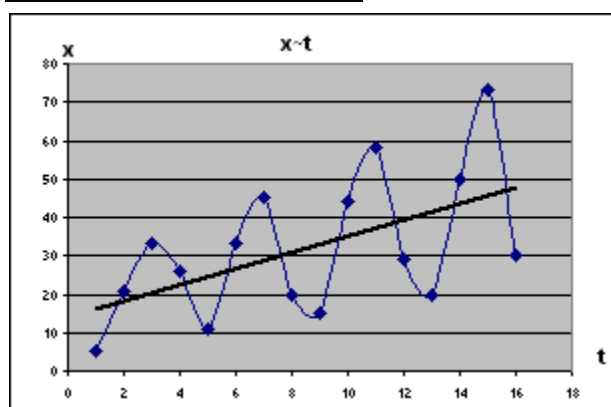
với

$$d_t = x_t / S_t$$

$$E_t = x_t - T_t \cdot S_t$$

Lập bảng tính các thành phần, M_t là trung bình cộng của 2 trung bình trượt 4 bước kế nhau của các quan sát x_t , $t = 3, 4, \dots, 14$.

t	x_t	4-MA	M_t	x_t/M_t	S_t	d_t	t^2	td_t	T_t	e_t	Dự báo
1	5				0.45	11.11	1	11.11	17.25	-2.76	7.76
2	21	21.25			1.19	17.59	4	35.18	19.09	-1.80	22.80
3	33	22.75	22	1.5	1.56	21.14	9	63.41	20.94	0.31	32.69
4	26	25.75	24.3	1.07	0.83	31.19	16	124.77	22.78	7.01	18.99
5	11	28.75	27.3	0.4	0.45	24.45	25	122.24	24.62	-0.08	11.08
6	33	27.25	28	1.18	1.19	27.64	36	165.84	26.46	1.40	31.60
7	45	28.25	27.8	1.62	1.56	28.82	49	201.75	28.31	0.80	44.20
8	20	31	29.6	0.68	0.83	24	64	191.96	30.15	-5.13	25.13
9	15	34.25	32.6	0.46	0.45	33.34	81	300.05	31.99	0.61	14.39
10	44	36.5	35.4	1.24	1.19	36.85	100	368.53	33.84	3.60	40.40
11	58	37.75	37.1	1.56	1.56	37.15	121	408.63	35.68	2.29	55.71
12	29	39.25	38.5	0.75	0.83	34.79	144	417.51	37.52	-2.28	31.28
13	20	43	41.1	0.49	0.45	44.45	169	577.88	39.37	2.29	17.71
14	50	43.25	43.1	1.16	1.19	41.88	196	586.30	41.21	0.80	49.20
15	73				1.56	46.76	225	701.34	43.05	5.78	67.22
16	30				0.83	35.99	256	575.88	44.89	-7.42	37.42
17					0.45		289		46.74	$F_{17} =$	21.03
136						497.1	1496	4852.39	18496		
b=	1.84	a=	15.4								



Nhận xét: Đồ thị của sai số e_t không có xu thế và không có định dạng (E_t là các đại lượng ngẫu nhiên).

2.6 BÀI TẬP CHƯƠNG 2

1. Số máy tính bán ra hàng tuần của một đại lí cho bởi bảng dưới đây

Tuần	1	2	3	4	5	6	7	8	9	10	11	12
Máy	75	74	79	83	69	78	71	80	77	85	81	70

Chứng tỏ rằng mô hình làm trơn dạng mũ đơn là thích hợp cho dữ liệu này.

Sử dụng mô hình làm trơn dạng mũ đơn với $\alpha = 0,1$ dự báo số máy tính sẽ bán ra trong tuần tới.

Có thể sử dụng mô hình này để cung cấp dự báo tin cậy cho tuần thứ 20 không?

2. Doanh thu bán hàng của một đại lí trong 6 tháng gần đây là

1000 3000 4500 5500 5100 5400 (tính bằng triệu đồng)

Sử dụng đồ thị chứng tỏ rằng việc sử dụng tính xu thế là cần thiết

Dự báo doanh thu cho 3 tháng tới, sử dụng lần lượt các phương pháp

Trung bình trượt dạng mũ kép

Mô hình Holt với $\alpha = 0,1$ và $\beta = 0,1$

Mô hình Holt với $\alpha = 0,9$ và $\beta = 0,9$

Tính các sai số MAE và MSE cho cả 3 phương pháp

3. Số lượng khách hàng đăng kí sử dụng truyền hình cáp trong 6 tháng đầu năm là

3403, 3790, 4025, 426, 4192, 4235.

Hãy sử dụng mô hình Holt với các tham số trơn $\alpha = \beta = 0,1$ dự báo lượng khách hàng đăng kí sử dụng dịch vụ này trong tháng tới.

Vẽ đồ thị minh họa các giá trị quan sát được và dự báo rồi rút ra nhận xét.

Tính sai số MSE. Có cách nào cải thiện dự báo không?

4. Trung tâm báo chí quốc gia muốn dự báo nhu cầu hội nghị cho từng quý của năm tới. Dữ liệu thu thập được của 4 năm gần đây là

Năm	Quý	Nhu cầu	Năm	Quý	Nhu cầu
2002	1	10	2004	1	13
	2	31		2	34
	3	43		3	48
	4	16		4	19
2003	1	11	2005	1	15
	2	33		2	37
	3	46		3	51
	4	17		4	21

Vẽ đồ thị và giải thích vì sao mô hình Winters với xu thế tuyến tính, thời vụ cộng tính là thích hợp cho mục đích dự báo.

Sử dụng mô hình này với các tham số trơn $\alpha = 0,2$; $\beta = 0,1$ và $\gamma = 0,1$ dự báo nhu cầu hội nghị cho các quý của năm tới.

5. Sử dụng phương pháp phân ly cổ điển, tách chuỗi thời gian ở bài 4 ra các thành phần T, S, E.

Vẽ đồ thị $x \sim t$ để tìm mô hình phân ly thích hợp.

Tính các dự báo cho 4 quý tiếp theo và đánh giá sai số.

6. Dữ liệu sau đây là tình hình tiêu thụ một loại sản phẩm theo quý của 6 năm gần đây

Năm	Quý	Sản phẩm	Năm	Quý	Sản phẩm	Năm	Quý	Sản phẩm
1999	1	362	2001	1	473	2003	1	628
	2	385		2	513		2	707
	3	432		3	582		3	773
	4	341		4	474		4	592
2000	1	382	2002	1	544	2004	1	627
	2	409		2	582		2	725
	3	498		3	681		3	854
	4	387		4	557		4	661

Vẽ đồ thị để kiểm tra sự biến đổi theo thời vụ ? Biến đổi này là cộng tính hay nhân tính?

Dữ liệu có thể hiện tính xu thế không ? Nếu có thì đó là xu thế tuyến tính hay dạng mũ?

Sử dụng mô hình Winters thích hợp dự báo tình hình tiêu thụ sản phẩm cho 4 quý tới.

3 CHƯƠNG 3 : PHÂN TÍCH CHUỖI THỜI GIAN VÀ CÁC MÔ HÌNH CỦA BOX-JENKINS

Trong chương trước chúng ta đã xét các kỹ thuật dự báo dựa trên cơ sở các phép làm trơn với giả thiết rằng giá trị trung bình của chuỗi thời gian là hàm xác định của thời gian và quan sát ở bất kì thời điểm nào cũng là tổng của trị trung bình với thành phần sai số ngẫu nhiên

$$x_t = \mu_t + e_t$$

tức là sai số e_t là các biến ngẫu nhiên độc lập đối với t , còn μ_t là hàm trung bình xác định theo t . Nhận xét rằng nếu dãy $\{e_t\}$ là các biến ngẫu nhiên độc lập thì dãy các quan sát $\{x_t\}$ cũng là các biến ngẫu nhiên độc lập. Nhưng giả thiết về sự độc lập của các quan sát $\{x_t\}$ lại thường không được đảm bảo vì có nhiều chuỗi thời gian mà các quan sát liên tiếp phụ thuộc vào nhau chặt chẽ. Trong các trường hợp đó, các kỹ thuật dự báo dựa trên các phép làm trơn có thể trở nên không thích hợp bởi chúng không tận dụng được ưu điểm của sự phụ thuộc giữa các quan sát một cách có hiệu quả nhất. Trên thực tế các phương pháp làm trơn thường chỉ cho kết quả tốt đối với các quan sát phụ thuộc vào thời vụ.

Dựa trên giả thiết rằng các giá trị liên tiếp của chuỗi thời gian có liên quan với nhau, Box và Jenkins cố gắng khám phá điều đó và sử dụng trong các mô hình dự báo. Các mô hình này do Box và Jenkins đề xướng nên thường được gọi là các mô hình Box-Jenkins. Các kỹ thuật phân tích chuỗi thời gian xét trong chương này sẽ khai thác sự phụ thuộc giữa các quan sát. Các giá trị tương lai của chuỗi thời gian sẽ được xác định từ tổ hợp của các giá trị quá khứ và sai số quá khứ.

3.1 Các mô hình chuỗi thời gian ARMA (AutoRegressive-Moving Average)

Trong phân tích chuỗi thời gian, phương pháp **Box – Jenkins**, được đặt tên sau khi hai nhà thống kê học George Box và Gwilym Jenkins áp dụng các mô hình tự hồi quy trung bình trượt ARMA hay ARIMA, tìm ra mô hình phù hợp nhất của chuỗi thời gian các giá trị thời gian trong quá khứ để tạo ra các dự báo.

3.1.1 Mô hình tự hồi quy bậc p - AR(p)

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \dots + \Phi_p x_{t-p} + a_t + \theta_0 \quad (3.1)$$

Vậy mô hình AR(p) là tổng có trọng số của các quan sát đã cho với các trọng số (các tham số) $\Phi_1, \Phi_2, \dots, \Phi_p$. Các tham số này cần được ước lượng để tìm hàm dự báo.

3.1.2 Mô hình trung bình trượt bậc q - MA(q)

$$x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + \theta_0 \quad (3.2)$$

3.1.3 Mô hình hỗn hợp tự hồi quy-trung bình trượt bậc (p,q) - ARMA(p,q)

$$x_t - \Phi_1 x_{t-1} - \Phi_2 x_{t-2} - \dots - \Phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + \theta_0 \quad (3.3)$$

trong đó x_t là các giá trị quá khứ và hiện tại của chuỗi thời gian,

θ_0 là hằng số (mức của quá trình),

Φ_t, θ_t là các hằng số (tham số) cần được ước lượng,

a_t là các nhiễu động (sai số) độc lập, có trung bình $\mu_A = 0$, phương sai σ_A^2 không đổi, và không nhất thiết phải có phân phối chuẩn.

➤ Một số mô hình thường gặp

$$\text{AR}(1) \quad x_t = \Phi_1 x_{t-1} + a_t + \theta_0$$

$$\text{AR}(2) \quad x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t + \theta_0$$

$$\text{MA}(1) \quad x_t = a_t - \theta_1 a_{t-1} + \theta_0 = (1 - \theta_1 B) a_t + \theta_0$$

$$\begin{aligned} \text{MA}(2) \quad x_t &= a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} + \theta_0 \\ &= (1 - \theta_1 B - \theta_2 B^2) a_t + \theta_0 \text{ (mô hình tron dạng mũ đơn)} \end{aligned}$$

$$\text{ARMA}(0,0) \quad x_t = a_t + \theta_0$$

$$\text{ARMA}(1,1) \quad x_t = \Phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} + \theta_0$$

$$\text{ARMA}(1,2) \quad x_t = \Phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} + \theta_0$$

$$\text{ARMA}(2,1) \quad x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t - \theta_1 a_{t-1} + \theta_0$$

3.2 Các điều kiện cần về tính dừng và tính khả nghịch

Phương trình (3.3) có thể viết lại dưới **dạng toán tử**

$$\Phi(B) x_t = \theta(B) a_t + \theta_0 \quad (3.4)$$

3.2.1 Điều kiện dừng

$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p$ được gọi là *toán tử tự hồi quy*,

B là toán tử lùi, tức là $Bx_t = x_{t-1}$; $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$, ...;

Box và Jenkins đã chứng minh rằng khi *quá trình* $AR(p)$ là *dừng* (có trung bình và phương sai bất biến), các nghiệm của phương trình $\Phi(z) = 0$ đều nằm ngoài vòng tròn đơn vị (đối với quá trình $MA(q)$ điều kiện dừng luôn luôn thỏa mãn với mọi giá trị của các tham số θ_i).

3.2.2 Điều kiện khả nghịch

$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ được gọi là *toán tử trung bình trượt*; khi quá trình $MA(q)$ là *khả nghịch*, các nghiệm của phương trình $\theta(z) = 0$ đều nằm ngoài vòng tròn đơn vị (đối với quá trình $AR(p)$, điều kiện khả nghịch luôn luôn thỏa mãn với mọi giá trị của các tham số Φ_i).

Các điều kiện về tính dừng và tính khả nghịch được áp dụng một cách độc lập, và nói chung $\Phi(B)$ và $\theta(B)$ sẽ không có cùng bậc ($p \neq q$).

Chú ý: Với các điều kiện trên, các mô hình ARMA đều *dừng* theo nghĩa trung bình $\mu_X = E(x_t)$ và phương sai $\sigma_X^2 = V(x_t) = E[(x_t - E(x_t))^2]$ là bất biến.

Ví dụ:

- Xét mô hình ARMA(1,0) hay AR(1): $x_t = \Phi_1 x_{t-1} + a_t + \theta_0$

Lấy kì vọng 2 vế

$$E x_t = \Phi_1 E x_{t-1} + E a_t + E \theta_0$$

hay

$$\mu_X - \Phi_1 \mu_X = \theta_0$$

$$\mu_X = \theta_0 / (1 - \Phi_1) = \text{const}$$

Lấy phương sai 2 vế

$$V x_t = \Phi_1^2 V x_{t-1} + V a_t + V \theta_0$$

hay

$$\sigma_X^2 = \Phi_1^2 \sigma_X^2 + \sigma_A^2 + 0$$

$$\sigma_X^2 = \sigma_A^2 / (1 - \Phi_1^2) = \text{const}$$

- Xét mô hình MA(q): $x_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + \theta_0$

Lấy kì vọng 2 vế ta được

$$\mu_X = \theta_0 = \text{const}$$

Lấy phương sai 2 vế ta được

$$\sigma_X^2 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_A^2 = \text{const}$$

- Xét mô hình ARMA(1,1):

$$x_t = \Phi_1 x_{t-1} + a_t - \theta_1 a_{t-1} + \theta_0$$

Lấy kì vọng 2 vế ta được

$$\mu_X = \Phi_1 \mu_X + \theta_0 \rightarrow \mu_X = \theta_0 / (1 - \Phi_1) = \text{const}$$

Lấy phương sai 2 vế ta được

$$\sigma_X^2 = \Phi_1^2 \sigma_X^2 + \sigma_A^2 + \theta_1^2 \sigma_A^2$$

$$\rightarrow \sigma_X^2 = \frac{(1 + \theta_1^2) \sigma_A^2}{1 - \Phi_1^2} = \text{const}$$

→ Từ nay về sau, để sử dụng được các mô hình ARMA, ta sẽ luôn luôn giả thiết *chuỗi thời gian là dừng*

3.3 Các trợ giúp cho việc phân tích chuỗi thời gian

3.3.1 Biểu diễn đồ họa chuỗi thời gian

Về đồ thị luôn luôn là khởi đầu cần thiết để xem xu thế, chu kì, thời vụ và các điểm ngoại lai của các quan sát.

3.3.2 Hệ số tự tương quan ACF (Auto Correlation Function)

Định nghĩa 1: Gọi x_1, x_2, \dots, x_N là tập các quan sát của chuỗi thời gian. Ta định nghĩa hệ số tự tương quan

$$\begin{aligned} \rho_k &= \rho(x_t, x_{t+k}) \\ &= [E(x_t - E(x_t))(x_{t+k} - E(x_{t+k}))] / [\sigma(x_t) \sigma(x_{t+k})] \end{aligned} \quad (3.5)$$

$$= [E(x_t x_{t+k}) - E(x_t) E(x_{t+k})] / [\sigma(x_t) \sigma(x_{t+k})] \text{ với } k = 1, 2, \dots \quad (3.5')$$

Nhận xét:

$$-1 \leq \rho_k \leq 1 \text{ với mọi } k = 1, 2, \dots$$

$$\rho_k \approx 1: x_t \text{ và } x_{t+k} \text{ có tương quan (dương) chặt;}$$

$$\rho_k \approx -1: x_t \text{ và } x_{t+k} \text{ có tương quan (âm) chặt;}$$

Đối với *chuỗi thời gian dừng*, trung bình μ_X và phương sai σ_X^2 là *bất biến* và do đó là *độc lập đối với t*.

$$\text{Vậy thì từ (3.5')} \text{ ta có } \rho_k = [E(x_t x_{t+k}) - \mu_X^2] / \sigma_X^2 \text{ với } k = 1, 2, \dots \quad (3.6)$$

là các hệ số tự tương quan lí thuyết đối với các bước nhảy $k = 1, 2, \dots$

Định nghĩa 2: Các hiệp phương sai (autocovariance) ở bước k là

$$\gamma_k = \text{Cov}(x_t, x_{t+k}) = E[(x_t - \mu_X)(x_{t+k} - \mu_X)] \quad (3.6')$$

từ (3.5) ta có công thức nữa tính ρ_k

$$\rho_k = \gamma_k / \gamma_0 \text{ (cần nhấn mạnh rằng } \rho_0 = 1) \quad (3.7)$$

Một số ví dụ:

- Đối với quá trình AR(1) người ta tính được $\gamma_k = \Phi_1^k \sigma_A^2 / (1 - \Phi_1^2)$ $k=0, 1, \dots$

→ $\rho_k = \Phi_1^k$ hay các hệ số tự tương quan của AR(1) tắt rất nhanh khi $k > 1$ vì $|\Phi_1| < 1$ (từ điều kiện cần $1 - \Phi_1 z = 0$ có nghiệm $|z| > 1 \rightarrow |\Phi_1| < 1$).

- Đối với quá trình MA(1) người ta tính được $\gamma_0 = \sigma_A^2 (1 + \theta_1^2)$

$$\rightarrow \rho_k = \begin{cases} \frac{-\theta_1}{1+\theta_1^2}, & k=1 \\ 0, & k>1 \end{cases} \rightarrow \rho_k \text{ biến mất sau bước } k=1$$

• Đối với quá trình ARMA(1,1) người ta tính được

$$\gamma_0 = \Phi_1 \gamma_1 + \sigma_A^2 [1 - \theta_1 (\Phi_1 - \theta_1)]$$

$$\gamma_1 = \Phi_1 \gamma_0 - \theta_1 \sigma_A^2$$

$$\gamma_k = \Phi_1 \gamma_{k-1} \text{ khi } k \geq 2$$

Từ đó suy ra

$$\rho_1 = \frac{(1 - \Phi_1 \theta_1)(\Phi_1 - \theta_1)}{1 + \theta_1^2 - 2\Phi_1 \theta_1}$$

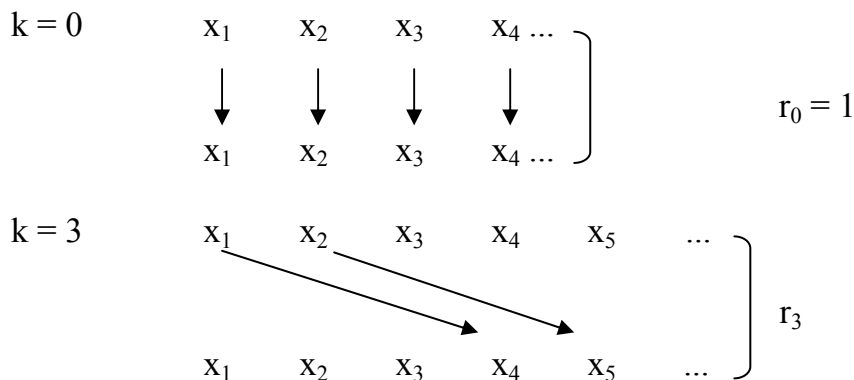
$$\rho_k = \Phi_1 \rho_{k-1} = \Phi_1^2 \rho_{k-2} = \dots = \Phi_1^{k-1} \rho_1 \text{ khi } k \geq 2$$

Ước lượng của ρ_k là hàm tự tương quan mẫu SACF (Sample AutoCorrelation Function). Hệ số tự tương quan mẫu ở bước k được định nghĩa:

$$r_k = \frac{\frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2}, \quad k = 1, 2, \dots \quad \text{với } \bar{x} = \frac{\sum_{t=1}^N x_t}{N} \text{ là trung bình mẫu}$$

$$\text{Cụ thể } r_1 = \frac{\frac{1}{N-1} \sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2}, \quad r_2 = \frac{\frac{1}{N-2} \sum_{t=1}^{N-2} (x_t - \bar{x})(x_{t+2} - \bar{x})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2}, \dots \quad (3.8)$$

Ý nghĩa: r_k đo mức độ tương quan giữa x_t và x_{t+k} . Vậy đối với chuỗi các biến ngẫu nhiên x_1, x_2, \dots, x_N thì $r_1, r_2, \dots \approx 0$

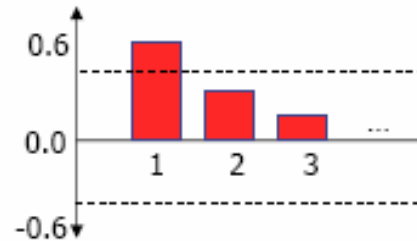


Đối với chuỗi có yếu tố thời vụ theo quý: r_3 là đáng kể

Đối với chuỗi có yếu tố thời vụ theo năm: r_{12} là đáng kể

Ví dụ: Chuỗi $x_t = \{7 \quad 8 \quad 7 \quad 6 \quad 5 \quad 4 \quad 5 \quad 6 \quad 4\}$ có các tự tương quan

Một bước		Hai bước		Ba bước	
7	8	7	7	7	6
8	7	8	6	8	5
7	6	7	5	7	4
6	5	6	4	6	5
5	4	5	5	5	6
4	5	4	6	4	4
5	6	5	4		
6	4				
$r_1 = 0.62$		$r_2 = 0.32$		$r_3 = 0.15$	



3.3.3 Hàm tự tương quan riêng phần PACF (Partial AutoCorrelation Function)

Định nghĩa 3: Hàm tự tương quan riêng phần ρ_{mm} (bậc m) là hệ số tự tương quan cuối cùng Φ_m của mô hình AR(m).

Ví dụ:

$$\begin{aligned}
 \text{AR}(1) \quad x_t &= \Phi_1 x_{t-1} + a_t & \rightarrow \rho_{11} &= \Phi_1 \\
 \text{AR}(2) \quad x_t &= \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t & \rightarrow \rho_{22} &= \Phi_2 \\
 \text{AR}(m) \quad x_t &= \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \dots + \Phi_m x_{t-m} + a_t & \rightarrow \rho_{mm} &= \Phi_m
 \end{aligned}$$

- Có thể tính được $\Phi_1, \Phi_2, \dots, \Phi_m$, ví dụ đối với AR(1), Φ_1 được tính như sau:

- Nhân 2 vế của mô hình AR(1) với x_{t-1} ta được

$$x_t x_{t-1} = \Phi_1 x_{t-1} x_{t-1} + a_t x_{t-1}$$

- Lấy kì vọng 2 vế

$$E(x_t x_{t-1}) = \Phi_1 E(x_{t-1} x_{t-1}) + E(a_t x_{t-1})$$

- Vì $E(a_t) = 0$ nên ta có phương trình tương đương

$$\gamma_1 = \Phi_1 \gamma_0 \rightarrow \Phi_1 = \rho_{11} = \gamma_1 / \gamma_0 = \rho_1 \rightarrow \text{ta sử dụng xấp xỉ } \rho_{11} = \Phi_1 \approx r_1$$

- Làm tương tự đối với mô hình AR(2) dẫn đến hệ phương trình Yule-Walker bậc 2

$$\begin{cases} \Phi_1 + \Phi_2 \rho_1 = \rho_1 \\ \Phi_1 \rho_1 + \Phi_2 = \rho_2 \end{cases}$$

hay viết dưới dạng phương trình ma trận

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

giải ra ta được nghiệm Φ_1 và Φ_2

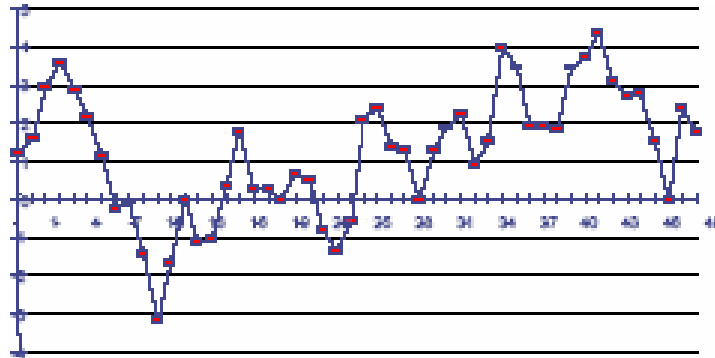
$$\Phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \text{ (bỏ qua)}$$

$$\Phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \rho_{22}$$

- Đối với mô hình AR(m): Giải hệ phương trình Yule-Walker bậc m

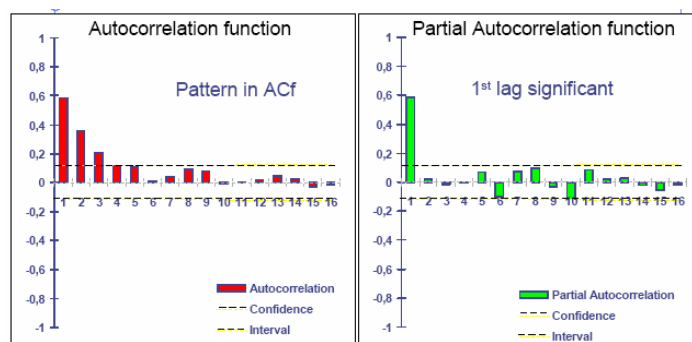
$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m-1} \\ \rho_1 & 1 & & \cdots & \rho_{m-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_m \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_m \end{pmatrix} \text{ và tính được } \Phi_m = \rho_{mm}$$

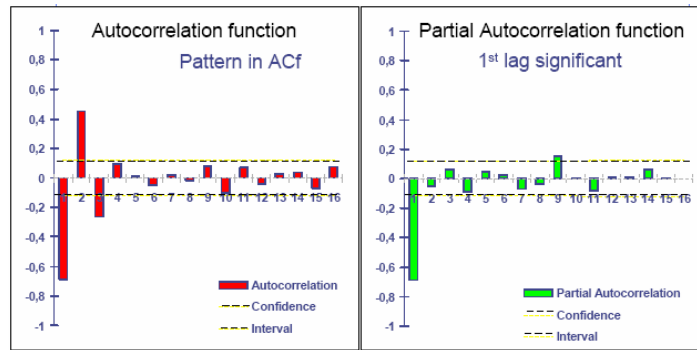
Ví dụ: Mô hình $x_t = 1.1 + 0.8 x_{t-1} + a_t$



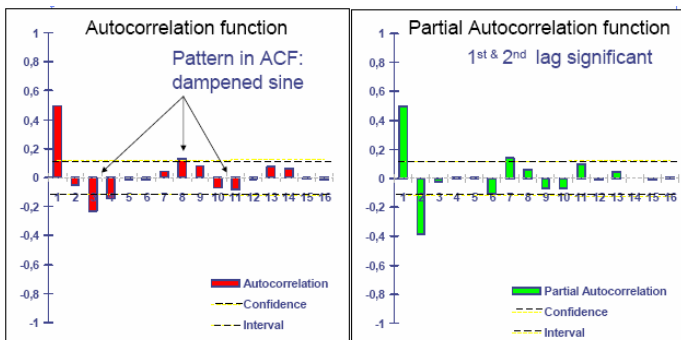
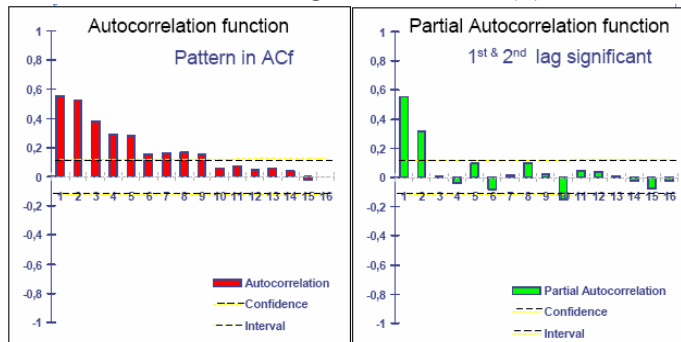
Ý nghĩa: hệ số tự tương quan riêng phần ρ_{mm} dùng để đo độ tương quan giữa x_t và x_{t+m} khi các ảnh hưởng của $m-1$ bước giữa chúng bị loại bỏ. Trong việc phân tích chuỗi thời gian, ρ_{mm} giúp nhận dạng một mô hình ARMA thích hợp cho dự báo, chẳng hạn đối với mô hình

- AR(1) chỉ có Φ_1 là đáng kể,
- AR(2) Φ_1, Φ_2 là đáng kể,
- MA(q) Φ_1, Φ_2, \dots giảm về 0 theo dạng mũ.





Nhận dạng mô hình AR(1)



Nhận dạng mô hình AR(2)

Mô hình	Trung bình μ_X	Phương sai σ_X^2	Tự tương quan ρ_k	Tự tương quan riêng ρ_{kk}
ARMA(0,0)	θ_0	σ_A^2	$\rho_k = 0, k \geq 1$	$\rho_{kk} = 0, k \geq 1$
ARMA(1,0)	$\theta_0 / (1 - \Phi_1)$	$\sigma_A^2 / (1 - \Phi_1^2)$	$\rho_k = \Phi_1^k, k \geq 1$	$\rho_{11} = \Phi_1$ $\rho_{kk} = 0, k \geq 1$
ARMA(2,0)	$\theta_0 / (1 - \Phi_1 - \Phi_2)$	$\frac{(1 - \Phi_2) \sigma_A^2}{(1 + \Phi_2)[(1 - \Phi_2)^2 - \Phi_1^2]}$	$\rho_1 = \Phi_1 / (1 - \Phi_2)$ $\rho_k = \Phi_1 \rho_{k-1} + \Phi_2 \rho_{k-2}, k \geq 1$	$\rho_{11} = \Phi_1 / (1 - \Phi_2)$ $\rho_{22} = \Phi_2$ $\rho_{kk} = 0, k > 2$

ARMA(0,1)	θ_0	$(1 + \theta_1^2) \sigma_A^2$	$\rho_k = \begin{cases} -\frac{\theta_1}{1 + \theta_1^2}, & k = 1 \\ 0, & k > 1 \end{cases}$	$\rho_{11} = -\theta_1 / (1 + \theta_1^2)$ $\rho_{kk} = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2k+2}}$
ARMA(0,2)	θ_0	$(1 + \theta_1^2 + \theta_2^2) \sigma_A^2$	$\rho_1 = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}$ $\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$ $\rho_k = 0, k > 2$	$\rho_{11} = \rho_1$ $\rho_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ ρ_{kk} giảm dần về 0
ARMA(1,1)	$\theta_0 / (1 - \Phi_1)$	$\frac{(1 + \theta_1^2 - 2\Phi_1\theta_1) \sigma_A^2}{1 - \Phi_1^2}$	$\rho_1 = \frac{(1 - \Phi_1\theta_1)(\Phi_1 - \theta_1)}{1 + \theta_1^2 - 2\Phi_1\theta_1}$ $\rho_k = \Phi_1 \rho_{k-1} = \Phi_1^2 \rho_{k-2} = \dots = \Phi_1^{k-1} \rho_1$ khi $k \geq 2$	giảm dần về 0

Bảng các thông số μ_X , σ_X^2 , ρ_k và ρ_{kk} của một số mô hình ARMA thường gặp:

3.3.4 Thống kê Q của Box-Pierce

Giả sử mô hình ARMA(p, q) được sử dụng để mô phỏng chuỗi thời gian x_1, x_2, \dots , Box và Pierce (1970) đã xây dựng một cách kiểm định giả thiết về sự phù hợp của mô hình: Các hệ số tự tương quan đầu tiên của phần dư đều bằng 0, tức là kiểm định giả thiết $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$

bằng cách sử dụng thống kê $Q = n \sum_{k=1}^m r_k^2$

với

m = số bước lớn nhất được xét,

$n = N - d$,

N = số các quan sát ban đầu,

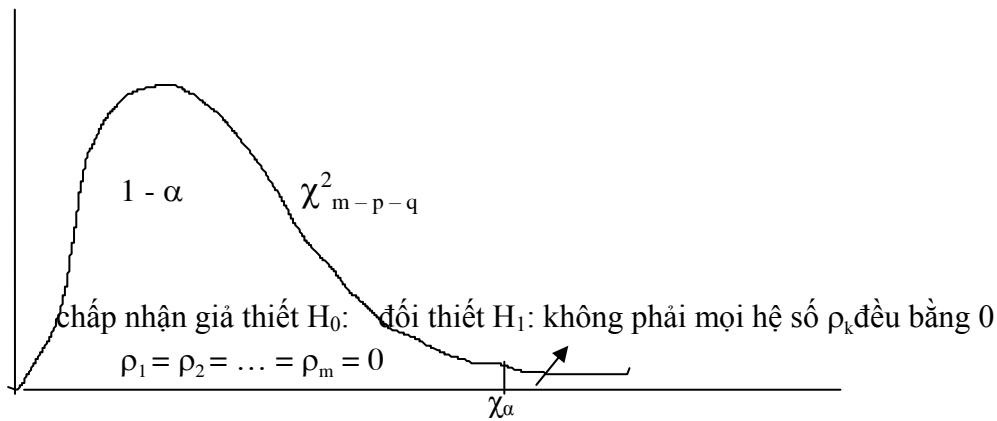
d = số lần lấy sai phân chuỗi thời gian (sẽ xét sau),

r_k = hệ số tự tương quan mẫu ở bước k của phần dư $\{e_t = x_t - f_t\}$.

Q có phân phối χ^2 với $m-p-q$ bậc tự do ($Q \sim \chi_{m-p-q}^2$)

Nếu mô hình ARMA(p, q) được chọn là thích hợp với chuỗi thời gian $\{x_t\}$ thì giá trị Q tính được là nhỏ hơn giá trị tới hạn χ_α (α là mức ý nghĩa của kiểm định) và vì thế giả thiết H_0 được chấp nhận, nếu trái lại, H_0 sẽ bị loại bỏ

Thống kê Q được sử dụng để kiểm định các hệ số tự tương quan của phần dư e_t sau khi tính được các dự báo f_t . Nên sử dụng kiểm định này cùng với việc vẽ biểu đồ các hệ số r_k của phần dư.



Sử dụng phân phối χ^2 để kiểm định giả thiết về sự phù hợp của mô hình ARMA(p,q)

3.4 Các ứng dụng của các hệ số tự tương quan

3.4.1 Kiểm tra tính ngẫu nhiên của dữ liệu và phần dư

- Có thể chứng minh rằng 95% các hệ số tự tương quan mẫu r_k nằm trong khoảng giới hạn
 $(-1.96 / \sqrt{n}, 1.96 / \sqrt{n})$
 với n là số thời điểm quan sát.
- Khi một mô hình dự báo đã được chọn, các hệ số r_k của phần dư (sai số) có thể tính được và do đó xét được tính ngẫu nhiên của chúng. Nếu chúng nằm trọn trong vùng giới hạn $\pm 2 / \sqrt{n}$, tức là không đáng kể, thì các phần dư là ngẫu nhiên. Nếu các hệ số r_k của phần dư có một định dạng xác định nào đó thì điều đó có nghĩa là mô hình dự báo đã chọn không phải là mô hình phù hợp.

3.4.2 Xác định tính dừng của chuỗi thời gian

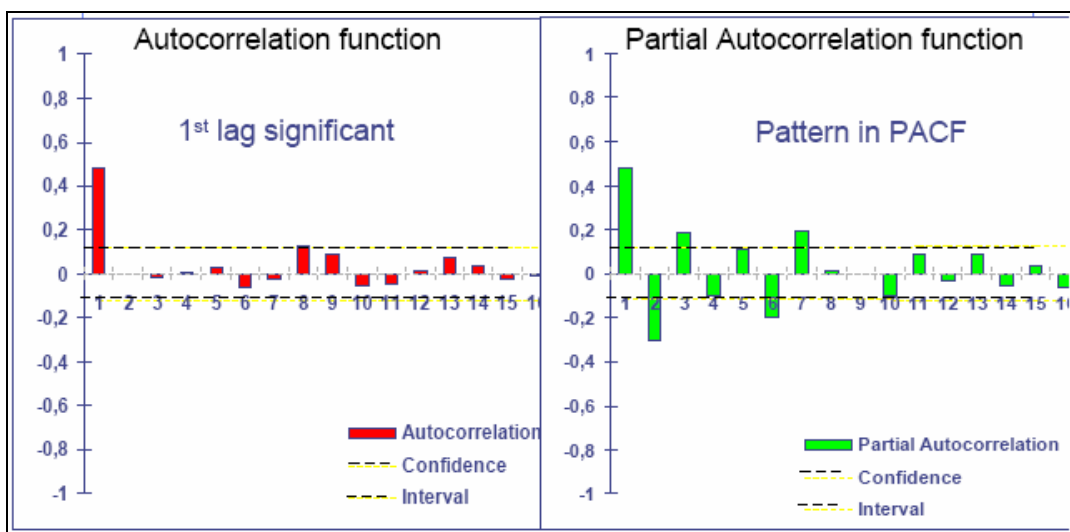
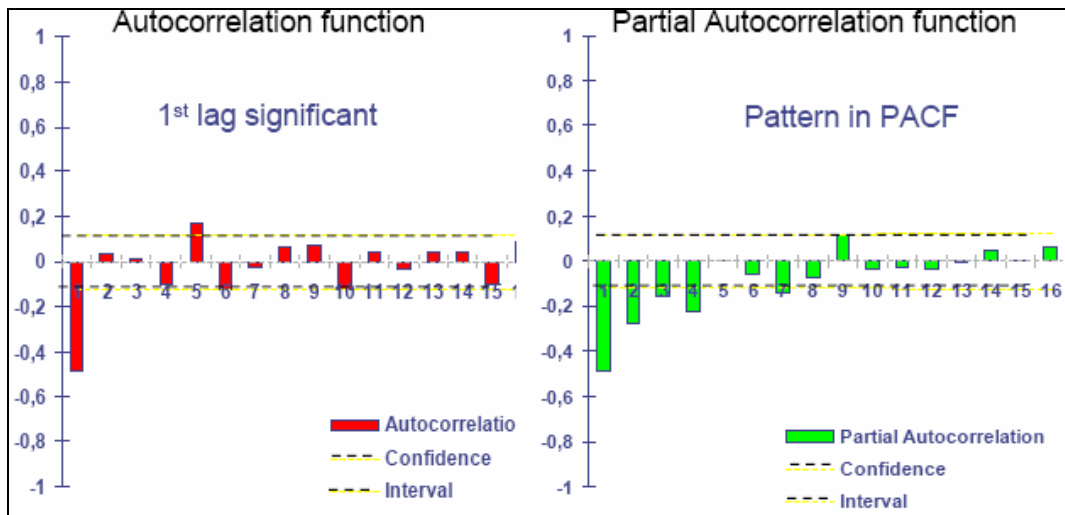
Các hệ số tự tương quan của chuỗi dừng sẽ giảm về 0 sau 2 hoặc 3 bước thời gian, trong khi đối với chuỗi không dừng, chúng vẫn còn khác 0 đáng kể sau một số bước. Trong việc phân tích chuỗi thời gian, nhận xét trên giúp người dự báo *nhận ra* một mô hình ARMA thích hợp. Chẳng hạn:

MA(1) chỉ có ρ_1 là đáng kể,

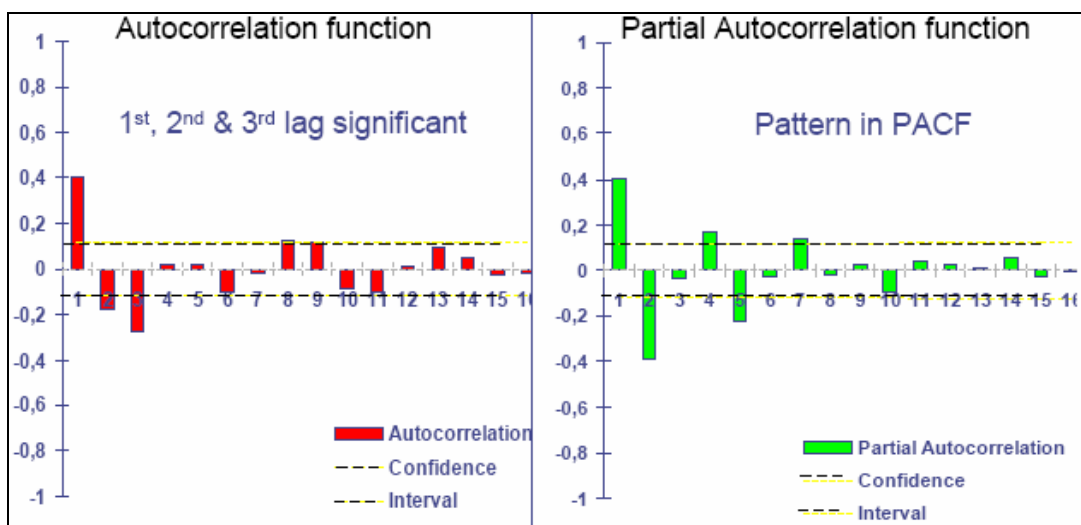
MA(2) ρ_1, ρ_2 là đáng kể,

AR(q) ρ_1, ρ_2, \dots giảm về 0 theo dạng mũ, chỉ có $\rho_{kk}, k = 1, \dots, q$ là đáng kể

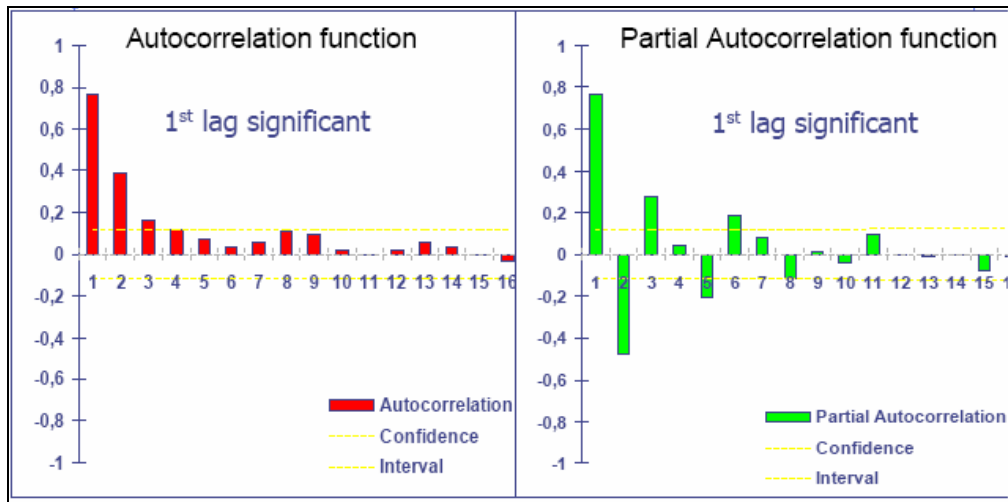
ARMA(1, 1) cả ρ_k and ρ_{kk} đều tắt dần theo dạng mũ với $k \geq 2$.



Nhận dạng mô hình MA(1)



Nhận dạng mô hình MA(3)



Nhận dạng mô hình ARMA(1, 1)

3.4.3 Loại bỏ tính không dừng của chuỗi thời gian

Bất cứ kiểu xu thế nào của các hệ số tự tương quan của $\{x_t\}$ đều do tính không xác thực của các tự tương quan chi phối. Nguyên nhân là do chuỗi thời gian đang xét chưa đảm bảo điều kiện dừng ban đầu (cả μ_X và σ_X^2 đều phải *bất biến* và do đó là *độc lập đối với t*). Vì vậy phải loại bỏ tính không dừng của dữ liệu trước khi tiến hành phân tích bất kì chuỗi thời gian nào. Điều này được thực hiện thông qua việc sai phân hóa.

Dấu hiệu nhận biết chuỗi thời gian không dừng là các hệ số r_k của nó không tắt đi nhanh chóng sau vài bước đầu. Cần phải xác định bậc của sai phân đều và sai phân thời vụ để biến đổi chuỗi đã cho thành chuỗi mới có tính dừng. Trên thực tế, bậc của sai phân đều thường là $d = 0, 1$, hoặc 2 và chỉ cần xét đến 20 hệ số r_k để kiểm tra kết quả. Chuỗi thời gian đã sai phân hóa thường được kí hiệu bởi

$$\begin{aligned} W_t &= X_t - X_{t-1} && \text{khi } d = 1 \\ &= X_t - B X_t \\ &= (1 - B) X_t && (\text{loại bỏ xu thế tuyến tính}) \end{aligned}$$

$$\begin{aligned} W_t &= (X_{t+1} - X_t) - (X_t - X_{t-1}) && \text{khi } d = 2 \\ &= X_t - 2X_{t-1} + X_{t-2} \\ &= (1 - B)^2 X_t && (\text{loại bỏ xu thế bậc hai}) \end{aligned}$$

ở đây B là toán tử lùi ($B X_t = X_{t-1}$, $B^k X_t = X_{t-k}$)

Nếu chuỗi thời gian không dừng do nguyên nhân phương sai σ_X^2 không bất biến, tức là tỉ lệ biến đổi của các quan sát không ổn định (được nhận ra khi vẽ đồ thị dữ liệu) thì cần thực hiện một phép biến đổi. Các phép biến đổi thường là

$$y_t = \log x_t$$

$$y_t = x_t^\alpha, \text{ với } \alpha = 0,5 \text{ hoặc } 0,25$$

Nếu có thể được, các phép biến đổi này nên thực hiện trước khi sai phân hóa. Các dự báo nhận được từ chuỗi đã biến đổi sẽ được chuyển lại (sử dụng phép biến đổi ngược) thành các dự báo cho chuỗi ban đầu.

3.4.4 Nhận biết tính thời vụ trong chuỗi thời gian

Yếu tố thời vụ là dạng tự lặp lại sau một số bước cố định của chuỗi thời gian. Đối với chuỗi dừng, có thể nhận ra tính thời vụ bởi dấu hiệu khác 0 đáng kể của các hệ số r_k sau một số bước thời gian $L \geq 2$ cố định. Bất cứ hệ số r_k nào khác không đáng kể đều chứng tỏ sự tồn tại của một định dạng dữ liệu. Để loại bỏ tính không dừng trong chuỗi thời gian chứa yếu tố thời vụ, ta cần sử dụng phép sai phân thời vụ. Chẳng hạn phép toán $(1 - B^{12})^D$ sẽ loại bỏ các thành phần năm r_{12} . Trên thực tế, bậc của sai phân thời vụ $D = 1$ và chỉ cần xét đến 36 hệ số r_k đầu để kiểm tra kết quả. Tránh lạm dụng phép sai phân thời vụ ($D \geq 2$) vì điều đó sẽ làm mất các quan sát thực.

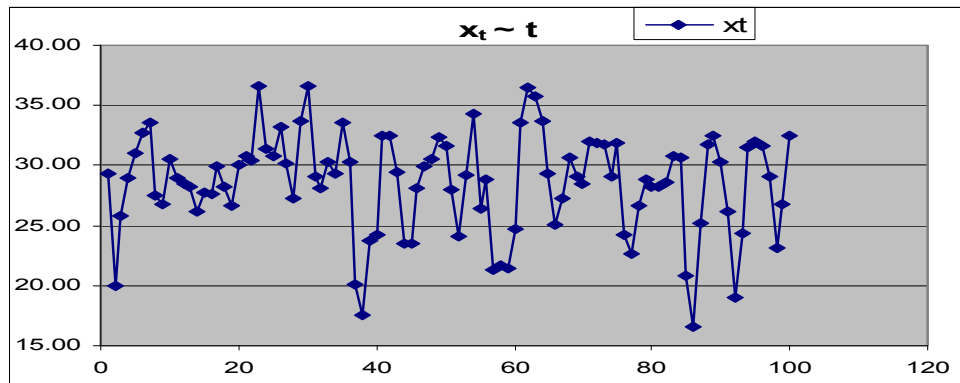
Ví dụ:

Độ dẻo của một loại sản phẩm cao su là một đặc tính quan trọng. Bảng dưới đây cho biết độ dẻo của 100 sản phẩm sau cùng. Hãy xây dựng một mô hình chuỗi thời gian ARMA phù hợp cho quá trình này (đọc từ trên xuống, từ trái sang phải)

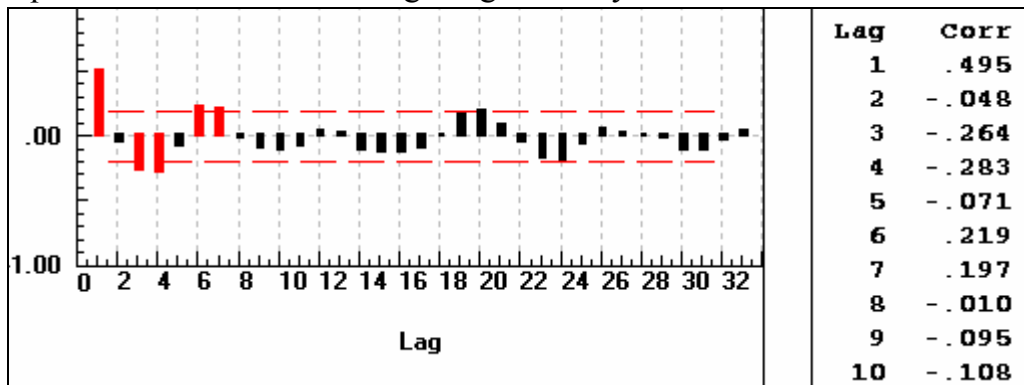
29.33	30.80	32.43	33.61	28.17
19.98	30.45	32.44	36.54	28.58
25.76	36.61	29.39	35.70	30.76
29.00	31.40	23.45	33.68	30.62
31.03	30.83	23.52	29.29	20.84
32.68	33.22	28.12	25.12	16.57
33.56	30.15	29.94	27.23	25.23
27.50	27.28	30.56	30.61	31.79
26.75	33.66	32.30	29.06	32.52
30.55	36.58	31.58	28.48	30.28
28.94	29.04	27.99	32.01	26.14
28.50	28.08	24.13	31.89	19.03
28.19	30.28	29.20	31.72	24.34
26.13	29.35	34.30	29.02	31.53
27.79	33.60	26.41	31.92	31.95
27.63	30.29	28.78	24.28	31.68
29.89	20.11	21.28	22.69	29.10
28.18	17.51	21.71	26.60	23.15
26.65	23.71	21.47	28.86	26.74
30.01	24.22	24.71	28.27	32.44

$$\rightarrow \bar{X} = 28,57$$

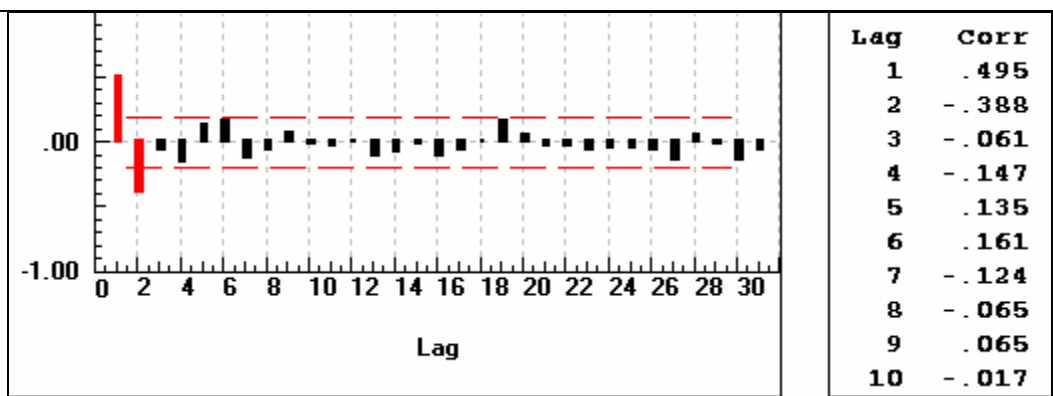
Nhìn vào đồ thị $x_t \sim t$ ta thấy chuỗi này không có tính xu thế và yếu tố thời vụ.



Các hàm tự tương quan mẫu và tự tương quan riêng phần mẫu cho chuỗi các quan sát trên được tính trong bảng dưới đây:



Các hệ số tự tương quan mẫu r_k tính bởi SIBYL



Các hệ số tự tương quan riêng phần mẫu r_{kk} tính bởi SIBYL

Từ kết quả trên có thể đoán nhận mô hình cho chuỗi này là AR(2) và tính được

$$r_1 = 0,495 \approx \Phi_1 \text{ và } r_2 = -0,048$$

Giải hệ phương trình Yule-Walker để nhận được các ước lượng cho Φ_1 và Φ_2 :

$$\begin{cases} \Phi_1 + \Phi_2 \rho_1 = \rho_1 \\ \Phi_1 \rho_1 + \Phi_2 = \rho_2 \end{cases} \Leftrightarrow \begin{cases} \Phi_1 + 0,495 \Phi_2 = 0,495 \\ 0,495 \Phi_1 + \Phi_2 = -0,048 \end{cases} \Rightarrow \begin{cases} \Phi_1 \approx 0,6871 \\ \Phi_2 \approx -0,388 \end{cases}$$

Ước lượng ban đầu của θ_0 được suy ra từ công thức $\mu_X = \theta_0 / (1 - \Phi_1 - \Phi_2)$

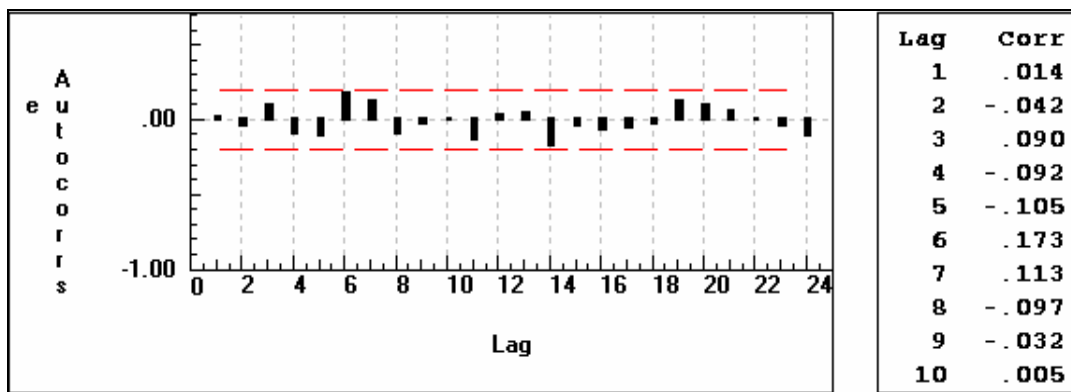
$$\rightarrow \theta_0 \approx \bar{x} (1 - \Phi_1 - \Phi_2) = 28,57(1 - 0,7 + 0,4) = 20$$

Ta có mô hình AR(2): $x_t = 20 + 0,495x_{t-1} - 0,388x_{t-2} + a_t$

Mô hình này có thể dùng để phát ra các giá trị f_t khớp với chuỗi thời gian và tính phần dư a_t . Nhìn vào biểu đồ các hệ số tự tương quan mẫu r_k của phần dư a_t tính bằng SIBYL ta thấy chúng đều không đáng kể và do đó a_t là các đại lượng ngẫu nhiên. Có thể xét thống kê Q của 10 hệ số tự tương quan đầu tiên:

$$Q = n \sum_{k=1}^m r_k^2 = 100(0,495^2 + 0,048^2 + \dots + 0,108^2) = 100 - 0,509714 = 50,97$$

So sánh $Q = 50,97$ với giá trị tới hạn $\chi_{0,05; 10-2}^2 = 15,51$ ta có thể kết luận mô hình $x_t = 20 + 0,495x_{t-1} - 0,388x_{t-2} + a_t$ là *chưa phù hợp* với chuỗi thời gian đã cho.



Các hệ số tự tương quan mẫu r_k của phần dư tính bởi SIBYL

Nhận xét:

Các tham số θ_0, Φ_1, Φ_2 của mô hình AR(2) cho chuỗi thời gian này có thể cải thiện được bằng phương pháp bình phương bé nhất (tuyến tính). Để ý rằng

$$x_t = \theta_0 + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t \text{ hay } a_t = x_t - \theta_0 - \Phi_1 x_{t-1} - \Phi_2 x_{t-2}$$

là mô hình tuyến tính đối với các thông số θ_0, Φ_1, Φ_2

→ Tìm các ước lượng cho θ_0, Φ_1, Φ_2 sao cho tổng các bình phương sai số

$$SSE = \sum_{t=3}^{100} a_t^2 = \sum_{t=3}^{100} (x_t - \theta_0 - \Phi_1 x_{t-1} - \Phi_2 x_{t-2})^2 \text{ đạt giá trị nhỏ nhất}$$

→ Hệ 3 phương trình đại số tuyến tính xác định các tham số θ_0, Φ_1, Φ_2 làm cực tiểu hàm SSE là

$$\begin{cases} n\theta_0 + \Phi_1 \sum x_{t-1} + \Phi_2 \sum x_{t-2} = \sum x_t \\ \theta_0 \sum x_{t-1} + \Phi_1 \sum x_{t-1}^2 + \Phi_2 \sum x_{t-1}x_{t-2} = \sum x_{t-1}x_t \\ \theta_0 \sum x_{t-2} + \Phi_1 \sum x_{t-1}x_{t-2} + \Phi_2 \sum x_{t-2}^2 = \sum x_{t-2}x_t \end{cases}$$

$$\iff \begin{cases} 98\theta_0 + 2795,20\Phi_1 + 2797,79\Phi_2 = 2807,66 \\ 2795,20\theta_0 + 81401,16\Phi_1 + 80643,65\Phi_2 = 80925,08 \\ 2797,79\theta_0 + 80643,65\Phi_1 + 81546,38\Phi_2 = 80074,25 \end{cases}$$

Giải hệ này ta tìm được $\theta_0 = 20,64 \quad \Phi_1 = 0,67 \quad \Phi_2 = -0,38$ là bộ thông số tối ưu

3.5 Các mô hình ARIMA

3.5.1 Các mô hình ARIMA không có tính thời vụ

Dạng mô hình chung nhất mô phỏng quá trình hỗn hợp (cả AR và MA) là

$$w_t - \Phi_1 w_{t-1} - \Phi_2 w_{t-2} - \dots - \Phi_p w_{t-p} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + \theta_0 \quad (3.7)$$

hay (dạng toán tử)

$$(1 - \Phi_1 B - \dots - \Phi_p B^p)w_t = (1 - \theta_1 B - \dots - \theta_q B^q)a_t + \theta_0 \quad (3.8)$$

được gọi là mô hình ARIMA bậc (p,d,q) hay ARIMA(p, d, q),

với

$$w_t = (1-B)^d X_t$$

p = bậc của quá trình AR

d = bậc của sai phân đều để $\{w_t\}$ là chuỗi dừng

q = bậc của quá trình MA

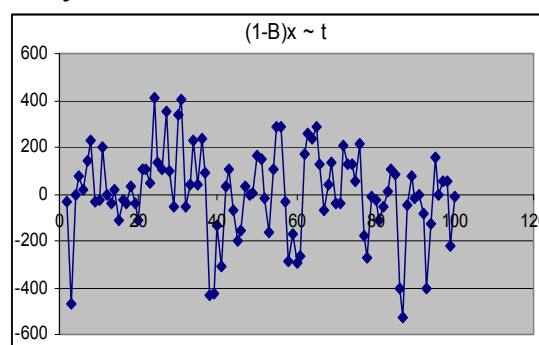
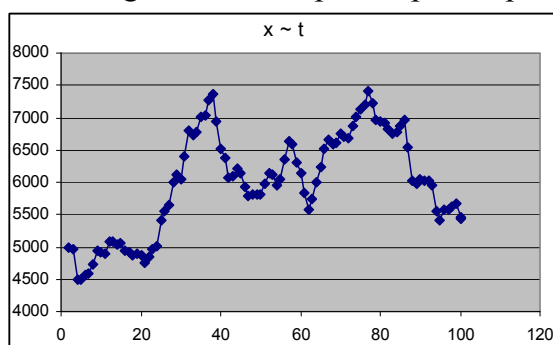
I: tích hợp (Integrated), vì $\{x_t\}$ đã được sai phân hóa để tạo ra chuỗi dừng.

θ_0 là hằng số, sẽ bằng 0 nếu và chỉ nếu trung bình của $\{w_t\}$ bằng 0.

Nói cách khác: Các mô hình chứa x_t (d = 0) gọi là các mô hình ARMA.

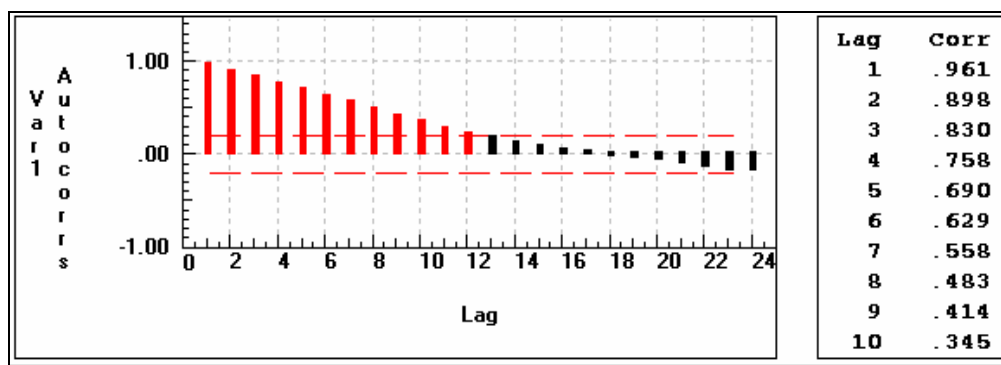
Các mô hình chứa w_t (d > 0) gọi là các mô hình ARIMA.

Ví dụ : Nhu cầu hàng tuần về nhựa (tính bằng tấn) của một công ty sản xuất dây cáp điện của 100 tuần liên tiếp được cho trong bảng dưới đây. Hãy xây dựng một mô hình chuỗi thời gian ARIMA phù hợp cho quá trình này



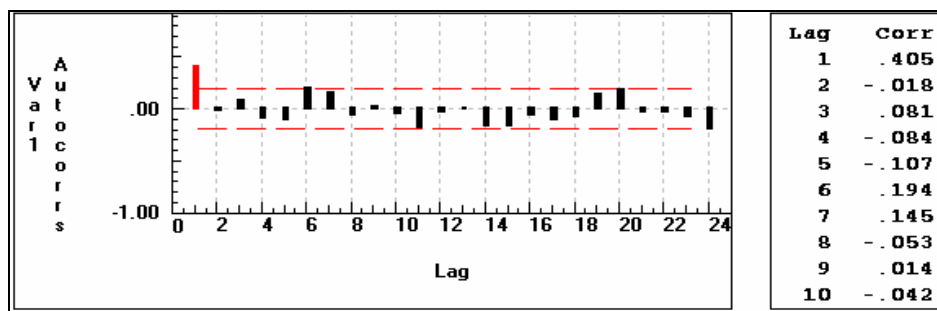
5000	5657	6132	7411
4965	6010	6111	7233
4496	6109	5948	6958
4491	6052	6056	6950
4566	6391	6342	6927
4585	6798	6626	6814
4724	6740	6591	6757
4951	6778	6302	6765
4917	7005	6132	6870
4888	7045	5837	6954
5087	7279	5572	6551
5082	7367	5744	6022
5039	6934	6005	5974
5054	6506	6239	6052
4940	6374	6523	6033
4913	6066	6652	6030
4871	6102	6585	5944
4901	6204	6622	5543
4864	6138	6754	5416
4750	5938	6712	5571
4856	5781	6675	5571
4959	5813	6882	5627
5004	5811	7011	5679
5415	5818	7140	5455
5550	5982	7197	5443

Nhìn vào đồ thị ta thấy chuỗi quan sát có khuynh hướng lệch ở vài chỗ nên không có được trị trung bình ổn định. Nếu vẽ đồ thị biểu diễn hàm tự tương quan ta thấy nó giảm về 0 rất chậm. Vậy chuỗi thời gian này không dừng.

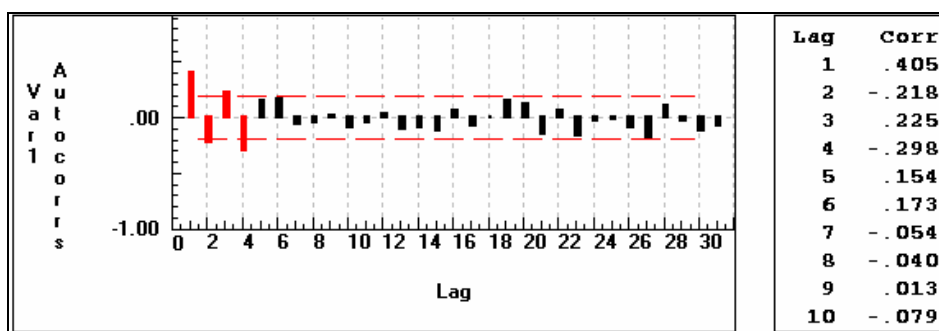


Lấy sai phân bậc nhất chuỗi x_t ta được chuỗi $w_t = (1 - B)x_t$. Đồ thị biểu diễn chuỗi w_t cho thấy đây là chuỗi dừng, không có tính thời vụ và có trung bình $\bar{w} = 4,47$.

Có thể coi hằng số $\theta_0 \approx 4,47$. Các hàm tự tương quan mẫu và tự tương quan riêng phần mẫu cho chuỗi w_t trên được tính trong bảng dưới đây:



Các hệ số tự tương quan mẫu r_k của chuỗi w_t tính bởi SIBYL



Các hệ số tự tương quan riêng phần mẫu r_{kk} của chuỗi w_t tính bởi SIBYL

Nhìn vào các đồ thị trên ta thấy các hệ số tự tương quan mẫu r_k là không đáng kể sau bước $k = 1$ còn các hệ số tự tương quan riêng phần mẫu r_{kk} giảm về 0 sau 4 bước đầu. Điều đó chỉ ra rằng có thể đoán nhận mô hình MA(1) là phù hợp cho chuỗi w_t hay mô hình ARIMA(0,1,1) là phù hợp cho chuỗi x_t

$$w_t = 4,47 + a_t - \theta_1 a_{t-1} \text{ hay } x_t = x_{t-1} + 4,47 + a_t - \theta_1 a_{t-1}$$

Tham số θ được tính từ công thức $\rho_1 = -\theta/(1+\theta^2)$ thay $\rho_1 \approx r_1 = 0,41$ ta tìm được

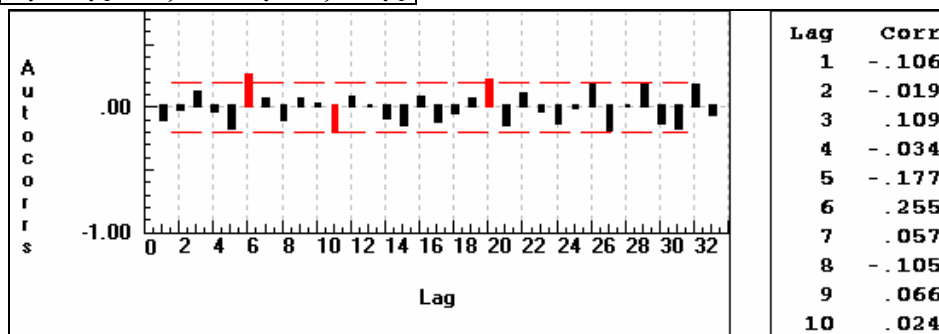
$$\theta_1 = -1,92 \text{ và } \theta_2 = -0,52$$

ta chọn $\theta_1 = -0,52$ thỏa mãn điều kiện $|\theta| < 1$ của mô hình ARMA(1,1) và nhận được mô hình thử nghiệm

$$w_t = 4,47 + a_t + 0,52a_{t-1}$$

hay

$$x_t = x_{t-1} + 4,47 + a_t + 0,52a_{t-1}$$



Các hệ số tự tương quan mẫu r_k của phần dư a_t tính bởi SIBYL

Nhìn vào biểu đồ các hệ số tự tương quan mẫu r_k của phần dư a_t tính bởi SIBYL ta thấy có 3 giá trị đáng kể là r_6, r_{11} và r_{19} do đó a_t chưa hẳn là các đại lượng ngẫu nhiên. Có thể xét thêm thống kê Q của 10 hệ số tự tương quan đầu tiên:

$$Q = n \sum_{k=1}^m r_k^2 = (100-1)(0,106^2 + 0,019^2 + \dots + 0,024^2) = 99 \cdot 0,1402 = 13,88$$

So sánh $Q = 13,88$ với giá trị tới hạn $\chi^2_{0,05; 10-1} = 16,92$ ta có thể kết luận mô hình $x_t = x_{t-1} + 4,47 + a_t + 0,52 a_{t-1}$ là phù hợp với chuỗi thời gian đã cho.

Để cải thiện giá trị của tham số θ_1 , người ta sử dụng phương pháp bình phương bé nhất phi tuyến, tìm θ_1 tối ưu theo mục tiêu làm cực tiểu hàm SSE. Nhận xét rằng từ mô hình MA(1)

$$w_t = 0,47 + a_t - \theta_1 a_{t-1}, t = 2, 3, \dots$$

đặt

$$z_t = w_t - 0,47 \text{ ta có mô hình}$$

$$z_t = a_t - \theta_1 a_{t-1}$$

hay

$$a_t = z_t + \theta_1 a_{t-1}$$

Sai số dự báo được cho bởi phần dư $a_t = z_t - f_t$. Giả sử \hat{a}_t là xấp xỉ của a_t , $t \leq n$, phụ thuộc vào sự lựa chọn thông số θ_1 và thỏa mãn phép đệ quy

$$\hat{a}_t = z_t + \theta_1 \hat{a}_{t-1} \text{ với các dự báo lùi } z_t = 0 \text{ và } \hat{a}_t = 0 \text{ với mọi } t \leq 1$$

$$\rightarrow \hat{a}_1 = 0$$

$$\hat{a}_2 = z_2 + \theta_1 \hat{a}_1 = z_2$$

$$\hat{a}_3 = z_3 + \theta_1 \hat{a}_2 = z_3 + \theta_1 z_2$$

$$\dots \dots \dots \hat{a}_n = z_n + \theta_1 \hat{a}_{n-1} = z_n + \theta_1(z_{n-1} + \theta_1 z_{n-2}) = \dots$$

$$= z_n + \theta_1 z_{n-1} + \theta_1^2 z_{n-2} + \dots + \theta_1^{n-2} z_2$$

Vậy \hat{a}_t là hàm đa thức của θ_1 . Khi đó thông số tối ưu θ_1^* là điểm làm cực tiểu hàm

$$SSE = \sum_{t=2}^n a_t^2 \text{ theo phương pháp bình phương bé nhất phi tuyến. Giá trị ước lượng ban}$$

đầu $\theta_1 = -0,52$ sẽ được dùng làm điểm xuất phát trong quá trình dò tìm.

3.5.2 Các mô hình ARIMA có tính thời vụ

Dạng mô hình thời vụ chung nhất với chu kỳ $L = 12$ mô phỏng quá trình hỗn hợp (cả AR và MA) gọi là mô hình ARIMA(p,d,q) (P,D,Q)₁₂

$$(1 - \Phi_1 B - \dots - \Phi_p B^p) (1 - \Lambda_1 B^{12} - \dots - \Lambda_P B^{12P}) w_t$$

$$= (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \gamma_1 B^{12} - \dots - \gamma_Q B^Q) a_t + \theta_0 \quad (3.9)$$

với

$$w_t = (1 - B)^d (1 - B^{12})^D x_t$$

p, d, q và θ_0 là tương tự như trên

P = bậc của quá trình thời vụ AR

D = bậc sai phân của quá trình thời vụ cần cho tính dừng

Q = bậc của quá trình thời vụ MA

I : tích hợp (Integrated), vì $\{x_t\}$ đã được sai phân hóa để tạo ra chuỗi dừng..

Trên thực tế, các mô hình chung nhất là

ARIMA(1,d,0) (1,D,0)₁₂,

ARIMA(1,d,0) (0,D,1)₁₂,

ARIMA(0,d,1) (1,D,0)₁₂,

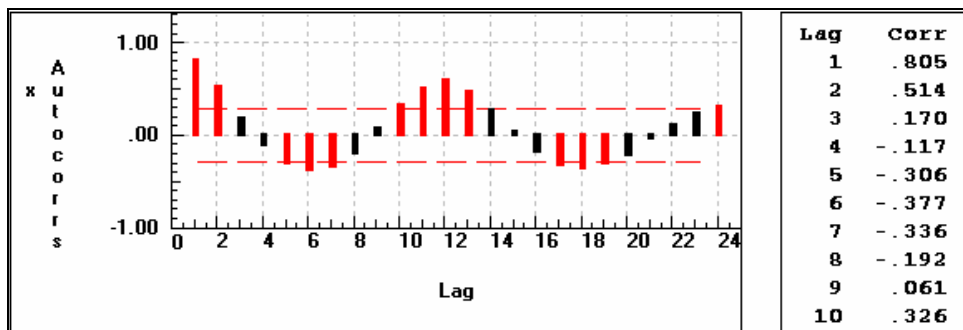
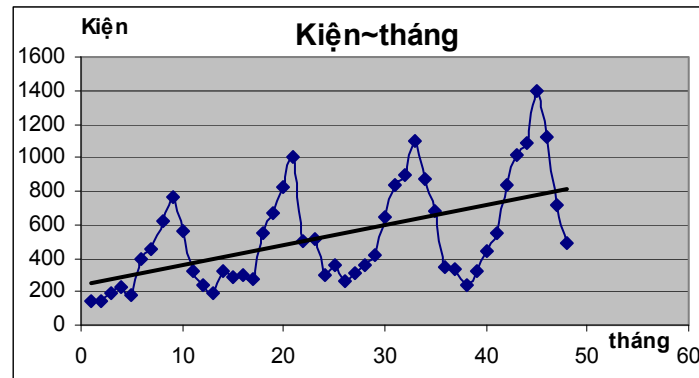
ARIMA(0,d,1) (0,D,1)₁₂

với d và D nhận các giá trị 0 hoặc 1.

Ví dụ 1:

Cho dãy 48 số liệu một loại nước giải khát đóng chai bán ra hàng tháng (tính theo kiện) của một hãng trong 4 năm liền. Hãy xây dựng một mô hình chuỗi thời gian ARIMA phù hợp cho quá trình này.

Tháng	1994	1995	1996	1997
1	143	189	359	332
2	138	326	264	244
3	195	289	315	320
4	225	293	361	437
5	175	279	414	544
6	389	552	647	830
7	454	674	836	1011
8	618	827	901	1081
9	770	1000	1104	1400
10	564	502	874	1123
11	327	512	683	713
12	235	300	352	487

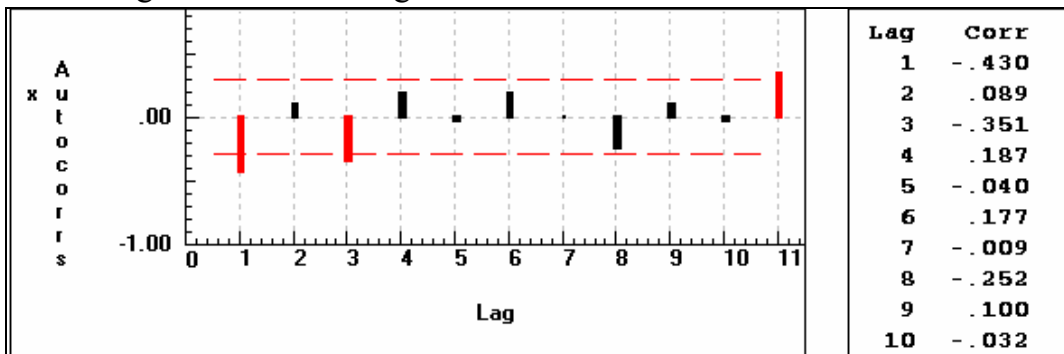


Hàm tự tương quan mẫu của chuỗi x_t

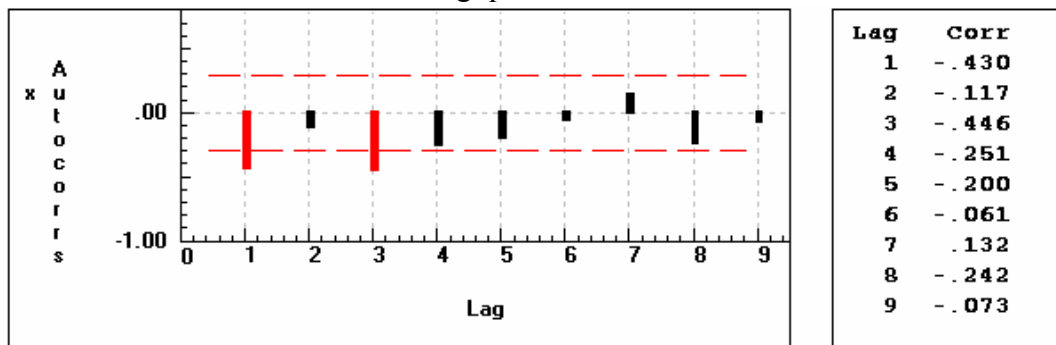
Từ đồ thị $x \sim t$ và hàm tự tương quan mẫu của x_t ta thấy chuỗi thời gian đã cho không dừng và có tính thời vụ với độ dài $L = 12$ do đó phép biến đổi

$$w_t = (1 - B)(1 - B^{12})x_t$$

được sử dụng để tạo chuỗi dừng.



Hàm tự tương quan mẫu của chuỗi w_t



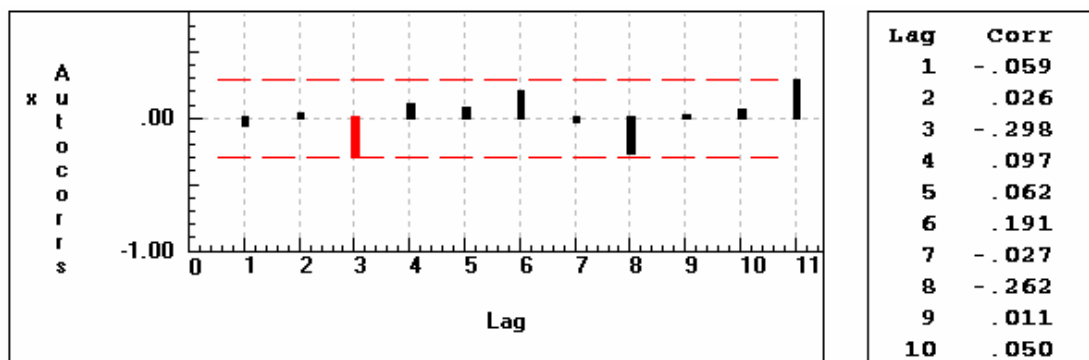
Hàm tự tương quan riêng phần mẫu của chuỗi w_t

Từ hai đồ thị trên ta thấy cả r_k lẫn r_{kk} đều giảm về 0 theo k nên có thể chọn mô hình thử nghiệm của w_t là ARMA(1,1), tức là

$$(1 - \Phi_1 B) w_t = (1 - \theta_1 B) a_t \leftrightarrow w_t = \Phi_1 w_{t-1} + a_t - \theta_1 a_{t-1}$$

hay $x_t = (1 + \Phi_1) x_{t-1} - \Phi_1 x_{t-2} - (1 + \Phi_1) x_{t-13} + \Phi_1 x_{t-14} + a_t - \theta_1 a_{t-1}$

Sử dụng dữ liệu đã cho ta tính được $\Phi_1 = 0,94$ và $\theta_1 = 0,14$



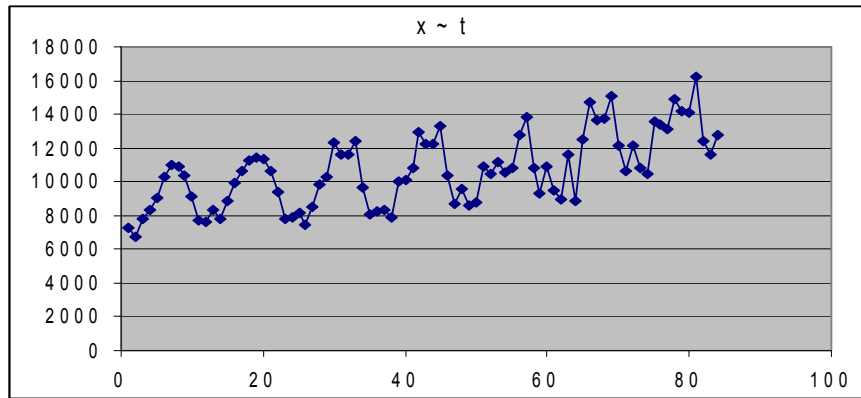
Hàm tự tương quan mẫu của phần dư a_t

Nhìn vào biểu đồ các hệ số tự tương quan mẫu r_k của phần dư a_t tính bởi SIBYL ta thấy chỉ có r_3 chạm vào đường giới hạn cho phép, do đó có thể xem a_t là các đại lượng ngẫu nhiên. Ta có thể kết luận mô hình ARIMA(1,1,1) (0,1,0)₁₂ là phù hợp với chuỗi thời gian x_t đã cho.

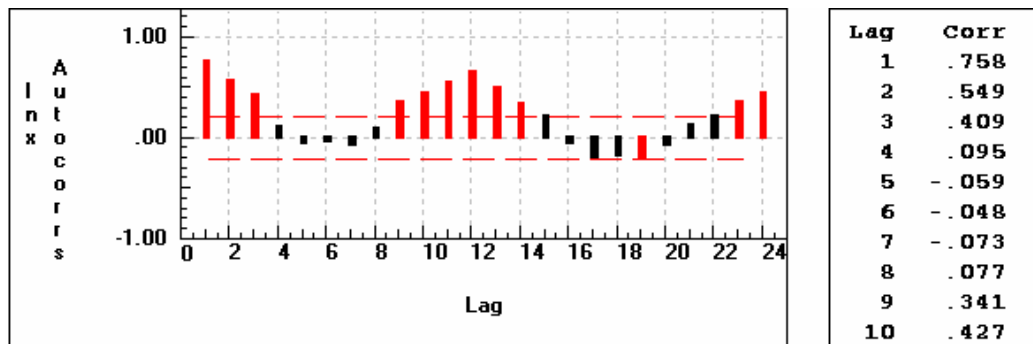
Ví dụ 2:

Quãng đường bay hàng tháng (đơn vị tính là triệu km) của một hãng hàng không trong 7 năm liên được cho trong bảng dưới đây. Hãy tìm mô hình ARIMA phù hợp với chuỗi thời gian này.

Tháng	1994	1995	1996	1997	1998	1999	2000
1	7269	8350	8186	8334	8639	9491	10840
2	6775	7829	7444	7899	8772	8919	10436
3	7819	8829	8484	9994	10894	11607	13589
4	8371	9948	9864	10078	10455	8852	13402
5	9069	10638	10252	10801	11179	12537	13103
6	10268	11253	12282	12953	10588	14759	14933
7	11030	11424	11637	12222	10794	13667	14147
8	10882	11391	11577	12246	12770	13731	14057
9	10333	10665	12417	13281	13812	15110	16234
10	9109	9396	9637	10366	10857	12185	12389
11	7685	7775	8094	8730	9290	10645	11594
12	7602	7933	8280	9614	10925	12161	12772

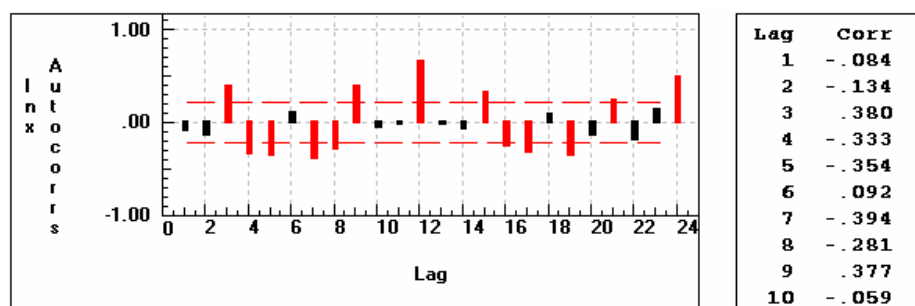


Từ đồ thị $x \sim t$ ta thấy chuỗi thời gian đã cho không dừng cả về trung bình và phương sai nên ta sử dụng phép biến đổi $y_t = \ln x_t$ để làm mất tính không dừng về độ lệch



Hàm tự tương quan mẫu của chuỗi y_t

Tuy nhiên đồ thị hàm tự tương quan mẫu cho thấy chuỗi y_t chưa phải là chuỗi dừng về trung bình, ta dùng phép biến đổi sai phân đều $z_t = y_t - y_{t-1}$

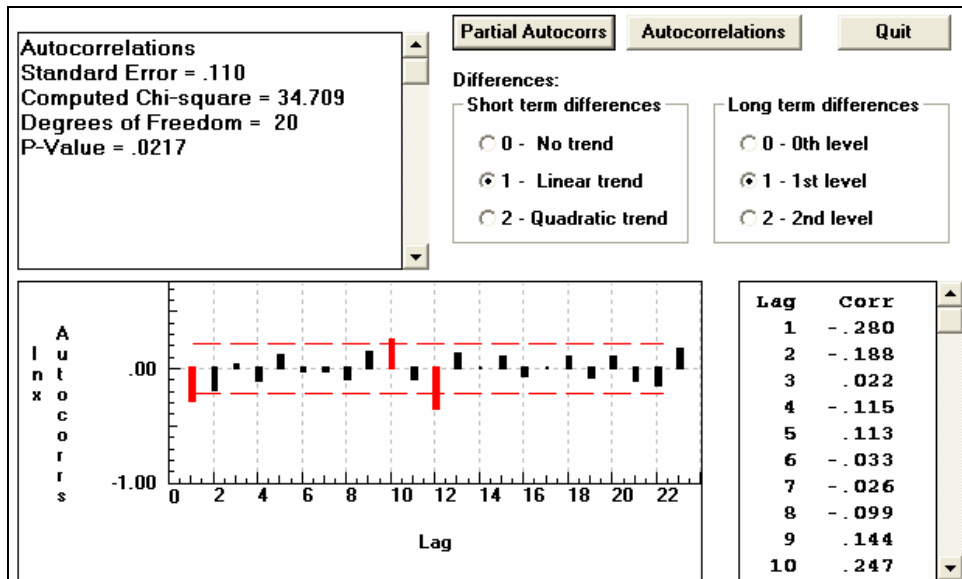


Hàm tự tương quan mẫu của chuỗi z_t

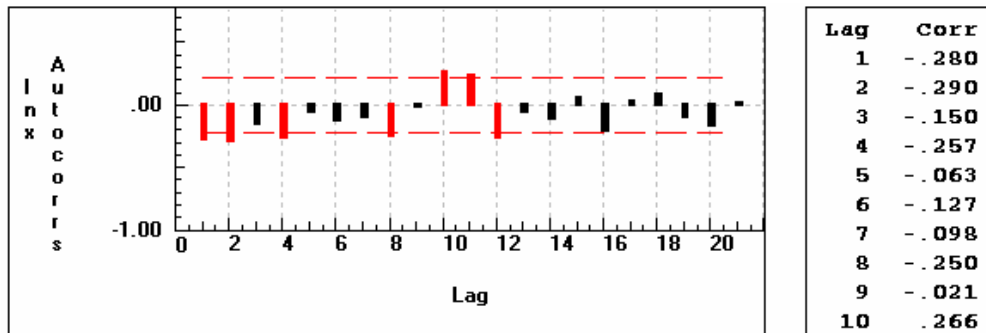
Hàm tự tương quan mẫu của chuỗi z_t vẫn giảm về 0 rất chậm do tính thời vụ, thể hiện ở độ lớn của tham số r_{12}, r_{24}, \dots . Điều đó cho thấy phép sai phân thời vụ

$$w_t = (1 - B^{12}) z_t = z_t - z_{t-12} = (y_t - y_{t-1}) - (y_{t-12} - y_{t-13})$$

là cần thiết để biến z_t thành chuỗi dừng w_t



Hàm tự tương quan mẫu của w_t



Hàm tự tương quan riêng phần mẫu của w_t

Từ hai đồ thị trên ta thấy mô hình thử nghiệm cho chuỗi w_t nên là

$$ARIMA(0,0,1) (0,0,1)_{12}$$

hay

$$w_t = (1 - \theta_1 B) (1 - \gamma_{12} B^{12}) a_t$$

Nói cách khác, mô hình thử nghiệm cho chuỗi y_t nên là $ARIMA(0,1,1) (0,1,1)_{12}$

$$\leftrightarrow (1 - B) (1 - B^{12}) y_t = (1 - \theta_1 B) (1 - \gamma_{12} B^{12}) a_t$$

hay

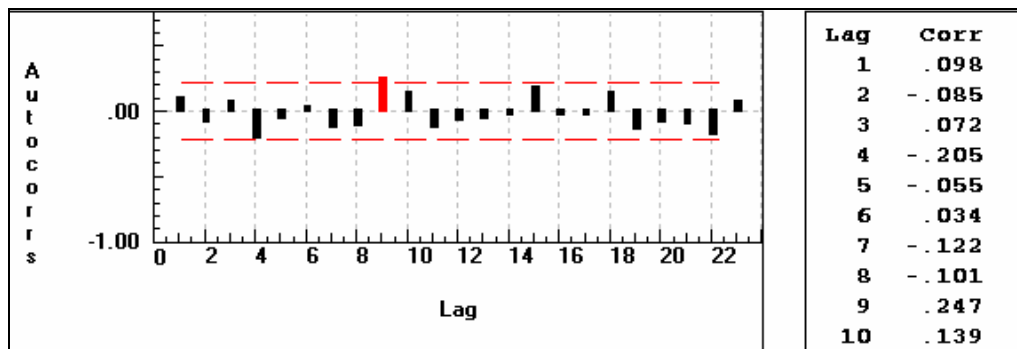
$$y_t - y_{t-1} - y_{t-12} + y_{t-13} = a_t - \theta_1 a_{t-1} - \gamma_{12} a_{t-12} + \theta_1 \gamma_{12} a_{t-13}$$

Sử dụng phương pháp bình phương bé nhất (phi tuyến), người ta tính được các hệ số của mô hình là $\theta_1 = 0,67$ và $\gamma_{12} = 0,56$, tức là

$$w_t = (1 - 0,67B) (1 - 0,56 B^{12}) a_t = a_t - 0,67 a_{t-1} - 0,56 a_{t-12} + 0,38 a_{t-13}$$

hay

$$y_t - y_{t-1} - y_{t-12} + y_{t-13} = a_t - 0,67 a_{t-1} - 0,56 a_{t-12} + 0,38 a_{t-13}$$



Hàm tự tương quan mẫu của phần dư a_t

Nhìn vào biểu đồ các hệ số tự tương quan mẫu r_k của phần dư a_t tính bởi SIBYL ta thấy chỉ có r_9 vượt ra khỏi đường giới hạn cho phép, do đó a_t chưa hẳn là các đại lượng ngẫu nhiên. Xét thêm thống kê Q của 10 hệ số tự tương quan đầu tiên

$$Q = n \sum_{k=1}^{10} r_k^2 = (72 - 2) (0,098^2 + 0,085^2 + \dots + 0,139^2) = 72 \times 0,1736 = 12,50$$

So sánh $Q = 12,5$ với giá trị tới hạn $\chi^2_{0,05; 10-2} = 15,51$ ta có thể kết luận mô hình

ARIMA(0,1,1) (0,1,1)₁₂ là phù hợp với chuỗi $y_t = \ln x_t$ đã cho.

3.6 BÀI TẬP CHƯƠNG 3

1. Chứng minh rằng đối với mô hình AR(1) ta có

$$\sigma_X^2 = \frac{\sigma_A^2}{1 - \Phi_1^2}$$

và do đó điều kiện dừng của mô hình là $-1 < \Phi_1 < 1$;

2. Xét mô hình AR(1): $x_t = 0,7 x_{t-1} + 30 + a_t$

i> $\{x_t\}$ có phải là chuỗi dừng không?

ii> Cho $x_{20} = 35$, hãy mô phỏng 5 giá trị tiếp theo của quá trình này biết $a_t \sim N(0, 4)$

3. Xét mô hình AR(1): $x_t = \Phi_1 x_{t-1} + a_t$. Chứng minh rằng mô hình này tương đương với mô hình

$$x_t = a_t + \Phi_1 a_{t-1} + \Phi_1^2 a_{t-2} + \Phi_1^3 a_{t-3} + \dots$$

4. Cho 20 quan sát của chuỗi thời gian (đọc theo hàng)

4.2	5.8	6.9	7.62	5.57	3.34	2	1.7	2.02	2.71
3.63	5.18	7.11	8.26	7.95	6.78	5.07	5.04	6.02	7.61

i> Tìm ước lượng Φ_1 cho mô hình AR(1) $x_t = \Phi_1 x_{t-1} + a_t$ của chuỗi thời gian trên bằng cách cực tiểu tổng bình phương các phần dư a_t ($f_t = \Phi_1 x_{t-1}$)

ii> Vẽ đồ thị phần dư $a_t \sim t$. Từ đó có thể kết luận a_t là các đại lượng ngẫu nhiên không? Tại sao?

5. Xét mô hình MA(1): $x_t = a_t - \theta_1 a_{t-1}$.

i> Chứng minh rằng mô hình này tương đương với mô hình

$$x_t = a_t - \theta_1 x_{t-1} - \theta_1^2 x_{t-2} - \theta_1^3 x_{t-3} - \dots$$

ii> Tìm khoảng giá trị chấp nhận được của θ_1 để mô hình này khả nghịch.

iii> Tìm ước lượng của θ_1 nếu biết r_1

6. Hiện thị các mô hình sau đây theo nghĩa x_t được biểu diễn theo các số hạng trước đó của nó và của các nhiễu động a_t

ARIMA(0,0,0)	ARIMA(1,0,0)	ARIMA(0,0,1)	ARIMA(0,0,2)
ARIMA(1,0,1)	ARIMA(1,1,0)		
ARIMA(0,1,1)	ARIMA(1,1,1)	ARIMA(2,1,0)	
ARIMA(2,1,1)	ARIMA(1,1,2)		

7. Doanh thu hàng quý của một tổng công ty trong 5 năm liền là (đọc theo hàng)

25758	26067	26390	26710	27023	27338	27657	27967	28290	28598
28922	29519	30209	30611	31021	31235	31551	32071	32268	32394

i> Dữ liệu đã cho có tính xu thế và yếu tố thời vụ không? Tại sao?

ii> Sử dụng phép biến đổi $w_t = (1 - B) x_t$

Chuỗi $\{w_t\}$ có phải là chuỗi dừng không?

iii> Tìm mô hình ARIMA phù hợp với chuỗi $\{x_t\}$

8. Số khách du lịch đến Thái Lan hàng tháng qua đường hàng không (tính theo ngàn người) của 3 năm gần đây là

Tháng	2000	2001	2002
1	112	115	145
2	118	126	150
3	132	141	178
4	129	135	163
5	121	125	172
6	135	149	178
7	148	170	199
8	148	170	199
9	136	158	184
10	119	133	162
11	104	114	146
12	118	140	168

i> Dữ liệu đã cho có tính xu thế và yếu tố thời vụ không? Tại sao?

ii> Sử dụng phép biến đổi

$$w_t = (1 - B)(1 - B^{12}) x_t$$

iii> Chuỗi $\{w_t\}$ có phải là chuỗi dừng không?

iv> Tìm mô hình ARIMA phù hợp với chuỗi x_t

9. Số người (ngàn) truy cập vào trang web của một công ty trong 5 năm là

Tháng	1996	1997	1998	1999	2000
1	9.56	41.47	72.66	62.61	96.30
2	12.48	46.14	71.25	69.07	96.09
3	13.64	52.62	65.48	77.36	99.27
4	18.80	59.01	62.68	80.39	104.77
5	25.04	60.20	56.60	85.28	105.51
6	30.33	58.53	49.90	84.44	105.19
7	34.08	56.98	49.82	86.59	109.16
8	40.10	57.82	51.87	88.05	110.78
9	42.40	60.50	57.74	90.83	115.77
10	41.36	63.29	58.24	93.05	122.75
11	39.25	66.55	58.31	94.65	126.85
12	38.20	68.65	59.91	96.66	132.57

Tìm mô hình ARIMA phù hợp với chuỗi $\{x_t\}$

4 CHƯƠNG 4: CÁC PHƯƠNG PHÁP DỰ BÁO CỦA BOX-JENKINS

Đây là quá trình xây dựng mô hình bằng cách chọn ra một mô hình từ lớp các mô hình ARIMA. Kỹ thuật của **Box-Jenkins** là tiến trình xây dựng mô hình chứ không chỉ đơn thuần là tiến trình tìm kiếm mô hình phù hợp bởi vì các mô hình được xác định trên cơ sở dữ liệu chứ không phải trên cơ sở giả thiết.

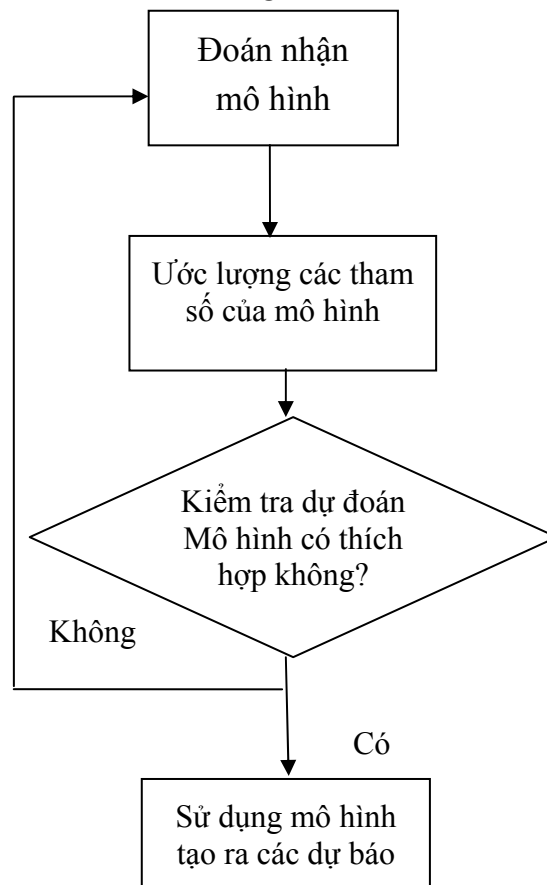
4.1 Các khâu chính trong phương pháp Box-Jenkins

Bước 1- Đoán nhận thăm dò: dữ liệu quá khứ được sử dụng để nhận dạng thử một mô hình ARIMA thích hợp.

Bước 2- Ước lượng: dữ liệu quá khứ được sử dụng để ước lượng các tham số của mô hình thử nghiệm.

Bước 3- Kiểm tra dự đoán: các đánh giá khác nhau được dùng để kiểm tra sự thích hợp của mô hình thử nghiệm, và nếu cần thiết, gợi ý một mô hình tốt hơn rồi sau đó mô hình này lại được xem như một mô hình thử nghiệm mới.

Bước 4- Dự báo: Khi đã chọn được mô hình cuối cùng, nó được sử dụng để dự báo các giá trị tương lai của chuỗi thời gian.



Trên thực tế, nhiều chuỗi thời gian có thể được biểu diễn bằng những mô hình đơn giản. Mô hình với số tham số ít nhất thường được ưa chuộng hơn. Thông thường đối với các mô hình ARMA(p,q) ta chỉ cần xét $p \leq 2$ và, hoặc $q \leq 2$. Có thể cải thiện việc biểu diễn mô hình bằng cách sử dụng một phép biến đổi dữ liệu gốc phù hợp. Dữ liệu đã được biến đổi, nếu chưa có tính dừng, sẽ được sai phân hóa cho đến khi đạt được tính dừng bởi vì ta *bắt buộc phải làm việc với chuỗi thời gian dừng*. Các hệ số tự tương quan của chuỗi thời gian dừng sẽ tắt rất nhanh và không có cấu trúc nào cả tức là không có bất kì một dấu hiệu định dạng nào.

4.2 Các nguyên tắc lựa chọn mô hình ARIMA(p,d,q) phù hợp

Phương pháp Box-Jenkins được xem là một trong những kỹ thuật có hiệu quả cao trong việc phát ra các dự báo chính xác và tin cậy. Sức mạnh của nó là ở chỗ nó đưa ra những thông tin giúp nhà phân tích chuỗi thời gian lựa chọn mô hình phù hợp với dữ liệu quan sát được. Đối với các phương pháp khác, nhà phân tích giả thiết một mô hình nào đó rồi tiến hành ước lượng các tham số của mô hình.

Trong giai đoạn đầu tiên, ta nhận dạng một mô hình thử nghiệm bằng cách so sánh các hàm tự tương quan mẫu và tự tương quan riêng phần mẫu của chuỗi thời gian dừng với các hàm tự tương quan và tự tương quan riêng lý thuyết của các mô hình ARMA.

Trong khi rút ra nhận xét về các hàm tự tương quan và tự tương quan riêng lý thuyết của các mô hình ARMA khác nhau, cần nhớ rằng

i> Chuỗi thời gian $\{w_t\}$ được xét là dừng theo nghĩa có trung bình không đổi (thường là 0) và phương sai bất biến.

iii> Phần dư a_t thường là đại lượng ngẫu nhiên có phân phối chuẩn $N(0, \sigma_A^2)$

iii> Các a_t là độc lập, tức là

$$\text{cov}(a_t, a_{t-k}) = 0 \rightarrow E(a_t \cdot a_{t-k}) = 0$$

iv> Các a_t không có tương quan với các quan sát trước đó, tức là

$$E(a_t \cdot w_{t-k}) = 0 \text{ với } k > 0$$

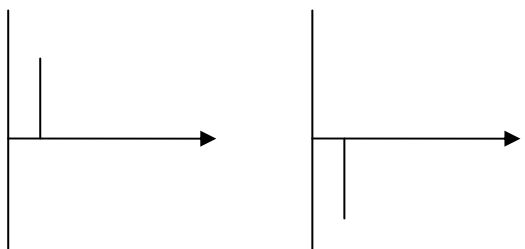
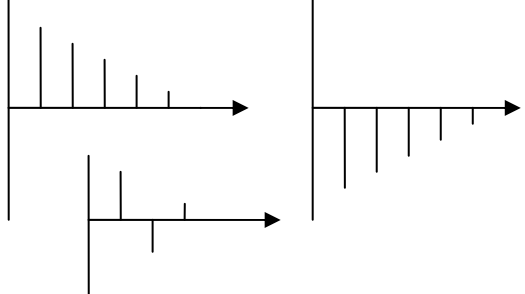
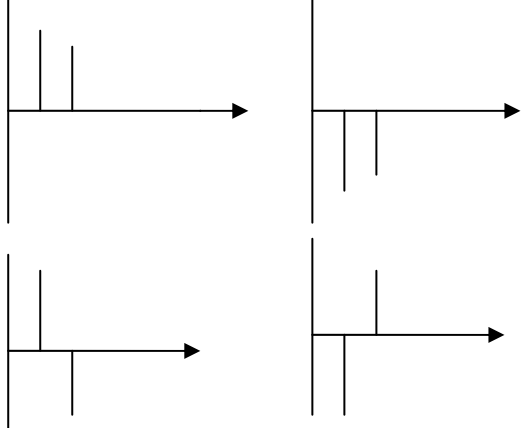
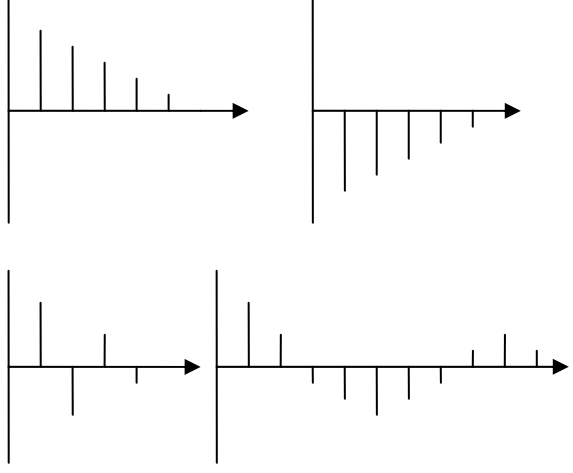
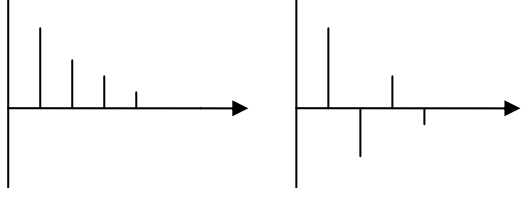
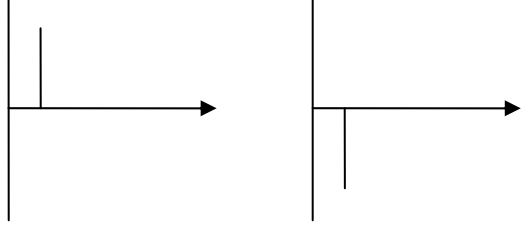
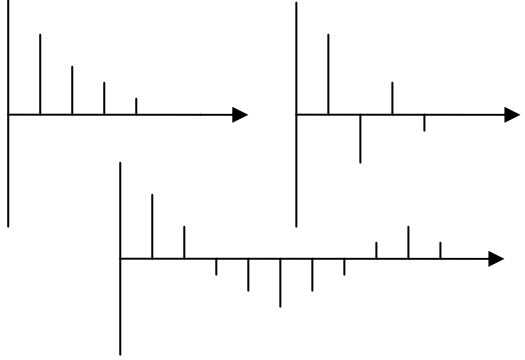
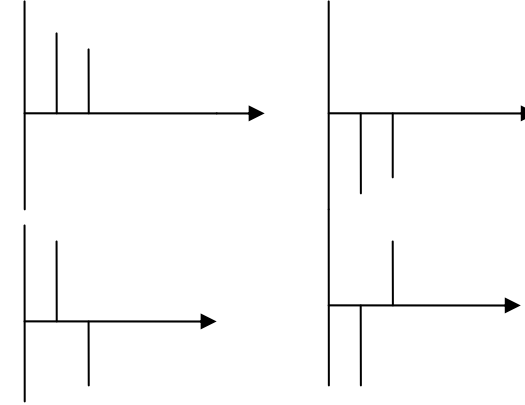
v> Các w_t có tương quan với các quan sát w_{t-k} trước đó (do định nghĩa của các mô hình ARIMA)

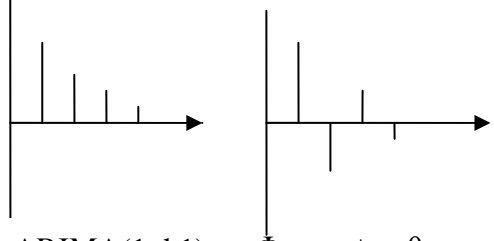
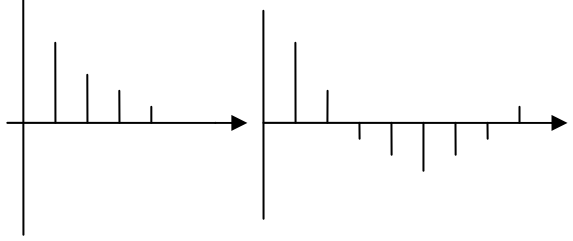
vi> Tham số tự tương quan lý thuyết ở bước k được định nghĩa là

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{E(w_t w_{t-k})}{E(w_t^2)} \quad k = 0, 1, 2, \dots$$

Dưới đây là bảng đồ thị của các hàm tự tương quan và tự tương quan riêng phần lý thuyết đối với các mô hình ARIMA của chuỗi thời gian $x_t: w_t = (1 - B)^d x_t$ không có thành phần $(P,D,Q)_L$ (x_t không có tính thời vụ)

ρ_k	ρ_{kk}
----------	-------------

 <p>ARIMA(0,d,1) $w_t = a_t - \theta_1 a_{t-1}$</p>	
 <p>ARIMA(0,d,2) $w_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$</p>	
 <p>ARIMA(1,d,0) $w_t = \Phi_1 w_{t-1} + a_t$</p>	
 <p>ARIMA(2,d,0) $w_t = \Phi_1 w_{t-1} + \Phi_2 w_{t-2} + a_t$</p>	
<p>ρ_k</p>	<p>ρ_{kk}</p>

 <p>ARIMA(1,d,1) $w_t = \Phi_1 w_{t-1} + a_t - \theta_1 a_{t-1}$</p>	
<p>ARMA(0,d,0) $w_t = a_t$ Tất cả các ρ_k đều bằng 0 (các w_t là độc lập với nhau)</p>	<p>Tất cả các ρ_{kk} đều bằng 0 (các w_t cách nhau k bước là độc lập với nhau)</p>

Một số mô hình ARIMA có tính thời vụ của chuỗi x_t : $w_t = (1 - B)^d (1 - B^{12})^D x_t$

Tên mô hình	Dạng mô hình
ARIMA(0, d, 0) (0, D, 0) ₁₂	$w_t = a_t$
ARIMA(0, d, 0) (0, D, 1) ₁₂	$w_t = (1 - \gamma_{12} B^{12}) a_t$
ARIMA(0, d, 0) (0, D, 2) ₁₂	$w_t = (1 - \gamma_{12} B^{12} - \gamma_{24} B^{24}) a_t$
ARIMA(0, d, 0) (1, D, 0) ₁₂	$(1 - \Lambda_{12} B^{12}) w_t = a_t$
ARIMA(0, d, 0) (2, D, 0) ₁₂	$(1 - \Lambda_{12} B^{12} - \Lambda_{24} B^{24}) w_t = a_t$
ARIMA(0, d, 0) (1, D, 1) ₁₂	$(1 - \Lambda_{12} B^{12}) w_t = (1 - \gamma_{12} B^{12}) a_t$
ARIMA(0, d, 0) (P, D, Q) ₁₂	$(1 - \Lambda_{12} B^{12} - \dots - \Lambda_{12P} B^{12P}) w_t = (1 - \gamma_{12} B^{12} - \dots - \gamma_{12Q} B^{12Q}) a_t$
ARIMA(0, d, 1) (0, D, 1) ₁₂	$w_t = (1 - \theta_1 B) (1 - \gamma_{12} B^{12}) a_t$
ARIMA(1, d, 0) (1, D, 0) ₁₂	$(1 - \Phi_1 B) (1 - \Lambda_{12} B^{12}) w_t = a_t$
ARIMA(1, d, 0) (0, D, 1) ₁₂	$(1 - \Phi_1 B) w_t = (1 - \gamma_{12} B^{12}) a_t$
ARIMA(0, d, 1) (1, D, 0) ₁₂	$(1 - \Lambda_{12} B^{12}) w_t = (1 - \theta_1 B) a_t$
ARIMA(1, d, 1) (1, D, 0) ₁₂	$(1 - \Phi_1 B) (1 - \Lambda_{12} B^{12}) w_t = (1 - \theta_1 B) a_t$
ARIMA(1, d, 1) (0, D, 1) ₁₂	$(1 - \Phi_1 B) w_t = (1 - \theta_1 B) (1 - \gamma_{12} B^{12}) a_t$
ARIMA(p, d, q) (P, D, Q) ₁₂	$(1 - \Phi_1 B - \dots - \Phi_p B^p) (1 - \Lambda_{12} B^{12} - \dots - \Lambda_{12P} B^{12P}) w_t = (1 - \theta_1 B - \dots - \theta_q B^q) (1 - \gamma_{12} B^{12} - \dots - \gamma_{12Q} B^{12Q}) a_t$

Trên thực tế, các mô hình sau đây thường được sử dụng:

ARIMA(0, d, 1) (0, D, 0)₁₂ ARIMA(1, d, 0) (0, D, 0)₁₂ ARIMA(1, d, 0) (0, D, 0)₁₂
ARIMA(0, d, 1) (0, D, 0)₁₂ ARIMA(0, d, 1) (0, D, 1)₁₂ ARIMA(1, d, 0) (1, D, 0)₁₂
ARIMA(1, d, 0) (0, D, 1)₁₂ ARIMA(0, d, 1) (1, D, 0)₁₂

với d và D nhận các giá trị 0 hoặc 1.

4.3 Các hàm dự báo của các mô hình ARMA(p,q)

Sau khi tìm được mô hình ARIMA phù hợp (tối ưu theo nghĩa làm cực tiểu hàm MSE), mô hình này sẽ được sử dụng để phát ra các dự báo cho các quan sát tương lai.

Kí hiệu T là thời điểm hiện tại và $T + \tau$ là thời điểm cần dự báo trong tương lai. Gọi $f_T(\tau)$ là dự báo cho thời điểm $T + \tau$. Giá trị $f_T(\tau)$ sẽ được xây dựng thông qua các dự báo liên tiếp cho các bước $T+1, T+2, \dots, T+\tau-1$. Các dự báo này được tính *bằng cách lấy kì vọng (có điều kiện) của chuỗi thời gian x_t theo mô hình ARIMA viết ở thời điểm xuất phát T cho các bước thời gian $T + i$* . Trong quá trình này

Các quan sát x_{T+i} chưa biết sẽ được thay thế bởi $f_T(i)$

Các phần dư đã xuất hiện $a_{T-i} = x_{T-i} - f_{T-i-1}(1)$

Các phần dư a_{T+i} chưa biết sẽ được thay thế bởi 0.

Ở thời điểm đầu tiên, ta giả thiết rằng $a_{T-i} = 0$ với mọi $T - i \leq 0$ (các phần dư ở các thời điểm 0, -1, -2, ... đều bằng 0).

4.3.1 Một số mô hình ARMA thường gặp:

- **ARMA(0,0)** $x_t = \theta_0 + a_t$

ta có $f_T(1) = E[x_{T+1}] = E[\theta_0 + a_{T+1}] = \mu_X$

$f_T(2) = E[x_{T+2}] = E[\theta_0 + a_{T+2}] = \mu_X$

...

$f_T(\tau) = E[x_{T+\tau}] = E[\theta_0 + a_{T+\tau}] = \mu_X$ (4.1)

- **ARMA(1,0)** $x_t = \theta_0 + \Phi_1 x_{t-1} + a_t$

ta có $f_T(1) = E[x_{T+1}] = E[\theta_0 + \Phi_1 x_T + a_{T+1}]$

$= \theta_0 + \Phi_1 x_T = \mu_X (1 - \Phi_1) + \Phi_1 x_T = \mu_X + \Phi_1 (x_T - \mu_X)$

$f_T(2) = E[x_{T+2}] = E[\theta_0 + \Phi_1 x_{T+1} + a_{T+2}] = E[\theta_0 + \Phi_1 f_T(1) + a_{T+2}]$

$= \theta_0 + \Phi_1 [\mu_X + \Phi_1 (x_T - \mu_X)] = \mu_X + \Phi_1^2 (x_T - \mu_X)$

.....

$f_T(\tau) = \mu_X + \Phi_1^\tau (x_T - \mu_X)$ (4.2)

- **ARMA(2,0)** $x_t = \theta_0 + \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + a_t$

ta có $f_T(1) = E[x_{T+1}] = E[\theta_0 + \Phi_1 x_T + \Phi_2 x_{T-1} + a_{T+1}] = \theta_0 + \Phi_1 x_T + \Phi_2 x_{T-1}$

$= \mu_X (1 - \Phi_1 - \Phi_2) + \Phi_1 x_T + \Phi_2 x_{T-1}$

$= \mu_X + \Phi_1 (x_T - \mu_X) + \Phi_2 (x_{T-1} - \mu_X)$ (4.3)

$f_T(2) = E[x_{T+2}] = E[\theta_0 + \Phi_1 x_{T+1} + \Phi_2 x_T + a_{T+2}] = E[\theta_0 + \Phi_1 f_T(1) + \Phi_2 x_T + a_{T+2}]$

$= \mu_X (1 - \Phi_1 - \Phi_2) + \Phi_1 f_T(1) + \Phi_2 x_T$

$= \mu_X + \Phi_1 [f_T(1) - \mu_X] + \Phi_2 (x_T - \mu_X)$ (4.3')

.....

$f_T(\tau) = \mu_X + \Phi_1 [f_T(\tau-1) - \mu_X] + \Phi_2 [(f_T(\tau-2) - \mu_X)]$ (4.3'')

- **ARMA(0,1)** $x_t = \theta_0 + a_t - \theta_1 a_{t-1}$

Tính các $a_t = x_t - \theta_0 + \theta_1 a_{t-1}$ với $a_0 = 0$ (và từ đó suy ra các a_{t-1})

Tính $f_T(1) \approx x_{T+1} = \theta_0 + a_{T+1} - \theta_1 a_T = \theta_0 - \theta_1 a_T$ (4.4)

$$f_{100}(\tau) \approx x_{100+k} = \theta_0 = \mu_X \forall \tau > 1 \text{ (ở đây } \theta_0 = \mu_X) \quad (4.4')$$

- **ARMA(0,2)** $x_t = \theta_0 + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ ở đây $\theta_0 = \mu_X$

Tính các a_t với $a_0 = 0, a_t = x_t - \theta_0 + \theta_1 a_{t-1} + \theta_2 a_{t-2}$

$$\begin{aligned} \text{ta có } f_T(1) &= E[x_{T+1}] = E[\theta_0 + a_{T+1} - \theta_1 a_T - \theta_2 a_{T-1}] \\ &= \mu_X - \theta_1 a_T - \theta_2 a_{T-1} \end{aligned} \quad (4.5)$$

$$\begin{aligned} f_T(2) &= E[x_{T+2}] = E[\theta_0 + a_{T+2} - \theta_1 a_{T+1} - \theta_2 a_T] \\ &= \mu_X - \theta_2 a_T \end{aligned} \quad (4.5')$$

$$f_T(\tau) = \theta_0 = \mu_X \text{ với mọi } \tau > 2 \quad (4.5'')$$

- **ARMA(1,1)** $x_t = \theta_0 + \Phi_1 x_{t-1} + a_t - \theta_1 a_{t-1}$

Tính các a_t với $a_0 = 0, a_t = x_t - \Phi_1 x_{t-1} - \theta_0 + \theta_1 a_{t-1} + \theta_1 a_{t-1}$ ở đây $\theta_0 = \mu_X(1 - \Phi_1)$

$$\begin{aligned} \text{ta có } f_T(1) &= E[x_{T+1}] = E[\theta_0 + \Phi_1 x_T + a_{T+1} - \theta_1 a_T] \\ &= \theta_0 + \Phi_1 x_T - \theta_1 a_T = \mu_X(1 - \Phi_1) + \Phi_1 x_T - \theta_1 a_T \\ &= \mu_X + \Phi_1(x_T - \mu_X) - \theta_1 a_T \end{aligned} \quad (4.6)$$

$$\begin{aligned} f_T(2) &= E[x_{T+2}] = E[\theta_0 + \Phi_1 x_{T+1} + a_{T+2} - \theta_1 a_{T+1}] \\ &= \theta_0 + \Phi_1 x_{T+1} = \mu_X(1 - \Phi_1) + \Phi_1 f_T(1) \\ &= \mu_X + \Phi_1(f_T(1) - \mu_X) \end{aligned}$$

.....

$$f_T(\tau) = \mu_X + \Phi_1(f_T(\tau-1) - \mu_X) \text{ với mọi } \tau \geq 2 \quad (4.6')$$

4.3.2 Giới hạn cho phép của các dự báo

Một trong các cách tính các giới hạn cho phép của dự báo là viết lại mô hình đã cho dưới dạng tổng theo trọng số

$$x_{T+\tau} = a_{T+\tau} + \Psi_1 a_{T+\tau-1} + \Psi_2 a_{T+\tau-2} + \dots + \Psi_{\tau-1} a_{T+1} + \Psi_{\tau} a_T + \Psi_{\tau+1} a_{T-1} + \dots \quad (4.7)$$

$$\rightarrow f_T(\tau) = E[x_{T+\tau}] = \Psi_{\tau} [x_T - f_{T-1}(1)] + \Psi_{\tau+1} [x_{T-1} - f_{T-2}(1)] + \dots \quad (4.8)$$

$$\rightarrow e_T(\tau) = x_{T+\tau} - f_T(\tau) = a_{T+\tau} + \Psi_1 a_{T+\tau-1} + \Psi_2 a_{T+\tau-2} + \dots + \Psi_{\tau-1} a_{T+1}$$

4.4 Các ví dụ minh họa

Ví dụ 1: Cho chuỗi các quan sát về độ dẻo của 100 sản phẩm có mô hình ARMA(2,0) phù hợp

$$x_t = 20,64 + 0,67 x_{t-1} - 0,38 x_{t-2} + a_t.$$

Hãy dự báo độ dẻo cho hai sản phẩm tiếp theo.

$$\text{Sử dụng công thức (4.3) và (4.3')} \text{ với } T = 100, \mu_X = \frac{20,64}{1 - 0,67 + 0,38} = 28,71$$

ta có

$$\begin{aligned} f_{100}(1) &= \mu_X + \Phi_1(x_{100} - \mu_X) + \Phi_2(x_{99} - \mu_X) \\ &= 28,71 + 0,67(32,44 - 28,71) - 0,38(26,74 - 28,71) = 31,95 \end{aligned}$$

$$f_{100}(2) = \mu_X + \Phi_1[f_{100}(1) - \mu_X] + \Phi_2(x_{100} - \mu_X)$$

$$= 28,71 + 0,67 (31,95 - 28,71) - 0,38 (32,44 - 28,71) = 29,44$$

Ví dụ 2: Cho các quan sát x_t nhu cầu về nhựa hàng tuần (tính bằng tấn) của một công ty sản xuất dây cáp điện trong 100 tuần liên tiếp. Mô hình ARMA(0,1) phù hợp cho chuỗi

$$w_t = (1 - B) x_t$$

là

$$w_t = 4,47 + a_t + 0,52 a_{t-1}$$

$$(\text{hay } x_t = x_{t-1} + 4,47 + a_t + 0,52 a_{t-1})$$

Hãy dự báo nhu cầu dùng nhựa cho hai tuần tới.

Tính cột các $a_t = w_t + \theta_1 a_{t-1}$ với $a_0 = 0$ và từ đó suy ra cột các a_{t-1}

Sử dụng các công thức (4.4) và (4.4') ta có

$$f_{100}(1) = E[w_{100+1}] = E[\theta_0 + a_{101} - \theta_1 a_{100}]$$

$$= \theta_0 - \theta_1 a_{100}$$

$$= 4,47 + 0,52 (-226,72) \rightarrow x_{101} \approx f_{100}(1) + x_{100} = 5502,54$$

$$f_{100}(2) = E[x_{100+2}] = E[x_{100+1} + \theta_0 + a_{100+2} - \theta_1 a_{100+1}]$$

$$= f_{100}(1) + 4,47 \rightarrow x_{102} \approx \theta_0 + f_{100}(1) = 5502,54 + 4,47 = 5507,01$$

Ví dụ 3: Cho 48 quan sát x_t về số kiện một loại nước giải khát đóng chai bán ra hàng tháng của một hãng trong 4 năm liền. Hãy dự báo sự tiêu thụ loại nước này trong 3 tháng tới.

Mô hình ARMA(1,1) là phù hợp với chuỗi thời gian $w_t = (1-B)(1-B^{12})x_t$. Nói cách khác, mô hình ARIMA(1,1,1) $(0,1,0)_{12}$ là phù hợp với chuỗi thời gian x_t .

$$(1 - \Phi_1 B) w_t = (1 - \theta_1 B) a_t \leftrightarrow w_t = \Phi_1 w_{t-1} + a_t - \theta_1 a_{t-1}$$

$$(\text{hay } x_t = (1 + \Phi_1) x_{t-1} - \Phi_1 x_{t-2} + x_{t-12} - (1 + \Phi_1) x_{t-13} + \Phi_1 x_{t-14} + a_t - \theta_1 a_{t-1})$$

với

$$\Phi_1 = 0,94; \theta_0 = 0 \text{ và } \theta_1 = 0,14$$

Tính cột các $a_t = w_t + \theta_1 a_{t-1}$ với $a_0 = 0$ (và từ đó suy ra cột các a_{t-1})

Ở thời điểm dự báo $T+\tau$ ta có

$$w_{T+\tau} = \Phi_1 w_{T+\tau-1} + a_{T+\tau} - \theta_1 a_{T+\tau-1}$$

Với $\tau = 1$ ta có

$$f_T(1) = E[w_{T+1}] = \Phi_1 w_T - \theta_1 a_T$$

Vậy

$$f_{48}(1) = (0,94) w_{48} - 0,14 a_{48}$$

$$= (0,94) (683) - (0,14) (-78,02) = 87,78$$

$$\rightarrow x_{49} \approx f_{48}(1) + x_{48} + x_{37} - x_{36} = 554,78$$

Với $\tau = 2$ ta có

$$f_{48}(2) = E[w_{48+2}] = \Phi_1 w_{48+1} = \Phi_1 f_{48}(1) = (0,94) (87,78) = 82,51$$

$$\rightarrow x_{50} \approx f_{48}(2) + x_{49} + x_{38} - x_{37} = 549,29$$

...

Ví dụ 4: Cho các quan sát x_t về quãng đường bay hàng tháng (đơn vị tính là triệu km) của một hãng hàng không trong 7 năm liền. Mô hình ARIMA phù hợp với chuỗi thời gian

$$y_t = \ln x_t$$

là

$$(1 - B)(1 - B^{12}) y_t = (1 - \theta_1 B)(1 - \gamma_{12} B^{12}) a_t \text{ với } \theta_1 = -0,67 \text{ và } \gamma_{12} = -0,56$$

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + a_t - 0,67 a_{t-1} - 0,56 a_{t-12} + 0,38 a_{t-13}$$

Hãy dự báo quãng đường bay của ba tháng tới.

Ở thời điểm dự báo $T + \tau = 84 + 1$ ta có

$$y_{T+1} = y_T + y_{T-11} - y_{T-12} + a_{T+1} - 0,67 a_T - 0,56 a_{T-11} + 0,38 a_{T-12}$$

$$\rightarrow f_T(1) = E[y_T + y_{T-11} - y_{T-12} + a_{T+1} - 0,67 a_T - 0,56 a_{T-11} + 0,38 a_{T-12}]$$

$$= y_T + y_{T-11} - y_{T-12} - 0,67 a_T - 0,56 a_{T-11} + 0,38 a_{T-12}$$

$$\text{Vậy } f_{84}(1) = y_{84} + y_{73} - y_{72} - 0,67 a_{84} - 0,56 a_{73} + 0,38 a_{72}$$

$$= 9,46 + 9,29 - 9,41 - (0,67) (a_{84}) - (0,56) (a_{73}) + (0,38) (a_{72})$$

$$= 9,3485$$

Ở thời điểm dự báo $T + \tau = 84 + 2$ ta có

$$y_{T+2} = y_{T+1} + y_{T-10} - y_{T-11} + a_{T+2} - 0,67 a_{T+1} - 0,56 a_{T-10} + 0,38 a_{T-11}$$

$$\rightarrow f_T(2) = E[y_{T+2}] = f_T(1) + y_{T-10} - y_{T-11} - 0,56 a_{T-10} + 0,38 a_{T-11}$$

$$= 9,3485 + 9,2530 - 9,2910 - (0,56) (0,0092) + (0,38) (0,0026)$$

$$= 9,3063$$

Ở thời điểm dự báo $T + \tau = 84 + 3$ ta có

$$y_{T+3} = y_{T+2} + y_{T-9} - y_{T-10} + a_{T+3} - 0,67 a_{T+2} - 0,56 a_{T-9} + 0,38 a_{T-10}$$

$$\rightarrow f_T(3) = E[y_{T+3}] = f_T(2) + y_{T-9} - y_{T-10} - 0,56 a_{T-9} + 0,38 a_{T-10}$$

$$= 9,3063 + 9,4716 - 9,2438 - (0,56) (0,0455) + (0,38) (0,0092)$$

$$= 9,5484$$

Trở lại biến cũ ta nhận được các dự báo cho ba tháng tiếp theo là

$$F_{84}(1) = 11482$$

$$F_{84}(2) = 11008$$

$$F_{84}(3) = 13128$$

4.5 BÀI TẬP CHƯƠNG 4

1. Cho các mô hình chuỗi thời gian x_t phù hợp

a) $x_t = 25 + 0,34 x_{t-1} + a_t$

Biết $x_{100} = 28$, tính các dự báo cho các thời điểm 101, 102, 103

b) $x_t = 15 + 0,36 x_{t-1} - 0,32 x_{t-2} + a_t$

Biết $x_{50} = 32$ và $x_{51} = 30$, tính các dự báo cho các thời điểm 51, 52

c) $x_t = 20 + a_t + 0,45 a_{t-1} - 0,35 a_{t-2}$

Biết $x_{100} = 620$ và $x_{99} = 624$, tính các dự báo cho các thời điểm 101, 102

2. a) Sau khi tìm được mô hình phù hợp với dữ liệu cho ở câu 7 chương 3, hãy phát dự báo cho 3 tháng sắp tới.
b) Sau khi tìm được mô hình phù hợp với dữ liệu cho ở câu 8 chương 3, hãy phát dự báo cho 3 tháng sắp tới.
c) Sau khi tìm được mô hình phù hợp với dữ liệu cho ở câu 9 chương 3, hãy phát dự báo cho 3 tháng sắp tới.

3. Doanh thu hàng tháng (đơn vị tính là triệu đồng) của một cửa hàng trong 8 năm liên được cho trong bảng dưới đây

Tháng	1994	1995	1996	1997	1998	1999	2000	2001
1	26.0	29.0	27.0	27.5	23.9	27.6	26.9	28.4
2	24.5	24.7	26.3	27.2	24.7	23.4	26.6	25.5
3	27.9	31.3	29.8	30.2	27.5	25.0	27.6	26.6
4	29.1	32.4	32.6	28.6	26.7	26.0	27.1	26.2
5	34.7	33.9	35.1	34.1	28.7	31.0	29.8	29.3
6	33.1	35.0	34.4	30.9	30.3	29.3	29.1	28.8
7	36.0	36.4	35.7	34.7	31.3	31.7	32.6	31.2
8	37.5	36.5	33.6	33.7	32.1	32.0	31.6	31.9
9	34.8	34.4	31.9	33.6	31.2	30.0	31.1	28.5
10	35.5	33.9	35.1	31.0	31.4	31.8	33.2	32.2
11	33.4	33.9	33.4	28.9	30.8	33.6	33.6	31.7
12	32.9	36.4	37.6	29.7	30.6	31.6	34.0	31.8

Tìm mô hình ARIMA phù hợp với chuỗi thời gian này. Phát dự báo doanh thu cho 3 tháng sắp tới.

4. Số lượng khách (đơn vị ngàn người) của một hãng hàng không trong 3 năm là

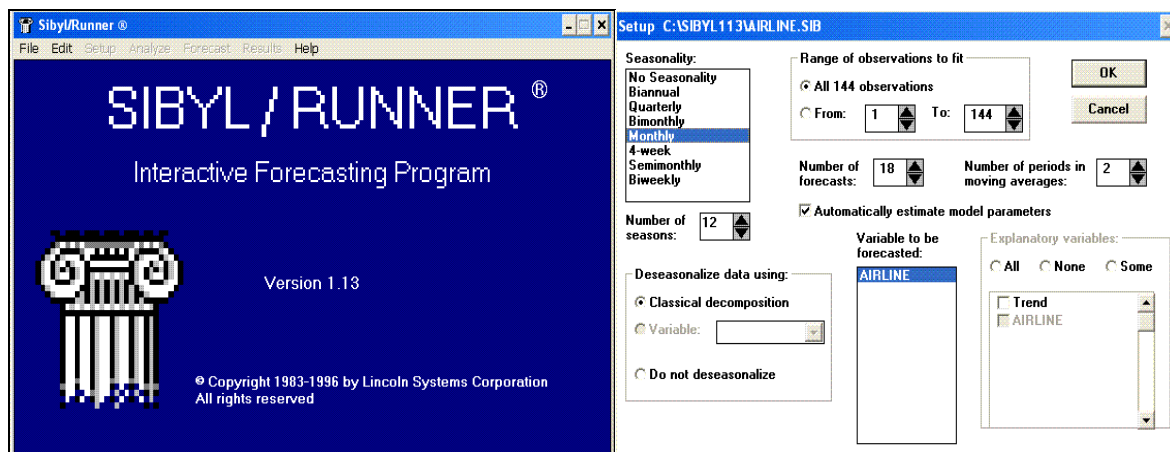
112	118	132	129	121	135	148	136	119	109	104	111
115	126	141	135	125	149	170	170	158	133	114	141
145	150	178	163	172	178	199	199	189	162	146	161

Tìm mô hình ARIMA phù hợp với chuỗi thời gian này. Phát dự báo doanh thu cho 3 tháng sắp tới.

5 PHỤ LỤC: GIỚI THIỆU PHẦN MỀM DỰ BÁO SIBYL

5.1 Môi trường làm việc của Sibyl

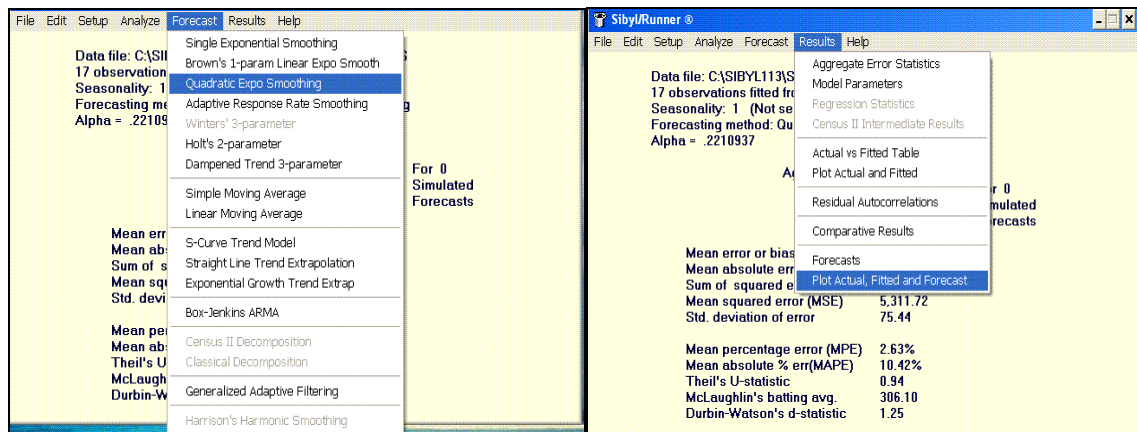
SIBYL/RUNNER là chương trình dự báo của Lincoln System Coporation gọn nhẹ, tương tác thân thiện với người sử dụng. Với Sibyl, ta có thể nhập dữ liệu mới từ bàn phím hay đọc từ file và ghi lại dữ liệu ra file với kiểu định sẵn là SIR



Có thể vào menu File để mở file dữ liệu có sẵn hay tạo file mới, hoặc vào menu Edit để soạn thảo dữ liệu. Sau đó chọn menu Analysis để xem các phân tích dạng dữ liệu (vẽ đường quan hệ $x \sim t$, phân tích yếu tố mùa, tính tự hồi quy, ... rồi vào Setup để thiết đặt các tính chất của dữ liệu và các tham số cần cho dự báo

Menu Forecast sẽ hiển thị những mô hình dự báo có thể sử dụng cho file dữ liệu bạn đang mở để lựa chọn.

Sau khi một mô hình dự báo đã được lựa chọn, Sibyl tự động tính toán và bạn mở menu Results để lựa chọn những kết quả muốn xem.



Nếu các tiêu chuẩn dự báo chưa đạt mức đề ra, bạn có thể chỉnh sửa các thiết lập dữ liệu trong menu Setup, chọn mô hình dự báo khác và xem lại kết quả.

5.2 Một số phương pháp dự báo trong Sibyl

5.2.1 Các phương pháp trung bình trượt

Trung bình trượt là lấy trung bình với số điểm được xét cố định. Trong chuỗi thời gian điều đó nghĩa là thông tin cũ nhất bị loại ra khi thông tin mới được cập nhật. Do đó giá trị trung bình nhận được luôn luôn là trung bình của một số không đổi các quan sát gần hiện tại nhất.

- Trọng số trung bình trượt là như nhau ($1/N$) cho n quan sát gần nhất. Vậy người sử dụng bắt buộc phải cung cấp đủ N quan sát gần nhất để dự báo bước tiếp theo. **MAVE** (Simple Moving Average). Phương pháp này thích hợp hơn cho chuỗi dữ liệu ổn định (không có tính xu thế)
- **MAVE2** (Linear Brown's Moving Average). Đây là phương pháp trung bình trượt kép. Nó thích hợp cho chuỗi thời gian có tính xu thế.

5.2.2 Các phương pháp hồi quy tìm đường cong phù hợp với chuỗi dữ liệu (Trend-Cycle Regression Curve-Fitting Methods)

- **SCURVE** (Life Cycle Analysis). Phương pháp này giả thiết chuỗi thời gian có quan hệ dạng:

$$x = e^{a + b/t}$$

với

x là chuỗi thời gian,

t là các thời điểm (1, 2, 3, ..., N),

a và b là các hệ số tính được bằng phương pháp hồi quy

Việc tìm đường cong phù hợp là thích hợp cho việc phân tích vòng đời của bất kỳ dạng chuỗi thời gian nào. Phương pháp này thường giả thiết rằng chuỗi dữ liệu xuất phát từ giá trị nhỏ nhưng lại thay đổi theo cấp số nhân, có một mức bão hòa.

- **SREG** (Straight-Line Trend Extrapolation). Phương pháp này giả thiết chuỗi thời gian có quan hệ dạng:

$$x = a + bt$$

với

x là chuỗi thời gian,

t là các thời điểm (1, 2, 3, ..., N),

a và b là các hệ số tính được bằng cách lấy hồi quy

- **EXGROW** (Exponential Growth Trend Extrapolation). Phương pháp này giả thiết có một hằng số tỉ lệ tăng trong chuỗi thời gian và vì thế chuỗi thời gian có quan hệ dạng

$$x = e^{a + bt}$$

với

x là chuỗi thời gian,

t là các thời điểm (1, 2, 3, ..., N),

a và b là các hệ số tính được bằng phương pháp hồi quy

Phương pháp này được dùng khi cần phải dự báo dài hạn một chuỗi thời gian có tỉ lệ tăng không đổi.

5.2.3 Các phương pháp làm trơn dạng mũ

Các phương pháp này làm trơn (lấy trung bình) các giá trị quá khứ theo kiểu mũ, tức là chúng cho các quan sát ở xa hiện tại các trọng số giảm dần dạng mũ.

- **EXPO** (Single Exponential Smoothing). Trong phương pháp này, giá trị dự báo được là trung bình trọng số của một số giá trị đứng trước nó. Các trọng số này giảm dần theo dạng mũ khi các giá trị càng xa hiện tại.

Công thức dự báo:

$$f(t+1) = a x(t) + (1-a) f(t)$$

với

$f(t+1)$ là dự báo,

a là hằng trơn,

t là thời điểm dự báo

$x(t)$ là giá trị của chuỗi tại thời điểm t .

Phương pháp này chỉ tốt cho chuỗi thời gian ổn định, không có xu thế tăng / giảm.

- **EXPO2** (Brown's 1-Parameter Linear Exponential Smoothing). Giống như EXPO, phương pháp này làm trơn (lấy trung bình) các giá trị quá khứ theo kiểu mũ, tức là chúng cho các quan sát ở xa hiện tại các trọng số giảm dần dạng mũ. Tuy nhiên nó thích hợp cho cả trường hợp dữ liệu có tính xu thế tuyến tính (bậc nhất).
- **EXPOQ** (Quadratic Exponential Smoothing). Đây cũng là phương pháp làm trơn dạng mũ nhưng ưu điểm của nó là cho phép dự báo đối với dữ liệu có xu thế là đường bậc hai.
- **EXPOTL** (Adaptive Response Rate Smoothing Trigg and Leach's exponential smoothing) là một phương pháp làm trơn dạng mũ với tỉ lệ thích hợp. Điểm khác biệt là nó không đòi hỏi người sử dụng phải chỉ ra một giá trị α (hoặc máy tính phải dò tìm một giá trị α tối ưu) phụ thuộc vào kiểu dữ liệu và biên độ biến đổi của nó. Vậy đây là một phương pháp tự thích nghi, không đòi hỏi thêm thông tin từ phía người sử dụng.
- **EXPOW** (Linear and Seasonal Winters' 3-Parameter Exponential Smoothing) Phương pháp này tương tự các phương pháp làm trơn dạng mũ kể trên, điểm khác biệt là nó xử lý được cả chuỗi dữ liệu vừa có nhân tố mùa vừa có tính xu thế.
- **EXPOH** (Linear Holt's Exponential Smoothing). Phương pháp Holt sử dụng nguyên tắc làm trơn dạng mũ đơn cho các quan sát quá khứ cho chuỗi dữ liệu có tính xu thế tuyến tính bằng cách sử dụng 2 tham số riêng biệt.
- **EXPOD** (Dampened-Trend Exponential Smoothing). Tương tự các mô hình của Winters và Holt là xem xét cả tính xu thế, nhưng nó thêm một tham số (là tổng của 3 tham số) để giảm thiểu tác động của xu thế khi dự báo dài hạn. Phương pháp này dùng cho trường hợp xu thế không có dạng đường thẳng mà là dạng dao động tắt dần theo thời gian.

5.2.4 Các phương pháp phân ly

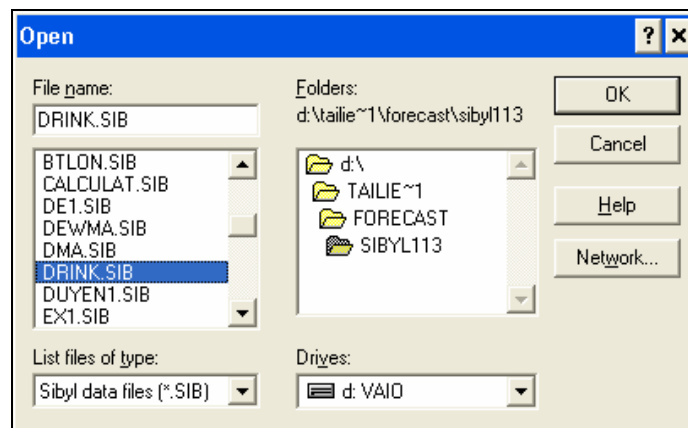
- **CENSUS** (Census Decomposition). Phương pháp này tách chuỗi thời gian thành các thành phần như thời vụ, xu thế và ngẫu nhiên, nó cung cấp cả ước lượng cho các chỉ số về thời vụ.
- **CENSUS II**. Phương pháp này cung cấp các dữ liệu thống kê giúp điều chỉnh mô hình sau khi xem xét kết quả dự báo. Dự báo được coi như ngoại suy tuyến tính của chuỗi thời gian trong quá khứ
- **DCOMP** (Classical Decomposition). Phương pháp này phân ly chuỗi thời gian ra các thành phần như thời vụ, xu thế Dự báo được coi như ngoại suy tuyến tính của chuỗi thời gian trong quá khứ.

5.2.5 Phương pháp Box-Jenkins

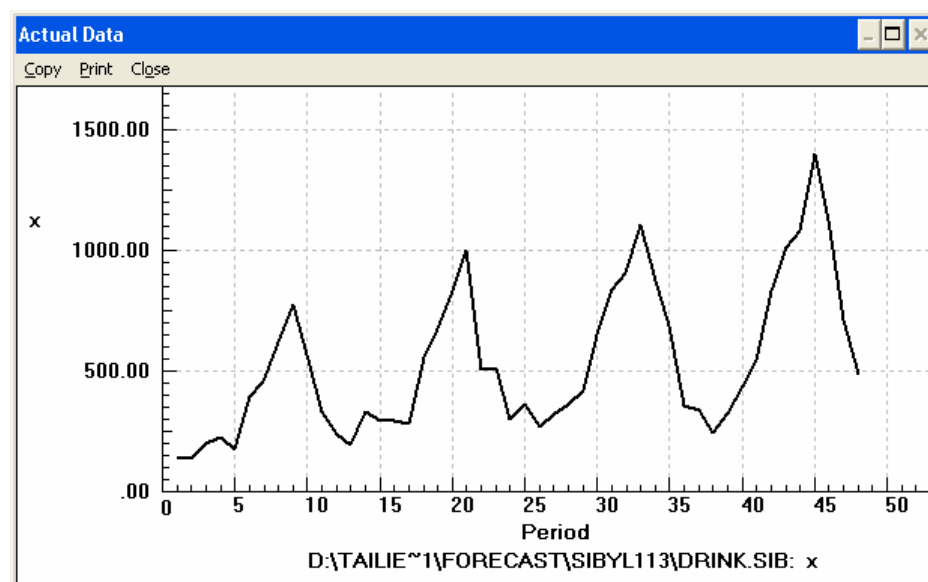
Box-Jenkins ARMA (Auto-Regressive Moving Average) là dạng chung nhất của các phương pháp dự báo chuỗi thời gian. Các giá trị tương lai của chuỗi thời gian được xác định từ tổ hợp của các giá trị quá khứ và sai số quá khứ. Dựa trên giả thiết rằng các giá trị liên tiếp của chuỗi thời gian có liên quan với nhau, Box-Jenkins cố gắng khám phá điều đó và sử dụng để dự báo. Phương pháp Box-Jenkins có thể dùng cho cả chuỗi thời gian có tính dừng cũng như không dừng.

➤ Cách sử dụng bộ công cụ ARIMA trong SIBYL

Bước 1: Sử dụng menu File và Edit nhập dữ liệu và biến đổi nếu cần thiết (dựa trên minh họa đồ thị chuỗi quan sát và hàm tự tương quan mẫu)



Bước 2: Mở menu Analyze để xem đồ thị chuỗi thời gian cũng như đồ thị hàm tự tương quan



Bước 3: Mở menu Setup để thiết đặt các tính chất của dữ liệu như chọn chuỗi thời gian, số quan sát được sử dụng, xác định tính thời vụ, có phân li tính thời vụ hay không, số bước thời gian cần dự báo.

Setup D:\TAILIE~1\FORCAST\SIBYL113\DRINK.SIB

Seasonality:
 No Seasonality
 Biannual
 Quarterly
 Bimonthly
Monthly
 4-week

Range of observations to fit
☒ All 48 observations
☐ From: 1 To: 48

Number of forecasts: 12 **Number of periods in moving averages:** 2

☒ Automatically estimate model parameters

Number of seasons: 12

Deseasonalize data using:
☐ Classical decomposition
☐ Variable:
☒ Do not deseasonalize

Variable to be forecasted:
 x

Explanatory variables:
☐ All ☐ None ☐ Some
☐ Trend
☒ x

OK Cancel

Bước 4: Mở menu Forecast chọn phương pháp Box-Jenkins. Sử dụng chức năng AutoCorrelation ... để hiển thị các thông tin về tính dừng, tính thời vụ cho dạng mô hình ARIMA phù hợp, ... Chọn một mô hình thử nghiệm.

Sibyl/Runner ®

File Edit Setup Analyze Forecast Results Help

Data file: D:\TAILIE~1\FORCAST\SIBYL113\DRINK.SIB
48 observations
Seasonality: Monthly
Forecasting: Box-Jenkins ARMA

Comp Per-t
Orig
Trend
Cy
Seas
Rand
*** Use**
Othe

Single Exponential Smoothing
 Brown's 1-param Linear Expo Smooth
 Quadratic Expo Smoothing
 Adaptive Response Rate Smoothing
 Winters' 3-parameter
 Holt's 2-parameter
 Dampened Trend 3-parameter
 Simple Moving Average
 Linear Moving Average
 S-Curve Trend Model
 Straight Line Trend Extrapolation
 Exponential Growth Trend Extrapolation
Box-Jenkins ARMA
 Census II Decomposition
 Classical Decomposition
 Generalized Adaptive Filtering
 Harrison's Harmonic Smoothing
 Mult Regress - Seasonal Dummies
 Multivariate Multiple Regression

Variable: x
Approximate % Explained
 100.00
 14.07
 40.33
 45.60
 this percentage is small.

Box-Jenkins Methods

Model:

- AR (1)
- AR (2)
- MA (1)
- MA (2)
- ARMA(1,1)
- AR(1) S in AR
- AR(1) S in MA
- AR(2) S in MA
- MA(1) S in AR
- MA(1) S in MA
- MA(2) S in AR
- MA(2) S in MA
- ARMA(1,1) S in AR
- ARMA(1,1) S in MA

Differences:

Short term differences

☐ 0 - No trend

☒ 1 - Linear trend

☐ 2 - Quadratic trend

Long term differences

☐ 0 - 0th level

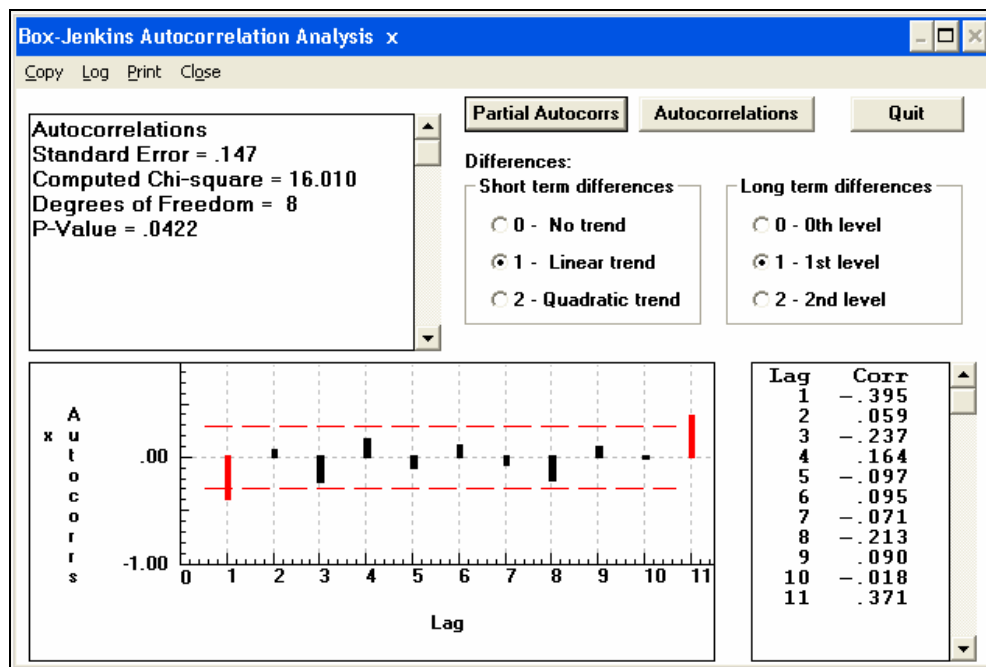
☒ 1 - 1st level

☐ 2 - 2nd level

OK

Cancel

Autocorr...



Box-Jenkins Methods

Model:

- AR (1)
- AR (2)
- MA (1)
- MA (2)
- ARMA(1,1)
- AR(1) S in AR
- AR(1) S in MA
- AR(2) S in MA
- MA(1) S in AR
- MA(1) S in MA
- MA(2) S in AR
- MA(2) S in MA
- ARMA(1,1) S in AR
- ARMA(1,1) S in MA

Differences:

Short term differences

☐ 0 - No trend

☒ 1 - Linear trend

☐ 2 - Quadratic trend

Long term differences

☐ 0 - 0th level

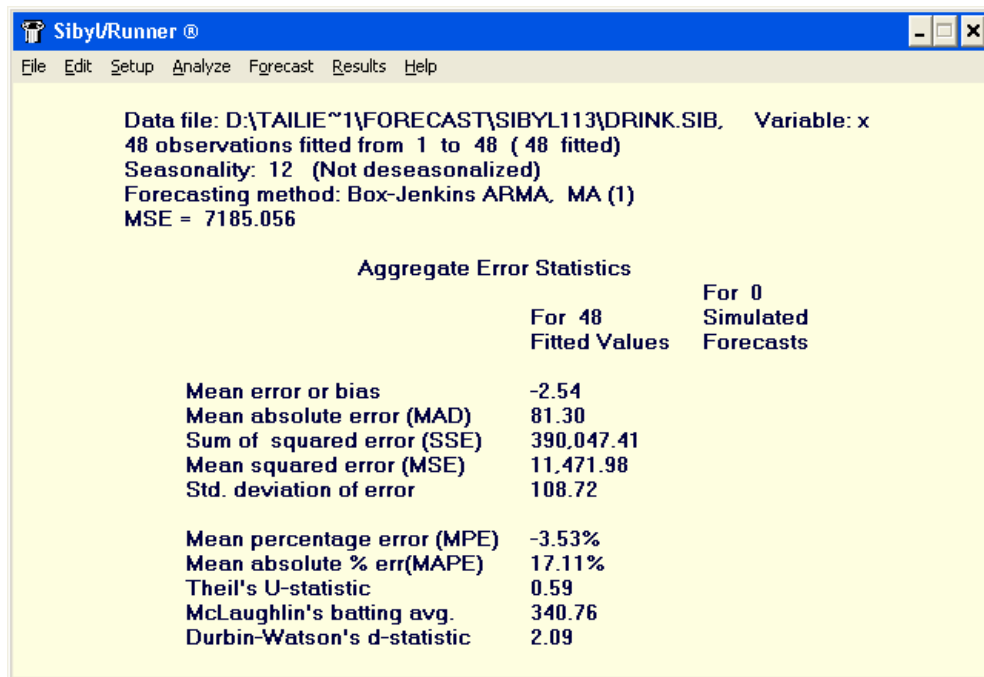
☒ 1 - 1st level

☐ 2 - 2nd level

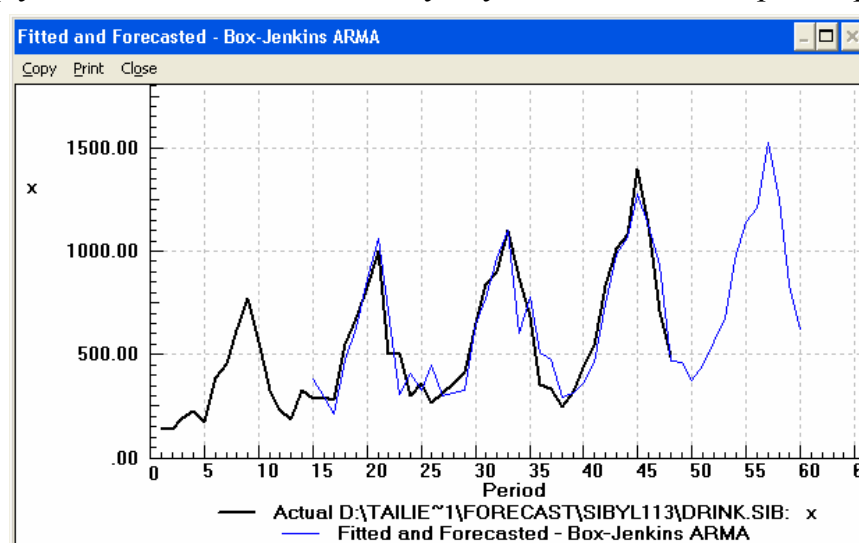
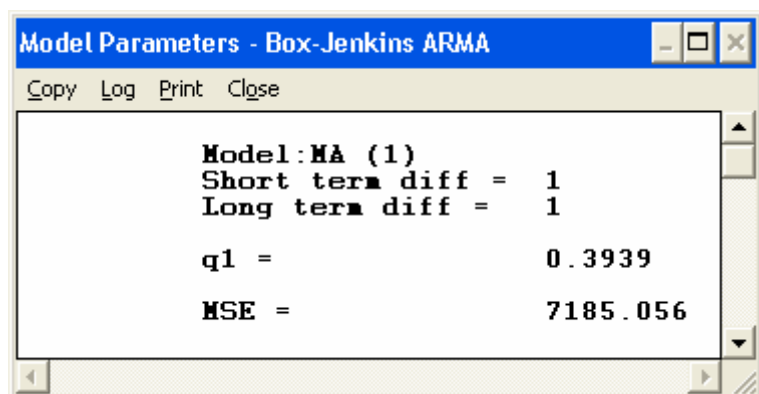
OK

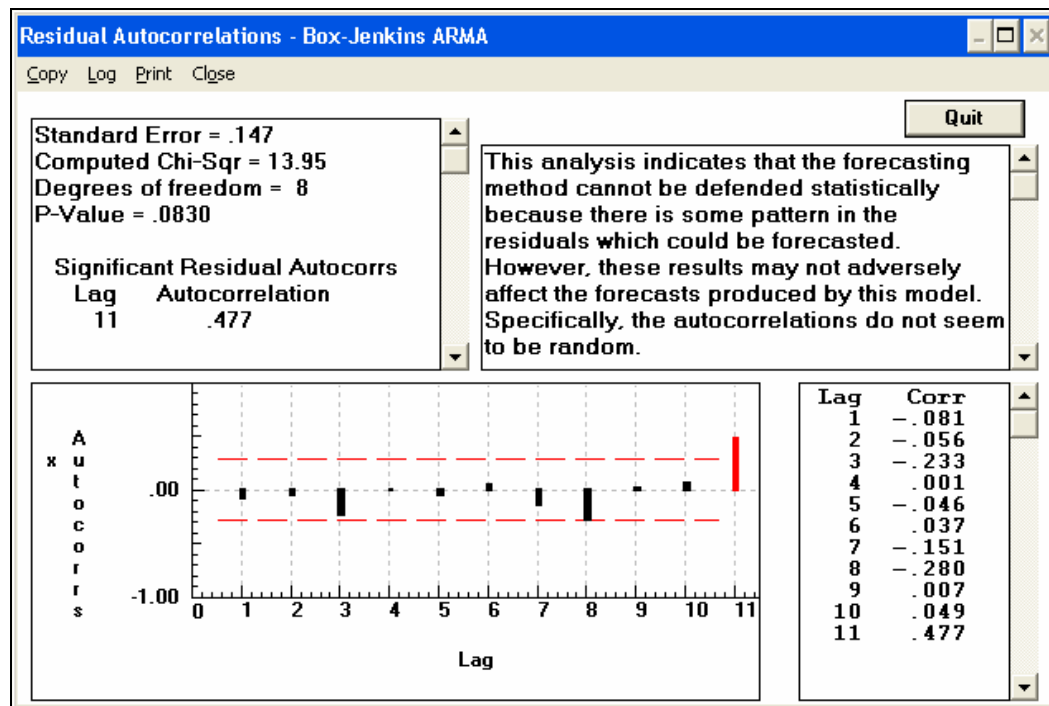
Cancel

Autocorr...



Bước 5: Mở menu Results xem các thông số, các giá trị của mô hình khớp với các quan sát, các sai số, các giá trị của hàm tự tương quan mẫu và tự tương quan riêng phần mẫu và đồ thị minh họa để quyết định lựa chọn mô hình này hay xét mô hình khác phù hợp hơn.





Forecast Values - Box-Jenkins ARMA

Copy Log Print Close

Period	Forecast
49	459.56
50	371.56
51	447.56
52	564.56
53	671.56
54	957.56
55	1138.56
56	1208.56
57	1527.56
58	1250.56
59	840.56
60	614.56

Bước 6: So sánh các dự báo do các mô hình tạo ra để lựa chọn kết quả tốt nhất.