



GT.0000025809

GIÁO DỤC VÀ ĐÀO TẠO  
ẠI HỌC THÁI NGUYÊN

NGUYỄN VĂN HUÂN - PHẠM VIỆT BÌNH

**GIÁO TRÌNH**  
**PHÂN TÍCH VÀ DỰ BÁO**  
**TRONG KINH TẾ**

NGUYỄN  
ĐC LIÊU

1



NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**  
**ĐẠI HỌC THÁI NGUYÊN**  
====o0o====  
**NGUYỄN VĂN HUÂN – PHẠM VIỆT BÌNH**

**Giáo trình**  
**PHÂN TÍCH VÀ DỰ BÁO**  
**TRONG KINH TẾ**



**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**  
**HÀ NỘI - 2011**

*Chịu trách nhiệm xuất bản* : TS. Phạm Văn Diễn  
*Biên tập* : Quỳnh Anh, Minh Luận  
*Trình bày bìa* : Trịnh Thị Thuỳ Dương

**NHÀ XUẤT BẢN KHOA HỌC VÀ KỸ THUẬT**  
**70 Trần Hưng Đạo – Hà Nội**

---

In 200 bản khổ 15.5 x 22.5 cm tại Công ty TNHH In Thanh Bình  
Số ĐKKHXB: -2011/CXB/388 - 11/KHKT, ngày 14/02/2011.  
Quyết định XB số: 72/QĐXB-NXBKHKT ngày 20 tháng 5 năm 2011.  
In xong và nộp lưu chiểu tháng 6/2011.

# *MỞ ĐẦU*

---

*Trong hoạt động kinh tế - xã hội của mỗi quốc gia, mỗi vùng lãnh thổ, mỗi đơn vị hoạt động kinh doanh đều vì mục đích lợi ích kinh tế, tức là mục tiêu lợi nhuận. Tuy nhiên, không chỉ bó hẹp trong phạm vi kinh tế mà còn trong tất cả các lĩnh vực khác của đời sống xã hội, chúng ta cần phải biết những gì về quá khứ, hiện tại và cả tương lai để từ đó chúng ta có thể đưa ra những định hướng, quyết định đúng đắn nhất. Để làm được điều này, hiện nay trong nước và trên thế giới người ta đã áp dụng những phương pháp khác nhau, nhưng trong đó có một phương pháp được sử dụng một cách hiệu quả và phổ biến, đó là phương pháp Phân tích dữ liệu và dự báo kinh tế.*

*Lịch sử phát triển của Phân tích dữ liệu và dự báo kinh tế đã được phát triển từ rất lâu. Cho đến nay, đã có nhiều những phương pháp khác nhau, nhưng những phương pháp dự báo phổ biến chỉ được phát triển gần đây: Phương pháp phân tích, phương pháp san mũ, phương pháp ARIMA... Cùng với sự phát triển của nhiều phương pháp dự báo phức tạp và các phần mềm, dự báo ngày càng nhận được nhiều sự quan tâm hơn, nhiều phương pháp dự báo mới tiếp tục được phát triển.*

*Dự báo là một yếu tố quan trọng của hầu hết các quyết định kinh doanh và lập kế hoạch kinh tế; Dự báo như một tập hợp các công cụ giúp người ra quyết định đưa ra các phán đoán tốt nhất về các sự kiện*

*tương lai (dựa vào quá khứ và hiện tại); Nhu cầu nhân sự có kiến thức về dự báo đang gia tăng.*

*Vì những quyết định hôm nay ảnh hưởng đến tương lai của tổ chức, nhưng tương lai là bất định, nên hầu như mọi tổ chức: lớn và nhỏ, tư và công đều sử dụng dự báo. Các bộ phận chức năng như tài chính, marketing, nhân sự, sản xuất, ngoài ra, các tổ chức Chính phủ, phi chính phủ, các CLB xã hội, ... cũng sử dụng dự báo.*

*Chúng ta có thể tiến hành dự báo hàng ngày, hàng tháng, hàng quý, hàng năm hay vài năm, ...*

*Ví dụ: Dự báo trong kinh doanh hàng ngày: Dự báo ngày càng trở nên quan trọng vì các công ty tập trung vào việc gia tăng mức độ hài lòng của khách hàng trong khi vẫn phải giảm chi phí của việc cung cấp hàng hóa và dịch vụ. Hầu như mọi lĩnh vực chức năng của doanh nghiệp đều sử dụng một loại dự báo nào đó, ví dụ:*

*Kế toán: Dự báo chi phí và doanh thu.*

*Phòng nhân sự: Dự báo nhu cầu tuyển dụng và những thay đổi trong công sở.*

*Quản đốc sản xuất: Dự báo nhu cầu nguyên vật liệu và tồn kho.*

*Giám đốc marketing: Dự báo doanh số để thiết lập ngân sách cho quảng cáo.*

*Dự báo doanh số thường là dự báo cơ bản cho các dự báo khác (ví dụ giữa những năm 1980, 94% sử dụng dự báo doanh số).*

*Như vậy, phân tích dữ liệu và dự báo kinh tế có một vai trò quan trọng trong các hoạt động của mỗi đơn vị, mỗi quốc gia, ...*

*Cuốn giáo trình này nhằm cung cấp cho sinh viên, nhà quản lý, nhà hoạch định, nhà kinh tế, ...những phương pháp phân tích dữ liệu và dự báo kinh tế để phần nào hỗ trợ cho công việc.*

*Cuốn giáo trình này sẽ trình bày một cách chi tiết về lý thuyết cũng như các bài thực hành ứng dụng, giúp cho người học dễ tiếp cận những kiến thức mới.*

*Giáo trình gồm có năm chương:*

**Chương 1: TỔNG QUAN VỀ PHÂN TÍCH VÀ DỰ BÁO KINH TẾ**

**Chương 2: CÁC PHƯƠNG PHÁP PHÂN TÍCH VÀ DỰ BÁO**

**Chương 3: PHƯƠNG PHÁP HỒI QUY ĐƠN VÀ HỒI QUY  
BỘI VÀ THỐNG KÊ HỒI QUY**

**Chương 4: PHƯƠNG PHÁP BOX - JENKINS (ARIMA)**

**Chương 5: DÃY SỐ THỜI GIAN**

*Tác giả xin chân thành cảm ơn TS. Trương Văn Tú – Trường Đại học Kinh tế Quốc dân Hà Nội, PGS. TS. Đỗ Năng Toàn – Trưởng phòng Thực tại ảo – Viện Công nghệ thông tin – Viện KH&CN Việt Nam, KS. Vũ Xuân Nam – Trường Đại học Công nghệ thông tin và truyền thông – ĐH Thái Nguyên đã dành nhiều thời gian đọc kỹ bản thảo và cho nhiều ý kiến quý báu trong quá trình tác giả biên soạn cuốn giáo trình này.*

*Tác giả cũng bày tỏ lòng biết ơn đối với Ban Lãnh đạo Trường Đại học Công nghệ thông tin và truyền thông – Đại học Thái Nguyên, Hội đồng Khoa học – Đào tạo, Bộ môn HTTT Kinh tế đã tạo mọi điều kiện để giáo trình này được ra mắt bạn đọc.*

*Tuy đã cố gắng nhưng giáo trình này không thể tránh khỏi những thiếu sót. Rất mong nhận được ý kiến đóng góp của độc giả để trong lần tái bản sau, giáo trình sẽ hoàn chỉnh hơn.*

*Thư góp ý xin gửi về: Nguyễn Văn Huân: Trường Đại học Công nghệ thông tin và truyền thông – Đại học Thái Nguyên - Xã Quyết Thắng, Tp. Thái Nguyên*

*Điện thoại: 0987 118 623; Email: [nvhuân@ictu.edu.vn](mailto:nvhuan@ictu.edu.vn)*

**Các tác giả**

# Mục lục

Mở đầu .....	3
--------------	---

## Chương 1. TỔNG QUAN VỀ PHÂN TÍCH VÀ DỰ BÁO KINH TẾ

1.1. Khái niệm .....	11
1.2. Ý nghĩa và vai trò của phân tích và dự báo trong quá trình ra quyết định kinh doanh .....	12
1.2.1. Ý nghĩa .....	12
1.2.2. Vai trò .....	13
1.3. Các loại dự báo .....	13
1.3.1. Căn cứ vào độ dài thời gian dự báo .....	13
1.3.2. Dựa vào các phương pháp dự báo .....	14
1.3.3. Căn cứ vào nội dung (đối tượng dự báo) .....	15
1.4. Các phương pháp dự báo .....	17
1.4.1. Phương pháp dự báo định tính .....	17
1.4.2. Phương pháp dự báo định lượng .....	20
1.5. Quy trình dự báo .....	36

## Chương 2. CÁC PHƯƠNG PHÁP PHÂN TÍCH VÀ DỰ BÁO

2.1. Dự báo từ các mức độ bình quân .....	42
2.1.1. Dự báo từ số bình quân trượt (di động) .....	42
2.1.2. Mô hình dự báo dựa vào lượng tăng (giảm) tuyệt đối bình quân .....	44
2.1.3. Mô hình dự báo dựa vào tốc độ phát triển bình quân .....	46



2.2. Mô hình dự báo theo phương trình hồi quy (dự báo dựa vào xu thế).....	49
2.2.1. Mô hình hồi quy theo thời gian.....	49
2.2.2. Mô hình hồi quy giữa các tiêu thức.....	54
2.3. Dự báo dựa vào hàm xu thế và biến động thời vụ .....	55
2.3.1. Dự báo dựa vào mô hình cộng .....	55
2.3.2. Dự báo dựa vào mô hình nhân .....	58
2.4. Dự báo theo phương pháp san bằng mũ.....	62
2.4.1. Mô hình đơn giản (phương pháp san bằng mũ đơn giản).....	62
2.4.2. Mô hình xu thế tuyến tính và không có biến động thời vụ (Mô hình san mũ Holt – Winters) .....	68
2.4.3. Mô hình xu thế tuyến tính và biến động thời vụ .....	71
2.5. Sử dụng chương trình SPSS để dự báo theo các mô hình....	75
2.5.1. Dự đoán bằng hàm xu thế .....	75
2.5.2. Dự đoán bằng san bằng mũ .....	76

### **Chương 3. PHƯƠNG PHÁP HỒI QUY ĐƠN VÀ HỒI QUY BỘI VÀ THỐNG KÊ HỒI QUY**

3.1. Phương pháp hồi quy đơn .....	79
3.2. Phương pháp hồi quy bội .....	89
3.3. Phương pháp thống kê hồi quy .....	90

### **Chương 4. PHƯƠNG PHÁP BOX - JENKINS (ARIMA)**

4.1. Tính ổn định của một chuỗi .....	98
4.2. Hàm số tự tương quan đơn và tự tương quan riêng phần.....	98
4.3. Kiểm định nhiễu trắng.....	101
4.3.1. Phân tích hàm tự tương quan .....	101
4.3.2. Tham số thống kê của Box-Pierce và Ljung-Box .....	102
4.4. Mô hình AR(p) (Auto Regression) .....	104

4.5. Mô hình MA(q) (Moving Average).....	107
4.6. Mô hình ARMA(p, q) .....	109
4.7. Mô hình ARMA mở rộng: ARIMA, SARIMA.....	111
4.8. Phương pháp Box - Jenkins .....	112
 <b>Chương 5. DÃY SỐ THỜI GIAN</b>	
5.1. Khái niệm .....	125
5.2. Các chỉ tiêu phân tích.....	127
5.2.1. Mức độ trung bình theo thời gian.....	127
5.2.2. Lượng tăng hoặc giảm tuyệt đối .....	129
5.2.3. Tốc độ phát triển .....	131
5.2.4. Tốc độ tăng hoặc giảm .....	132
5.2.5. Trị tuyệt đối của 1% tăng (hoặc giảm).....	133
5.3. Các phương pháp biểu hiện xu hướng phát triển của hiện tượng .....	133
5.3.1. Phương pháp mở rộng khoảng cách thời gian.....	133
5.3.2. Phương pháp số trung bình trượt.....	134
5.3.3. Phương pháp hồi quy .....	136
5.3.4. Phương pháp biểu hiện biến động thời vụ .....	139
 <b>Tài liệu tham khảo .....</b>	 <b>145</b>

## *Chương 1*

# **TỔNG QUAN VỀ PHÂN TÍCH VÀ DỰ BÁO KINH TẾ**

### **1.1. Khái niệm**

Dự báo đã hình thành từ đầu những năm 60 của thế kỉ 20. Khoa học dự báo với tư cách một ngành khoa học độc lập có hệ thống lí luận, phương pháp luận và phương pháp hệ riêng nhằm nâng cao tính hiệu quả của dự báo. Người ta thường nhấn mạnh rằng một phương pháp tiếp cận hiệu quả đối với dự báo là phần quan trọng trong hoạch định. Khi các nhà quản trị lên kế hoạch, trong hiện tại họ xác định hướng tương lai cho các hoạt động mà họ sẽ thực hiện. Bước đầu tiên trong hoạch định là dự báo hay là ước lượng nhu cầu tương lai cho sản phẩm hoặc dịch vụ và các nguồn lực cần thiết để sản xuất sản phẩm hoặc dịch vụ đó.

Như vậy, dự báo là một khoa học và nghệ thuật tiên đoán những sự việc sẽ xảy ra trong tương lai, trên cơ sở phân tích khoa học về các dữ liệu đã thu thập được.

Khi tiến hành dự báo, ta căn cứ vào việc thu thập, xử lý số liệu trong quá khứ và hiện tại để xác định xu hướng vận động của các hiện tượng trong tương lai nhờ vào một số mô hình toán học.

Dự báo có thể là một dự đoán chủ quan hoặc trực giác về tương lai. Nhưng để cho dự báo được chính xác hơn, người ta cố loại trừ những tính chủ quan của người dự báo.

Ngày nay, dự báo là một nhu cầu không thể thiếu được của mọi hoạt động kinh tế - xã hội, khoa học - kỹ thuật, được tất cả các ngành khoa học quan tâm nghiên cứu.

## **1.2. Ý nghĩa và vai trò của phân tích và dự báo trong quá trình ra quyết định kinh doanh**

### **1.2.1. Ý nghĩa**

- Dùng để dự báo các mức độ tương lai của hiện tượng, qua đó giúp các nhà quản trị doanh nghiệp chủ động trong việc đề ra các kế hoạch và các quyết định cần thiết phục vụ cho quá trình sản xuất kinh doanh, đầu tư, quảng bá, quy mô sản xuất, kênh phân phối sản phẩm, nguồn cung cấp tài chính... và chuẩn bị đầy đủ điều kiện cơ sở vật chất, kỹ thuật cho sự phát triển trong thời gian tới (kế hoạch cung cấp các yếu tố đầu vào như: lao động, nguyên vật liệu, tư liệu lao động... cũng như các yếu tố đầu ra dưới dạng sản phẩm vật chất và dịch vụ).

- Trong các doanh nghiệp, nếu công tác dự báo được thực hiện một cách nghiêm túc còn tạo điều kiện nâng cao khả năng cạnh tranh trên thị trường.

- Dự báo chính xác sẽ giảm bớt mức độ rủi ro cho doanh nghiệp nói riêng và toàn bộ nền kinh tế nói chung.

- Dự báo chính xác là căn cứ để các nhà hoạch định hoạch định các chính sách phát triển kinh tế, văn hoá, xã hội trong toàn bộ nền kinh tế quốc dân.

- Nhờ có dự báo, các chính sách kinh tế, các kế hoạch và chương trình phát triển kinh tế được xây dựng có cơ sở khoa học và mang lại hiệu quả kinh tế cao.

- Nhờ có dự báo thường xuyên và kịp thời, các nhà quản trị doanh nghiệp có khả năng kịp thời đưa ra những biện pháp điều chỉnh các hoạt động kinh tế của đơn vị mình nhằm thu được hiệu quả sản xuất kinh doanh cao nhất.

### **1.2.2. Vai trò**

- Dự báo tạo ra lợi thế cạnh tranh.

- Công tác dự báo là một bộ phận không thể thiếu trong hoạt động của các doanh nghiệp, trong từng phòng ban như: phòng Kinh doanh hoặc Marketing, phòng Sản xuất hoặc phòng Nhân sự, phòng Kế toán – tài chính.

## **1.3. Các loại dự báo**

### **1.3.1. Căn cứ vào độ dài thời gian dự báo**

Dự báo có thể phân thành ba loại:

- *Dự báo dài hạn*: Là những dự báo có thời gian dự báo từ 5 năm trở lên. Thường dùng để dự báo những mục tiêu, chiến lược về kinh tế chính trị, khoa học - kỹ thuật trong thời gian dài ở tầm vĩ mô.

- *Dự báo trung hạn*: Là những dự báo có thời gian dự báo từ 3 đến 5 năm. Thường phục vụ cho việc xây dựng những kế hoạch trung hạn về kinh tế, văn hoá, xã hội... ở tầm vi mô và vĩ mô.

- *Dự báo ngắn hạn*: Là những dự báo có thời gian dự báo dưới 3 năm, loại dự báo này thường dùng để dự báo hoặc lập các kế hoạch kinh tế, văn hoá, xã hội chủ yếu ở tầm vi mô và vĩ mô trong khoảng thời gian ngắn nhằm phục vụ cho công tác chỉ đạo kịp thời.

Cách phân loại này chỉ mang tính tương đối tùy thuộc vào từng loại hiện tượng để quy định khoảng cách thời gian cho phù hợp với

loại hiện tượng đó. Ví dụ trong dự báo kinh tế, dự báo dài hạn là những dự báo có tầm dự báo trên 5 năm, nhưng trong dự báo thời tiết, khí tượng học chỉ là một tuần. Thang thời gian đối với dự báo kinh tế dài hơn nhiều so với thang thời gian dự báo thời tiết. Vì vậy, thang thời gian có thể đo bằng những đơn vị thích hợp (ví dụ: quý, năm đối với dự báo kinh tế và ngày đối với dự báo dự báo thời tiết).

### ***1.3.2. Dựa vào các phương pháp dự báo***

Dự báo có thể chia thành ba nhóm:

- Dự báo bằng phương pháp chuyên gia: Loại dự báo này được tiến hành trên cơ sở tổng hợp, xử lý ý kiến của các chuyên gia thông thạo với hiện tượng được nghiên cứu, từ đó có phương pháp xử lý thích hợp để ra các dự đoán, các dự đoán này được cân nhắc và đánh giá chủ quan từ các chuyên gia. Phương pháp này có ưu thế trong trường hợp dự đoán những hiện tượng hay quá trình bao quát rộng, phức tạp, chịu sự chi phối của khoa học - kỹ thuật, sự thay đổi của môi trường, thời tiết, chiến tranh trong khoảng thời gian dài. Một cải tiến của phương pháp Delphi là phương pháp dự báo dựa trên cơ sở sử dụng một tập hợp những đánh giá của một nhóm chuyên gia. Mỗi chuyên gia được hỏi ý kiến và rồi dự báo của họ được trình bày dưới dạng thống kê tóm tắt. Việc trình bày những ý kiến này được thực hiện một cách gián tiếp (không có sự tiếp xúc trực tiếp) để tránh những sự tương tác trong nhóm nhỏ qua đó tạo nên những sai lệch nhất định trong kết quả dự báo. Sau đó, người ta yêu cầu các chuyên gia duyệt xét lại những dự báo của họ trên cơ sở tóm tắt tất cả các dự báo, có thể có những bổ sung thêm.

- Dự báo theo phương trình hồi quy: Theo phương pháp này, mức độ cần dự báo phải được xây dựng trên cơ sở xây dựng mô hình

hồi quy, mô hình này được xây dựng phù hợp với đặc điểm và xu thế phát triển của hiện tượng nghiên cứu. Để xây dựng mô hình hồi quy, đòi hỏi phải có tài liệu về hiện tượng cần dự báo và các hiện tượng có liên quan. Loại dự báo này thường được sử dụng để dự báo trung hạn và dài hạn ở tầm vĩ mô.

- Dự báo dựa vào dãy số thời gian: Là dựa trên cơ sở dãy số thời gian phản ánh sự biến động của hiện tượng ở những thời gian đã qua để xác định mức độ của hiện tượng trong tương lai.

### ***1.3.3. Căn cứ vào nội dung (đối tượng dự báo)***

Có thể chia dự báo thành: Dự báo khoa học, dự báo kinh tế, dự báo xã hội, dự báo tự nhiên, thiên văn học...

- Dự báo khoa học: Là dự kiến, tiên đoán về những sự kiện, hiện tượng, trạng thái nào đó có thể hay nhất định sẽ xảy ra trong tương lai. Theo nghĩa hẹp hơn, đó là sự nghiên cứu khoa học về những triển vọng của một hiện tượng nào đó, chủ yếu là những đánh giá số lượng và chỉ ra khoảng thời gian mà trong đó hiện tượng có thể diễn ra những biến đổi.

- Dự báo kinh tế: Là khoa học dự báo các hiện tượng kinh tế trong tương lai. Dự báo kinh tế được coi là giai đoạn trước của công tác xây dựng chiến lược phát triển kinh tế - xã hội và dự án kế hoạch dài hạn; không đặt ra những nhiệm vụ cụ thể, nhưng chứa đựng những nội dung cần thiết làm căn cứ để xây dựng những nhiệm vụ đó. Dự báo kinh tế bao trùm sự phát triển kinh tế và xã hội của đất nước có tính đến sự phát triển của tình hình thế giới và các quan hệ quốc tế. Thường được thực hiện chủ yếu theo những hướng sau: dân số, nguồn lao động, việc sử dụng và tái sản xuất chúng, năng suất lao động; tái sản xuất xã hội trước hết là vốn sản xuất cố định: sự phát triển của

cách mạng khoa học – kỹ thuật và công nghệ và khả năng ứng dụng vào kinh tế; mức sống của nhân dân, sự hình thành các nhu cầu phi sản xuất, động thái và cơ cấu tiêu dùng, thu nhập của nhân dân; động thái kinh tế quốc dân và sự chuyển dịch cơ cấu (nhịp độ, tỉ lệ, hiệu quả); sự phát triển các khu vực và ngành kinh tế (khối lượng động thái, cơ cấu, trình độ kỹ thuật, bộ máy, các mối liên hệ liên ngành); phân vùng sản xuất, khai thác tài nguyên thiên nhiên và phát triển các vùng kinh tế trong nước, các mối liên hệ liên vùng; dự báo sự phát triển kinh tế thế giới. Các kết quả dự báo kinh tế cho phép hiểu rõ đặc điểm của các điều kiện kinh tế - xã hội để đặt chiến lược phát triển kinh tế đúng đắn, xây dựng các chương trình, kế hoạch phát triển một cách chủ động, đạt hiệu quả cao và vững chắc.

- Dự báo xã hội: Dự báo xã hội là khoa học nghiên cứu những triển vọng cụ thể của một hiện tượng, một sự biến đổi, một quá trình xã hội, để đưa ra dự báo hay dự đoán về tình hình diễn biến, phát triển của một xã hội.

- Dự báo tự nhiên, thiên văn học, loại dự báo này thường bao gồm:

+ *Dự báo thời tiết*: Thông báo thời tiết dự kiến trong một thời gian nhất định trên một vùng nhất định. Trong dự báo thời tiết có dự báo chung, dự báo khu vực, dự báo địa phương,... Về thời gian, có dự báo thời tiết ngắn (1-3 ngày) và dự báo thời tiết dài (tới một năm).

+ *Dự báo thủy văn*: Là loại dự báo nhằm tính để xác định trước sự phát triển các quá trình, hiện tượng thủy văn xảy ra ở các sông, hồ, dựa trên các tài liệu liên quan tới khí tượng thủy văn. Dự báo thủy văn dựa trên sự hiểu biết những quy luật phát triển của các quá trình, khí tượng thủy văn, dự báo sự xuất hiện của hiện tượng hay yếu tố cần quan tâm. Căn cứ thời gian dự kiến, dự báo thủy văn được chia thành



dự báo thủy văn hạn ngắn (thời gian không quá 2 ngày), hạn vừa (từ 2 đến 10 ngày); dự báo thủy văn mùa (thời gian dự báo vài tháng); cấp báo thủy văn: thông tin khẩn cấp về hiện tượng thủy văn gây nguy hiểm. Theo mục đích dự báo, có các loại: dự báo thủy văn phục vụ thi công, phục vụ vận tải, phục vụ phát điện,... Theo yếu tố dự báo, có: dự báo lưu lượng lớn nhất, nhỏ nhất, dự báo lũ, ...

+ *Dự báo địa lý*: Là việc nghiên cứu về hướng phát triển của môi trường địa lý trong tương lai, nhằm đề ra trên cơ sở khoa học những giải pháp sử dụng hợp lý và bảo vệ môi trường.

+ *Dự báo động đất*: Là loại dự báo trước địa điểm và thời gian có khả năng xảy ra động đất. Động đất không đột nhiên xảy ra mà là một quá trình tích lũy lâu dài, có thể hiện ra trước bằng những biến đổi địa chất, những hiện tượng vật lý, những trạng thái sinh học bất thường ở động vật,... Việc dự báo thực hiện trên cơ sở nghiên cứu bản đồ phân vùng động đất và những dấu hiệu báo trước. Cho đến nay, chưa thể dự báo chính xác về thời gian động đất sẽ xảy ra.

## **1.4. Các phương pháp dự báo**

### ***1.4.1. Phương pháp dự báo định tính***

Các phương pháp này dựa trên cơ sở nhận xét của những nhân tố nhân-quả, dựa theo doanh số của từng sản phẩm hay dịch vụ riêng biệt và dựa trên những ý kiến về các khả năng có liên hệ của những nhân tố nhân-quả này trong tương lai. Những phương pháp này có liên quan đến mức độ phức tạp khác nhau, từ những khảo sát ý kiến được tiến hành một cách khoa học để nhận biết về các sự kiện tương lai.

♦ *Ưu điểm:*

Dễ dàng thực hiện, không đòi hỏi kiến thức về các mô hình toán hoặc kinh tế lượng, thường được chấp nhận.

♦ *Nhược điểm:* Mang tính chủ quan rất cao, không chuẩn, mất nhiều năm để trở thành người có khả năng phán đoán đúng. Không có phương pháp hệ thống để đánh giá và cải thiện mức độ chính xác.

Dưới đây là các dự báo định tính thường dùng:

*1.4.1.1. Lấy ý kiến của ban điều hành*

Phương pháp này được sử dụng rộng rãi ở các doanh nghiệp. Khi tiến hành dự báo, họ lấy ý kiến của các nhà quản trị cấp cao, những người phụ trách các công việc, các bộ phận quan trọng của doanh nghiệp, và sử dụng các số liệu thống kê về những chỉ tiêu tổng hợp: doanh số, chi phí, lợi nhuận...Ngoài ra, cần lấy thêm ý kiến của các chuyên gia về marketing, tài chính, sản xuất, kỹ thuật.

Ưu điểm của phương pháp này là: Thu thập được nhiều kinh nghiệm từ nhiều chuyên gia khác nhau.

Nhược điểm lớn nhất của phương pháp này là có tính chủ quan của các thành viên và ý kiến của người có chức vụ cao nhất thường chi phối ý kiến của những người khác.

*1.4.1.2. Lấy ý kiến của người bán hàng*

Ưu điểm của phương pháp này là: Những người bán hàng tiếp xúc thường xuyên với khách hàng, do đó họ hiểu rõ nhu cầu, thị hiếu của người tiêu dùng. Họ có thể dự đoán được lượng hàng tiêu thụ tại khu vực mình phụ trách.

Tập hợp ý kiến của nhiều người bán hàng tại nhiều khu vực khác nhau, ta có được lượng dự báo tổng hợp về nhu cầu đối với loại sản phẩm đang xét.

Nhược điểm của phương pháp này là phụ thuộc vào đánh giá chủ quan của người bán hàng. Một số có khuynh hướng lạc quan đánh giá cao lượng hàng bán ra của mình. Ngược lại, một số khác lại muốn giảm xuống để dễ đạt định mức.

#### *1.4.1.3. Phương pháp chuyên gia (Delphi)*

Phương pháp này thu thập ý kiến của các chuyên gia trong hoặc ngoài doanh nghiệp theo những mẫu câu hỏi được in sẵn và được thực hiện như sau:

- Mỗi chuyên gia được phát một thư yêu cầu trả lời một số câu hỏi phục vụ cho việc dự báo.

- Nhân viên dự báo tập hợp các câu trả lời, sắp xếp chọn lọc và tóm tắt lại các ý kiến của các chuyên gia.

- Dựa vào bảng tóm tắt này, nhân viên dự báo lại tiếp tục nêu ra các câu hỏi để các chuyên gia trả lời tiếp.

- Tập hợp các ý kiến mới của các chuyên gia. Nếu chưa thỏa mãn thì tiếp tục quá trình nêu trên cho đến khi đạt yêu cầu dự báo.

Ưu điểm của phương pháp này là tránh được các liên hệ cá nhân với nhau, không xảy ra va chạm giữa các chuyên gia và họ không bị ảnh hưởng bởi ý kiến của một người nào đó có ưu thế trong số người được hỏi ý kiến.

#### *1.4.1.4. Phương pháp điều tra người tiêu dùng*

Phương pháp này sẽ thu thập nguồn thông tin từ đối tượng người tiêu dùng về nhu cầu hiện tại cũng như tương lai. Cuộc điều tra nhu

cầu được thực hiện bởi những nhân viên bán hàng hoặc nhân viên nghiên cứu thị trường. Họ thu thập ý kiến khách hàng thông qua phiếu điều tra, phỏng vấn trực tiếp hay điện thoại... Cách tiếp cận này không những giúp cho doanh nghiệp về dự báo nhu cầu mà cả trong việc cải tiến thiết kế sản phẩm. Phương pháp này mất nhiều thời gian, việc chuẩn bị phức tạp, khó khăn và tốn kém, có thể không chính xác trong các câu trả lời của người tiêu dùng.

♦ *Ưu điểm:* Cách tốt nhất để dự báo nhu cầu, sở thích của người tiêu dùng thông qua dự định mua sắm của họ, điều tra được thị hiếu của khách hàng để cải tiến sản phẩm.

♦ *Nhược điểm:* Phù hợp cho các sản phẩm công nghiệp, tính chính xác của dữ liệu.

#### ***1.4.2. Phương pháp dự báo định lượng***

Mô hình dự báo định lượng dựa trên số liệu quá khứ, những số liệu này giả sử có liên quan đến tương lai và có thể tìm thấy được. Tất cả các mô hình dự báo theo định lượng có thể sử dụng thông qua chuỗi thời gian và các giá trị này được quan sát đo lường các giai đoạn theo từng chuỗi .

♦ *Ưu điểm:*

- Kết quả dự báo hoàn toàn khách quan.
- Có phương pháp đo lường độ chính xác dự báo.
- Tốn ít thời gian để tìm ra kết quả dự báo.

♦ *Nhược điểm:*

- Chỉ dự báo tốt trong thời gian ngắn và trung hạn.
- Không có phương pháp nào có thể đưa đầy đủ những yếu tố bên ngoài có tác động đến kết quả dự báo vào mô hình.

**\* Tính chính xác của dự báo**

Tính chính xác của dự báo đề cập đến độ chênh lệch của dự báo với số liệu thực tế. Bởi vì dự báo được hình thành trước khi số liệu thực tế xảy ra, vì vậy tính chính xác của dự báo chỉ có thể đánh giá sau khi thời gian đã qua đi. Nếu dự báo càng gần với số liệu thực tế, ta nói dự báo có độ chính xác cao và lỗi trong dự báo càng thấp.

Gọi:  $Y_t$ : Giá trị thực tại giai đoạn  $t$ .

$\hat{Y}_t$ : Giá trị dự báo tại giai đoạn  $t$ .

$n$ : Số giai đoạn.

Sai số dự báo:  $e_t = Y_t - \hat{Y}_t$

Nếu một mô hình được đánh giá là tốt thì sai số dự báo phải tương đối nhỏ.

Tiêu chí	Công thức tính
1. Sai số trung bình	$ME = \frac{\sum e_t}{n}$
2. Sai số trung bình tuyệt đối	$ME = \frac{\sum  e_t }{n}$
3. Sai số % trung bình	$MPE = \frac{\sum \frac{e_t}{Y_t}}{n}$
4. Sai số % trung bình tuyệt đối	$MPE = \frac{\sum \left  \frac{e_t}{Y_t} \right }{n}$
5. Sai số bình phương trung bình	$MSE = \frac{\sum e_t^2}{n}$

### 1.4.2.1. Dự báo ngắn hạn

Dự báo ngắn hạn ước lượng tương lai trong thời gian ngắn. Dự báo ngắn hạn cung cấp cho các nhà quản lý tác nghiệp những thông tin để đưa ra quyết định về các vấn đề như:

- Cần dự trữ bao nhiêu đối với một loại sản phẩm cụ thể nào đó cho tháng tới?

- Lên lịch sản xuất từng loại sản phẩm cho tháng tới như thế nào?

- Số lượng nguyên vật liệu cần đặt hàng để nhận vào tuần tới là bao nhiêu?

#### \* Dự báo sơ bộ

Thích hợp với các doanh nghiệp mới thành lập vì có rất ít dữ liệu. Dự báo sơ bộ khẳng định rằng các giai đoạn gần nhất là ước lượng tốt nhất cho tương lai.

Mô hình đơn giản là:  $\hat{Y}_{t+1} = Y_t$

$Y_t$ : Số liệu ở thời kỳ  $t$ .

$\hat{Y}_{t+1}$ : Số liệu ở thời kỳ  $t+1$ .

Mô hình dự báo sơ bộ là loại dự báo nhanh, không cần chi phí và dễ sử dụng. Ví dụ như:

- Sử dụng số liệu hàng bán ngày hôm nay làm dự báo cho lượng hàng bán ở ngày mai.

- Sử dụng số liệu ngày này ở năm rồi như là dự báo lượng hàng bán cho ngày ấy ở năm nay.

Mô hình dự báo sơ bộ quá đơn giản cho nên thường hay gặp những sai sót trong dự báo.

### *\* Phương pháp bình quân di động có quyền số*

Trong phương pháp bình quân di động này, chúng ta xem vai trò của các số liệu trong quá khứ là khác nhau. Vì trong một vài trường hợp, các số liệu này có ảnh hưởng khác nhau trên kết quả dự báo, vì thế, người ta thích sử dụng quyền số không đồng đều cho các số liệu quá khứ. Quyền số hay trọng số là các con số được gán cho các số liệu quá khứ để chỉ mức độ quan trọng của chúng ảnh hưởng đến kết quả dự báo. Quyền số lớn được gán cho số liệu gần với kỳ dự báo nhất để ám chỉ ảnh hưởng của nó là lớn nhất. Việc chọn các quyền số phụ thuộc vào kinh nghiệm và sự nhạy cảm của người dự báo.

Công thức tính toán:

$$F_t = \frac{\sum_{i=1}^n A_{t-i} k_i}{\sum_{i=1}^n k_i}$$

Với:  $F_t$ : Dự báo thời kỳ thứ  $t$ .

$A_{t-i}$ : Số liệu thực tế thời kỳ trước ( $i = 1, 2, \dots, n$ ).

$k_i$ : Quyền số tương ứng ở thời kỳ  $i$ .

Cả hai phương pháp bình quân di động và bình quân di động có quyền số đều có ưu điểm là san bằng được các biến động ngẫu nhiên trong dãy số. Tuy vậy, chúng đều có nhược điểm sau:

- Do việc san bằng các biến động ngẫu nhiên nên làm giảm độ nhạy cảm đối với những thay đổi thực đã được phản ánh trong dãy số.

- Số bình quân di động chưa cho chúng ta xu hướng phát triển của dãy số một cách tốt nhất. Nó chỉ thể hiện sự vận động trong quá khứ chứ chưa thể kéo dài sự vận động đó trong tương lai.

### *\* Phương pháp điều hòa mũ*

Điều hòa mũ đưa ra các dự báo cho giai đoạn trước và thêm vào đó một lượng điều chỉnh để có được lượng dự báo cho giai đoạn kế tiếp. Sự điều chỉnh này là một tỷ lệ nào đó của sai số dự báo ở giai đoạn trước và được tính bằng cách nhân số dự báo của giai đoạn trước với hệ số nằm giữa 0 và 1. Hệ số này gọi là hệ số điều hòa.

Công thức tính như sau:  $F_t = F_{t-1} + \alpha (A_{t-1} - F_{t-1})$

Trong đó:  $F_t$ : Dự báo cho giai đoạn thứ t, giai đoạn kế tiếp.

$F_{t-1}$ : Dự báo cho giai đoạn thứ t-1, giai đoạn trước.

$A_{t-1}$ : Số liệu thực tế của giai đoạn thứ t-1.

*Ví dụ:* Ông B trong ví dụ sau, nói với nhà phân tích ở công ty mẹ rằng, phải dự báo nhu cầu hàng tuần cho dự trữ trong nhà kho của ông. Nhà phân tích đề nghị ông B xem xét việc sử dụng phương pháp điều hòa mũ với các hệ số điều hòa 0,1 ; 0,2 ; 0,3. Ông B quyết định so sánh mức độ chính xác của dự báo ứng với từng hệ số cho giai đoạn 10 tuần lễ gần đây nhất.

### *Kết quả bài toán:*

Chúng ta tính toán dự báo hàng tuần cho tuần lễ thứ 8 đến tuần lễ thứ 17. Tất cả dự báo của tuần lễ thứ 7 được chọn một cách ngẫu nhiên, dự báo khởi đầu thì rất cần thiết trong phương pháp điều hòa mũ. Thông thường, người ta cho các dự báo này bằng với giá trị thực của giai đoạn.

Tính toán mẫu - dự báo cho tuần lễ thứ 8:

$$F_8 = 85 + 0,1(85-85) = 85$$

$$F_9 = 85 + 0,1(102 - 85) = 86,7$$

$$F_9 = 85 + 0,2(102 - 85) = 88,4$$



Sau đó ta tính sai số trung bình ME cho ba dự báo nói trên:

Tuần lễ	Nhu cầu dự trữ thực tế	$\alpha = 0,1$		$\alpha = 0,2$		$\alpha = 0,3$	
		Dự báo	$e^t$	Dự báo	$e^t$	Dự báo	$e^t$
8	102	85,0	17,0	85,0	17,0	85,0	17,0
9	110	86,7	23,3	88,4	21,6	90,1	19,9
10	90	89,0	1,0	92,7	2,7	96,1	6,1
11	105	89,1	15,9	92,2	12,8	94,3	10,7
12	95	90,7	4,3	94,8	0,2	97,5	2,5
13	115	91,1	23,9	94,8	20,2	96,8	18,2
14	120	93,5	26,5	98,8	21,2	102,3	17,7
15	80	96,2	16,2	103,0	23,0	107,6	27,8
16	95	94,6	0,4	98,4	3,4	99,3	4,3
17	100	94,6	5,4	97,7	2,3	98,0	2,0
Tổng sai số tuyệt đối			133,9		124,4		126,0
ME			13,39		12,44		12,6

– Hệ số điều hòa  $\alpha = 0,2$  cho chúng ta độ chính xác cao hơn  $\alpha = 0,1$  và  $\alpha = 0,3$ .

Sử dụng  $\alpha = 0,2$  (vì  $ME_{\min}$ ) để tính dự báo cho tuần thứ 18:

$$\begin{aligned}
 F_{18} &= F_{17} + \alpha (A_{17} - F_{17}) \\
 &= 97,7 + 0,2 (100 - 97,7) \\
 &= 98,2 \text{ hay } 982 \text{ triệu đồng.}
 \end{aligned}$$

*\* Phương pháp điều hòa mũ theo xu hướng*

Chúng ta thường xem xét kế hoạch ngắn hạn, thì mùa vụ và xu hướng là nhân tố không quan trọng. Khi chúng ta chuyển từ dự báo

ngắn hạn sang dự báo trung hạn thì mùa vụ và xu hướng trở nên quan trọng hơn. Kết hợp nhân tố xu hướng vào dự báo điều hòa mũ được gọi là điều hòa mũ theo xu hướng hay điều hòa đôi.

Vì ước lượng cho số trung bình và ước lượng cho xu hướng cho số trung bình và hệ số điều hòa  $\alpha$  được điều hòa cả hai. Hệ số điều hòa cho xu hướng được sử dụng trong mô hình này là  $\beta$

Công thức tính toán như sau:

*a. Chuỗi san bằng số mũ hoặc ước lượng giá trị hiện hành*

$$F_t = \alpha A_t + (1-\alpha)(F_{t-1} + T_{t-1})$$

*b. Ước lượng xu thế*

$$T_t = \beta(F_t - F_{t-1}) + (1-\beta)T_{t-1}$$

*c. Dự báo cho m giai đoạn tiếp theo*

$$H_{t+m} = F_t + mT_t$$

Trong đó:

$F_t$ : Giá trị san bằng mới.

$\alpha$ : Hệ số san bằng mũ ( $0 < \alpha < 1$ ).

$\beta$ : Hệ số san bằng số mũ để ước lượng xu thế ( $0 < \beta < 1$ ).

$A_t$ : Giá trị thực tại thời điểm t.

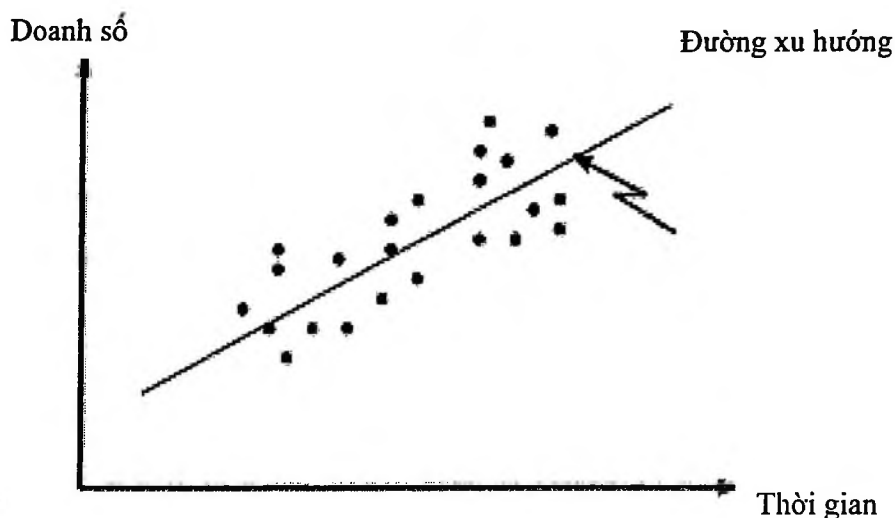
m: Số lượng giai đoạn dự báo trong tương lai.

$H_{t+m}$ : Giá trị dự báo cho m giai đoạn tiếp theo.

#### *1.4.2.2. Dự báo dài hạn*

Dự báo dài hạn là ước lượng tương lai trong thời gian dài, thường hơn một năm. Dự báo dài hạn rất cần thiết trong quản trị sản

xuất để trợ giúp các quyết định chiến lược về hoạch định sản phẩm, quy trình công nghệ và các phương tiện sản xuất. Ví dụ như:



- Thiết kế sản phẩm mới.

Xác định năng lực sản xuất cần thiết là bao nhiêu? Máy móc, thiết bị nào cần sử dụng và chúng được đặt ở đâu?

- Lên lịch trình cho những nhà cung ứng theo các hợp đồng cung cấp nguyên vật liệu dài hạn.

Dự báo dài hạn có thể được xây dựng bằng cách vẽ một đường thẳng đi xuyên qua các số liệu quá khứ và kéo dài nó đến tương lai. Dự báo trong giai đoạn kế tiếp có thể được vẽ vượt ra khỏi đồ thị thông thường. Phương pháp tiếp cận theo kiểu đồ thị đối với dự báo dài hạn có thể dùng trong thực tế, nhưng điểm không thuận lợi của nó là vấn đề vẽ một đường tương ứng hợp lý nhất đi qua các số liệu quá khứ này.

Phân tích hồi qui sẽ cung cấp cho chúng ta một phương pháp làm việc chính xác để xây dựng đường dự báo theo xu hướng.

### *\* Phương pháp hồi qui tuyến tính*

Phân tích hồi qui tuyến tính là một mô hình dự báo thiết lập mối quan hệ giữa biến phụ thuộc với hai hay nhiều biến độc lập. Trong phần này, chúng ta chỉ xét đến một biến độc lập duy nhất. Nếu số liệu là một chuỗi theo thời gian thì biến độc lập là giai đoạn thời gian và biến phụ thuộc thông thường là doanh số bán ra hay bất kỳ chỉ tiêu nào khác mà ta muốn dự báo.

Mô hình này có công thức:  $y = ax + b$

$$a = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$$

Trong đó: y: Biến phụ thuộc cần dự báo.

x: Biến độc lập.

a: Độ dốc của đường xu hướng.

b: Tung độ gốc.

n: Số lượng quan sát.

Trong trường hợp biến độc lập x được trình bày thông qua từng giai đoạn theo thời gian và chúng phải cách đều nhau (như: x = 0. Vì vậy  $\sum 2002, 2003, 2004 \dots$ ) thì ta có thể điều chỉnh lại để sao cho việc tính toán sẽ trở nên đơn giản và dễ dàng hơn nhiều.

Nếu có một số lẻ lượng mốc thời gian: chẳng hạn x = 0 và  $\sum$  là 5, thì giá trị của x được ấn định như sau: -2, -1, 0, 1, 2 và như thế giá trị của x được sử dụng cho dự báo trong năm tới là +3.

Nếu có một số chẵn lượng mốc thời gian: chẳng hạn  $x = 0$  và  $\Sigma$  là 6 thì giá trị của  $x$  được ấn định là: -5, -3, -1, 1, 3, 5. Như thế, giá trị của  $x$  được dùng cho dự báo trong năm tới là +7.

*Ví dụ:* Một hãng sản xuất loại động cơ điện tử cho các van khởi động trong ngành công nghiệp, nhà máy hoạt động gần hết công suất suốt một năm nay. Ông J, người quản lý nhà máy nghĩ rằng sự tăng trưởng trong doanh số bán ra vẫn còn tiếp tục và ông ta muốn xây dựng một dự báo dài hạn để hoạch định nhu cầu về lượng hàng bán ra trong 3 năm tới. Số lượng bán ra trong 10 năm qua được ghi lại như sau:

Năm	Số lượng bán	Năm	Số lượng bán
1	1000	6	2000
2	1300	7	2200
3	1800	8	2600
4	2000	9	2900
5	2000	10	3200

Kết quả bài toán:

Ta xây dựng bảng tính để thiết lập các giá trị:

Năm	Lượng bán (y)	Thời gian (x)	$x^2$	xy
1	1000	-9	81	-9000
2	1300	-7	49	-9100
3	1800	-5	25	-9000
4	2000	-3	9	-6000
5	2000	-1	1	-2000
6	2000	1	1	2000
7	2200	3	9	6600
8	2600	5	25	13000
9	2900	7	49	20300
10	3200	9	81	28800
<b>Tổng</b>	<b>21000</b>	<b>0</b>	<b>330</b>	<b>35600</b>

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{\sum xy}{\sum x^2} = \frac{35600}{330} = 107,8$$

$$b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = \frac{\sum y}{n} = \frac{21000}{10} = 2100$$

- Dùng phương trình hồi qui tuyến tính để dự báo hàng bán ra trong tương lai:

$$y = ax + b = 107,8x + 2100$$

Để dự báo cho hàng bán ra trong 3 năm tới ta thay giá trị của x lần lượt là 11, 13, 15 vào phương trình.

$$y_{11} = 107,8 \times 11 + 2.100 = 3285 \approx 3290 \text{ đơn vị}$$

$$y_{12} = 107,8 \times 13 + 2.100 = 3501 \approx 3500 \text{ đơn vị}$$

$$y_{13} = 107,8 \times 15 + 2.100 = 3717 \approx 3720 \text{ đơn vị}$$

Trường hợp biến độc lập không phải là biến thời gian, hồi qui tuyến tính là một nhóm các mô hình dự báo được gọi là mô hình nhân quả. Mô hình này đưa ra các dự báo sau khi thiết lập và đo lường các biến phụ thuộc với một hay nhiều biến độc lập.

*Ví dụ:* Ông B, nhà tổng quản lý của công ty kỹ nghệ chính xác nghĩ rằng các dịch vụ kỹ nghệ của công ty ông ta được cung ứng cho các công ty xây dựng thì có quan hệ trực tiếp đến số hợp đồng xây dựng trong vùng của ông ta. Ông B yêu cầu kỹ sư dưới quyền tiến hành phân tích hồi qui tuyến tính dựa trên các số liệu quá khứ và vạch ra kế hoạch như sau:

- Xây dựng một phương trình hồi qui cho dự báo mức độ nhu cầu về dịch vụ của công ty ông.

- Sử dụng phương trình hồi qui để dự báo mức độ nhu cầu trong 4 quý tới. Ước lượng trị giá hợp đồng 4 quý tới là 260, 290, 300 và 270 (ĐVT:10 triệu đồng).

- Xác định mức độ chặt chẽ, các mối liên hệ giữa nhu cầu và hợp đồng xây dựng được đưa ra.

Biết số liệu từng quý trong 2 năm qua cho trong bảng (đơn vị: 10 triệu đồng).

Năm	Quý	Nhu cầu của công ty	Trị giá hợp đồng thực hiện
1	1	8	150
	2	10	170
	3	15	190
	4	9	170
2	1	12	180
	2	13	190
	3	12	200
	4	16	220

*Kết quả bài toán:* Xây dựng phương trình hồi qui: Ông A xây dựng bảng tính như sau:

Thời gian	Nhu cầu (y)	Trị giá hợp đồng (x)	$x^2$	xy	$y^2$
1	8	150	22500	1.200	64
2	10	170	28900	1.700	100
3	15	190	36100	2.850	225
4	9	170	28900	1.530	81
5	12	180	32400	2.160	144
6	13	190	36100	2.470	169
7	12	200	40000	2.400	144
8	16	220	48400	3.520	256
<b>Tổng</b>	<b>95</b>	<b>1.470</b>	<b>273300</b>	<b>17.830</b>	<b>1.183</b>

Sử dụng công thức ta tính toán được hệ số  $a = 0,1173$ ;  $b = -9,671$

Phương trình hồi qui tìm được là:  $y = 0,1173x - 9,671$

Dự báo nhu cầu cho 4 quý tới: Ông A dự báo nhu cầu của công ty bằng cách sử dụng phương trình trên cho 4 quý tới như sau:

$$y_1 = (0,1173 \times 260) - 9,671 = 20,827$$

$$y_2 = (0,1173 \times 290) - 9,671 = 24,346$$

$$y_3 = (0,1173 \times 300) - 9,671 = 25,519$$

$$y_4 = (0,1173 \times 270) - 9,671 = 22,000$$

Dự báo tổng cộng cho năm tới là:

$$y = y_1 + y_2 + y_3 + y_4$$

$$= 20,827 + 24,346 + 25,519 + 22,000 \approx 92,7 \text{ triệu đồng} \approx 930 \text{ triệu đồng}$$

Đánh giá mức độ chặt chẽ mối liên hệ của nhu cầu với số lượng hợp đồng xây dựng:

$$r = \frac{n\sum xy - \sum x \sum y}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$
$$= \frac{8 \times 17.830 - 1.470 \times 95}{(8 \times 273.300 - 1470^2)(8 \times 1.183 - 95^2)} = \frac{2990}{3345,8} \approx 0,894$$

$r^2 = 0,799$ ; Trong đó  $r$  là hệ số tương quan và  $r^2$  là hệ số xác định.

Rõ ràng là số lượng hợp đồng xây dựng có ảnh hưởng khoảng 80% ( $r^2 = 0,799$ ) của biến số được quan sát về nhu cầu hàng quý của công ty.



Hệ số tương quan  $r$  giải thích tầm quan trọng tương đối của mối quan hệ giữa  $y$  và  $x$ ; dấu của  $r$  cho biết hướng của mối quan hệ giữa các biến. Dấu trị tuyệt đối của  $r$  chỉ cường độ của mối quan hệ,  $r$  có giá trị (từ -1) của  $r$  luôn luôn cùng với dấu của hệ số  $a$ . Nếu  $r$  âm chỉ ra rằng giá trị của  $y$  và  $x$  có khuynh hướng đi ngược chiều nhau, nếu  $r$  dương cho thấy giá trị của  $y$  và  $x$  đi cùng chiều nhau.

Dưới đây là vài giá trị của  $r$ :

$r = -1$ . Quan hệ ngược chiều hoàn toàn, khi  $y$  tăng lên thì  $x$  giảm xuống và ngược lại.

$r = +1$ . Quan hệ cùng chiều hoàn toàn, khi  $y$  tăng lên thì  $x$  cũng tăng và ngược lại.

$r = 0$ . Không có mối quan hệ giữa  $x$  và  $y$ .

*\* Tính chất mùa vụ trong dự báo chuỗi thời gian*

Loại mùa vụ thông thường là sự lên xuống xảy ra trong vòng một năm và có xu hướng lặp lại hàng năm. Những vụ mùa này xảy ra có thể do điều kiện thời tiết, địa lý hoặc do tập quán của người tiêu dùng khác nhau...

Cách thức xây dựng dự báo với phân tích hồi qui tuyến tính khi vụ mùa hiện diện trong chuỗi số theo thời gian. Ta thực hiện các bước:

- Chọn lựa chuỗi số liệu quá khứ đại diện.
- Xây dựng chỉ số mùa vụ cho từng giai đoạn thời gian.

$$I_i = \frac{\bar{Y}_i}{\bar{Y}_0}$$

Với  $\bar{Y}_i$  : Số bình quân của các thời kỳ cùng tên.

$\bar{Y}_0$ : Số bình quân chung của tất cả các thời kỳ trong dãy số.

$I_i$ : Chỉ số mùa vụ kỳ thứ  $i$ .

- Sử dụng các chỉ số mùa vụ để hóa giải tính chất mùa vụ của số liệu.
- Phân tích hồi qui tuyến tính dựa trên số liệu đã phi mùa vụ.
- Sử dụng phương trình hồi qui để dự báo cho tương lai.
- Sử dụng chỉ số mùa vụ để tái ứng dụng tính chất mùa vụ cho dự báo.

*Ví dụ:* Ông J, nhà quản lý nhà máy động cơ đặc biệt đang cố gắng lập kế hoạch tiền mặt và nhu cầu nguyên vật liệu cho từng quý của năm tới. Số liệu về lượng hàng bán ra trong vòng 3 năm qua phản ánh khá tốt kiểu sản lượng mùa vụ và có thể giống như trong tương lai. Số liệu cụ thể như sau:

Năm	Số lượng bán hàng quý (1.000 đơn vị)			
	Q1	Q2	Q3	Q4
1	520	730	820	530
2	590	810	900	600
3	650	900	1	650

Kết quả bài toán:

Đầu tiên ta tính toán các chỉ số mùa vụ.

Kế tiếp, hóa giải tính chất mùa vụ của số liệu bằng cách chia giá trị của từng quý cho chỉ số mùa vụ tương ứng. Chẳng hạn:  $520/0,809 = 642,8$ ;  $730/1,122 = 605,6 \dots$

Năm	Quý 1	Quý 2	Quý 3	Quý 4	Cả năm
1	520	730	820	530	2.600
2	590	810	900	600	2.900
3	650	900	1.000	650	3.200
Tổng	1.760	2.440	2.720	1.780	8.700
Trung bình quý	586,67	813,33	906,67	593,33	725
Chỉ số mùa vụ	0,809	1,122	1,251	0,818	-

Ta được bảng số liệu như sau:

Năm	Số liệu hàng quý đã phi mùa vụ			
	Quý 1	Quý 2	Quý 3	Quý 4
1	642,8	650,6	655,5	647,9
2	729,2	721,9	719,4	733,5
3	803,5	802,1	799,4	794,6

Chúng ta phân tích hồi qui trên cơ sở số liệu phi mùa vụ (12 quý) và xác định phương trình hồi qui.

Quý	x	y	$x^2$	xy
Q11	1	642,8	1	642,8
Q12	2	650,6	4	1.301,2
Q13	3	655,5	9	1.966,5
Q14	4	647,9	16	2.591,6
Q21	5	729,3	25	3.646,5
Q22	6	721,9	36	4.331,4
Q23	7	719,4	49	5.035,8
Q24	8	733,5	64	5.868,0
Q31	9	803,5	81	7.231,5
Q32	10	802,1	100	8.021,0
Q33	11	799,4	121	8.793,4
Q34	12	794,6	144	8.535,2
Tổng	78	8.700,5	650	58.964,9

Xác định được hệ số  $a = 16,865$  và  $b = 615,421$

Phương trình có dạng:  $y = 16,865x + 615,421$

Bây giờ chúng ta thay thế giá trị của  $x$  cho 4 quý tới bằng 13, 14, 15, 16 vào phương trình. Đây là dự báo phi mùa vụ trong 4 quý tới.

$$y_{41} = (16,865 \times 13) + 615,421 = 834,666$$

$$y_{42} = (16,865 \times 14) + 615,421 = 851,531$$

$$y_{43} = (16,865 \times 15) + 615,421 = 868,396$$

$$y_{44} = (16,865 \times 16) + 615,421 = 885,261$$

Tiếp theo, ta sử dụng chỉ số mùa vụ để mùa vụ hóa các số liệu.

Quý	Chỉ số mùa vụ (I)	Dự báo phi mùa vụ ( $Y_i$ )	Dự báo mùa vụ hóa ( $Y_{mv}$ )
1	0,809	834,666	675
2	1,122	851,531	955
3	1,251	868,396	1.086
4	0,818	885,261	724

### 1.5. Quy trình dự báo

Quy trình dự báo được chia thành chín bước. Các bước này bắt đầu và kết thúc với sự trao đổi (communication), hợp tác (cooperation) và cộng tác (collaboration) giữa những người sử dụng và những người làm dự báo

*Bước 1:* Xác định mục tiêu

- Các mục tiêu liên quan đến các quyết định cần đến dự báo phải được nói rõ. Nếu quyết định vẫn không thay đổi bất kể có dự báo hay không thì mọi nỗ lực thực hiện dự báo cũng vô ích.

- Nếu người sử dụng và người làm dự báo có cơ hội thảo luận các mục tiêu và kết quả dự báo sẽ được sử dụng như thế nào, thì kết quả dự báo sẽ có ý nghĩa quan trọng.

#### *Bước 2: Xác định dự báo cái gì*

- Khi các mục tiêu tổng quát đã rõ, ta phải xác định chính xác là dự báo cái gì (cần có sự trao đổi)

+ Ví dụ: Chỉ nói dự báo doanh số không thì chưa đủ, mà cần phải hỏi rõ hơn là: Dự báo doanh thu bán hàng (sales revenue) hay số đơn vị doanh số (unit sales). Dự báo theo năm, quý, tháng hay tuần.

+ Nên dự báo theo đơn vị để tránh những thay đổi của giá cả.

#### *Bước 3: Xác định khía cạnh thời gian*

Có hai loại khía cạnh thời gian cần xem xét:

- Thứ nhất: Độ dài dự báo, cần lưu ý:

+ Đối với dự báo theo năm: từ 1 đến 5 năm.

+ Đối với dự báo quý: từ 1 hoặc 2 năm.

+ Đối với dự báo tháng: từ 12 đến 18 tháng.

- Thứ hai: Người sử dụng và người làm dự báo phải thống nhất tính cấp thiết của dự báo.

#### *Bước 4: Xem xét dữ liệu*

- Dữ liệu cần để dự báo có thể từ hai nguồn:

+ Nguồn thông tin sơ cấp:

Thu thập qua các cuộc khảo sát, chọn mẫu hoặc các số liệu ghi chép các biến số trong doanh nghiệp.

Các phương pháp thu thập: phỏng vấn trực tiếp, gửi thư, điện thoại.

+ Nguồn thông tin thứ cấp:

Bên trong: nội bộ công ty, sổ sách kế toán.

Bên ngoài: sách báo, tạp chí, internet, các tài liệu thống kê,...

- Cần phải lưu ý dạng dữ liệu sẵn có (thời gian, đơn vị tính,...)

- Dữ liệu thường được tổng hợp theo cả biến và thời gian, nhưng tốt nhất là thu thập dữ liệu chưa được tổng hợp.

- Cần trao đổi giữa người sử dụng và người làm dự báo.

*Bước 5: Lựa chọn mô hình*

- Để chọn một phương pháp dự báo thích hợp người làm dự báo phải:

+ Xác định bản chất của vấn đề dự báo.

+ Bản chất của dữ liệu đang xem xét.

+ Mô tả các khả năng và hạn chế của các phương pháp dự báo tiềm năng.

+ Xây dựng các tiêu chí để ra quyết định lựa chọn.

+ Một nhân tố chính ảnh hưởng đến việc lựa chọn mô hình dự báo là nhận dạng và hiểu được bản chất số liệu lịch sử.

*Bước 6: Đánh giá mô hình*

- Đối với các phương pháp định tính thì bước này ít phù hợp hơn so với phương pháp định lượng.

- Đối với các phương pháp định lượng, cần phải đánh giá mức độ phù hợp của mô hình (trong phạm vi mẫu dữ liệu).

- Đánh giá mức độ chính xác của dự báo (ngoài phạm vi mẫu dữ liệu).

- Nếu mô hình không phù hợp, quay lại bước 5.

#### *Bước 7: Chuẩn bị dự báo*

- Nếu có thể nên sử dụng hơn một phương pháp dự báo, và nên là những loại phương pháp khác nhau (ví dụ mô hình hồi quy và san mũ Holt, thay vì cả 2 mô hình hồi quy khác nhau)

- Các phương pháp được chọn nên được sử dụng để chuẩn bị cho một số các dự báo (ví dụ trường hợp xấu nhất, tốt nhất và có thể nhất).

#### *Bước 8: Trình bày kết quả dự báo*

- Kết quả dự báo phải được trình bày rõ ràng cho ban quản lý sao cho họ hiểu các con số được tính toán như thế nào và chỉ ra sự tin cậy trong kết quả dự báo.

- Người dự báo phải có khả năng trao đổi các kết quả dự báo theo ngôn ngữ mà các nhà quản lý hiểu được.

- Trình bày cả ở dạng viết và dạng nói.

- Bảng biểu phải ngắn gọn, rõ ràng.

- Chỉ cần trình bày các quan sát và dự báo gần đây thôi.

- Chuỗi dữ liệu dài có thể được trình bày dưới dạng đồ thị (cả giá trị thực và dự báo).

- Trình bày thuyết trình nên theo cùng hình thức và cùng mức độ với phần trình bày viết.

#### *Bước 9: Theo dõi kết quả dự báo*

- Lệch giữa giá trị dự báo và giá trị thực phải được thảo luận một cách tích cực, khách quan và cởi mở.

- Mục tiêu của việc thảo luận là để hiểu tại sao có các sai số, để xác định độ lớn của sai số.

- Trao đổi và hợp tác giữa người sử dụng và người làm dự báo có vai trò rất quan trọng trong việc xây dựng và duy trì quy trình dự báo thành công.

*Một số khái niệm sử dụng trong phân tích dự báo:*

+ Xu thế: một thành phần dài hạn phản ánh xu hướng tăng hoặc giảm của chuỗi thời gian trong khoảng thời gian dài.

+ Chu kỳ: thành phần tăng, giảm có dạng sóng xung quanh đường xu thế.

+ Mùa vụ: thành phần thay đổi lặp đi lặp lại từ năm này sang năm khác.

+ Tính ngẫu nhiên: thành phần thay đổi bất thường

- Các phương pháp dự báo đối với dữ liệu dừng (dự báo thô, trung bình di động, ARIMA,...). Được sử dụng khi:

+ Môi trường của đối tượng dự báo không thay đổi.

+ Thiếu dữ liệu.

+ Thực hiện những điều chỉnh đơn giản có thể đạt được sự ổn định.

+ Chuỗi dữ liệu có thể chuyển đổi sang một dạng ổn định.

+ Dữ liệu là các sai số dự báo.

- Các phương pháp dự báo đối với dữ liệu xu thế (trung bình di động, san mũ Holt, hồi quy đơn, ARIMA). Được sử dụng khi:

+ Tăng năng suất hay công nghệ mới làm thay đổi lối sống.

+ Dân số tăng làm tăng nhu cầu hàng hóa, dịch vụ.



+ Các biến bị ảnh hưởng bởi lạm phát như: lương, chi phí sản xuất, sinh hoạt.

+ Mức độ chấp nhận của thị trường gia tăng.

- Các phương pháp dự báo đối với dữ liệu mùa vụ (sản phẩm mùa vụ, hồi quy bội, ARIMA,...). Được sử dụng khi:

+ Thời tiết ảnh hưởng đến biến đang xem xét.

+ Niên lịch ảnh hưởng đến biến đang xem xét.

- Các phương pháp được sử dụng với dữ liệu chu kỳ (hồi quy bội, ARIMA,...). Được sử dụng khi:

+ Chu kỳ kinh doanh bị ảnh hưởng bởi biến đang xem xét.

+ Chuyển biến trong sở thích chung như thời trang, âm nhạc,...

+ Chuyển biến trong dân số như đói, thiên tai,...

+ Chuyển dịch trong chu kỳ vòng đời sản phẩm.

## *Chương 2*

# **CÁC PHƯƠNG PHÁP PHÂN TÍCH VÀ DỰ BÁO**

Có nhiều phương pháp dự báo thống kê khác nhau (phương pháp lấy ý kiến chuyên gia, dự báo từng mức độ bình quân, ngoại suy hàm xu thế..., nhưng không phải phương pháp nào cũng được sử dụng phổ biến như nhau. Vì vậy, trong phần này chỉ trình bày một số phương pháp thông dụng nhất và giới thiệu một số phương pháp đang có xu hướng sử dụng nhiều trong thực tế hiện nay.

### **2.1. Dự báo từ các mức độ bình quân**

#### ***2.1.1. Dự báo từ số bình quân trượt (di động)***

Phương pháp số bình quân di động là một trong những phương pháp biểu hiện xu hướng phát triển cơ bản của hiện tượng nghiên cứu, hay nói cách khác, mô hình hoá sự phát triển thực tế của hiện tượng nghiên cứu dưới dạng dãy các số bình quân di động.

Phương pháp bình quân di động còn được sử dụng trong dự báo thống kê. Trên cơ sở xây dựng một dãy số bình quân di động, người ta xây dựng mô hình dự báo.

*Ví dụ:* Có dãy số thời gian về sản lượng thép của doanh nghiệp A trong 12 tháng theo bảng sau:

Thời gian	Sản lượng (triệu tấn) ( $y_i$ )	Doanh số trung bình di động (triệu tấn) ( $M_i$ )
1	79	-
2	82	-
3	85	82
4	82	83
5	88	85
6	86	85,3
7	98	90,6
8	105	96,3
9	110	104,3
10	115	110
11	120	115
12	118	117,6

Như vậy, ứng với tháng 3 ta có số bình quân di động là 82 triệu tấn, tháng 4 là 83 triệu tấn, v.v... và cuối cùng tháng 12 là 117,6 triệu tấn. Ta gọi các số bình quân di động mới này là  $M_i$  ( $i = k, k + 1, k + 2, \dots, n$ ), trong đó  $k$  là khoảng cách thời gian san bằng (ở đây  $k = 3$ , bình quân từ 3 mức độ thực tế).

Mô hình dự báo là:  $\hat{y}_{n+1} = M_n$

Khoảng dự báo sẽ được xác định theo công thức sau:

$$\hat{y}_{n+L} \pm t_{\alpha} S \sqrt{1 + \frac{1}{k}} \quad (2.1)$$

Trong đó:  $t_{\alpha}$  là giá trị tra trong bảng tiêu chuẩn t-Student với  $(k-1)$  bậc tự do và xác suất tin cậy  $(1-\alpha)$ . Độ lệch tiêu chuẩn mẫu điều chỉnh được tính theo công thức sau:

$$S = \sqrt{\frac{\sum (y_t - M_t)^2}{k - 1}} \quad (2.2)$$

Theo ví dụ trên ta tính được:

$$S = \sqrt{\frac{(85 - 82)^2 + (82 - 83)^2 + (88 - 85)^2 + (86 - 85,3)^2 + (98 - 90,6)^2 + (105 - 96,3)^2 + (110 - 104,3)^2 + (115 - 110)^2 + (120 - 115)^2 + (118 - 117,6)^2}{3 - 1}}$$

$$= 10,78$$

Trong ví dụ trên, dự đoán sản lượng thép cho tháng 1 năm sau là:

$$y_{13} = 117,6 \text{ triệu tấn}$$

Theo công thức trên ta tính được  $S = 10,78$  nghìn tấn và  $t_\alpha = 2,92$  với xác suất tin cậy  $(1-\alpha) = 0,95$  (xác suất đạt 95%) và số bậc tự do bằng 2. Do đó khoảng dự đoán về sản lượng thép tháng 1 năm sau sẽ nằm trong khoảng:

$$117,6 \pm (2,92 \times 10,78) \sqrt{1 \div \frac{1}{3}} = 117,6 \pm 36,35$$

### 2.1.2. Mô hình dự báo dựa vào lượng tăng (giảm) tuyệt đối bình quân

- Phương pháp này được sử dụng trong trường hợp lượng tăng (giảm) tuyệt đối liên hoàn xấp xỉ nhau qua các năm (dãy số thời gian có dạng gần giống như cấp số cộng):

$$\Delta_y = y_t - y_{y-1} \text{ xấp xỉ nhau } (I = z \div n).$$

Mô hình dự báo theo phương trình:

$$\hat{y}_{n+L} = y_n + \bar{\Delta}_y \cdot L \quad (2.3)$$

Trong đó:

$\hat{y}_{n+L}$ : Mức độ dự đoán ở thời gian  $(n+L)$ .

$y_n$ : Mức độ cuối cùng của dãy số thời gian.

$\bar{\Delta}_y$ : Lượng tăng, giảm tuyệt đối bình quân.

$L$ : Tầm xa của dự đoán ( $L=1,2,3,\dots$  năm).

Trong đó:

$$\bar{\Delta}_y = \frac{\sum(y_i - y_{i-1})}{n-1} \quad (i = \overline{2,n}) = \frac{y_n - y_1}{n-1}$$

Ví dụ: Giá trị sản xuất (GO). của một doanh nghiệp A qua các năm như sau:

Thời gian Chỉ tiêu	2001	2002	2003	2004	2005	2006
Giá trị sản xuất (GO) (tỷ đồng)	32	36	39	41	43	45

Ta có:  $\bar{\Delta}_y = \frac{45 - 32}{6 - 1} = 2,6$  tỷ

Dự báo GO của doanh nghiệp cho năm 2007 ( $L=1$ ). Ta có phương trình:

$$\hat{y}_{2006+1} = y_{2006} + 2,6 \times 1$$

$$\hat{y}_{2007} = 45 + 2,6 = 47,6 \text{ (tỷ)}$$

Dự báo GO của doanh nghiệp năm 2008

$$\hat{y}_{2008} = 45 + 2,6 \times 2 = 50,2 (\text{tỷ})$$

Tương tự, dự báo cho GO năm 2011 (tầm xa của dự báo là 5)

$$\hat{y}_{2011} = 45 + 2,6 \times 5 = 58 (\text{tỷ})$$

### 2.1.3. Mô hình dự báo dựa vào tốc độ phát triển bình quân

Thường áp dụng trong trường hợp các mức độ của dãy số biến động theo thời gian có tốc độ phát triển (hoặc tốc độ tăng, giảm) từng kỳ gần nhau (dãy số thời gian có dạng gần như cấp số nhân).

Có hai mô hình dự đoán:

\* *Dự đoán mức độ hàng năm (có thể dùng để dự báo trong dài hạn)*

- Phương pháp này được áp dụng khi tốc độ phát triển hoàn toàn xấp xỉ nhau.

- Mô hình dự đoán:

$$\hat{y}_{n+L} = y_n \cdot \bar{t}^{-(L)} \quad (2.4)$$

$\hat{y}_{n+L}$ : Mức độ dự đoán ở thời gian  $(n+L)$ .

$y_n$ : Mức độ được dùng làm kỳ gốc để ngoại suy.

$L$ : Tầm xa của dự đoán ( $L=1, 2, 3, \dots$  năm).

$\bar{t}$ : Tốc độ phát triển bình quân hàng năm.

$$\bar{t} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Với ví dụ trên ta có:

$$\bar{t} = \sqrt[n-1]{\frac{45}{32}} \approx 1,071$$

Dự đoán cho năm 2007 (ta chọn năm gốc là năm cuối cùng trong dãy số - 2006):

Theo công thức trên, GO của doanh nghiệp là:

$$\text{Năm 2007: } \hat{y}_{2007} = 45 \times (1,071)^1 = 48,18 \text{ (tỷ)}$$

$$\text{Năm 2008: } \hat{y}_{2008} = 45 \times (1,071)^2 = 51,5 \text{ (tỷ)}$$

Tương tự, GO của năm 2011 là:  $\hat{Y}_{2011} = 45 \times (1,071)^5 = 63,4 \text{ (tỷ)}$

*\* Dự đoán mức độ của khoảng thời gian dưới 1 năm (quý, tháng-dự báo ngắn hạn)*

$$\hat{y}_{ij} = y_j \frac{(\bar{t})^{i-1}}{S_t} \quad (2.5)$$

Trong đó:

$\hat{y}_{ij}$ : Mức độ của hiện tượng ở thời gian  $j$  ( $j=1, m$ ) của năm  $i$

$$y_j = \sum_{i=1}^n y_{ij} : \text{Tổng các mức độ của thời gian } j \text{ của năm } i$$

( $i = 1 \dots n$ ).

$$\bar{t} = \sqrt[n-1]{\frac{y_n}{y_1}}; \text{ Tốc độ phát triển bình quân hàng tháng.}$$

$$S_r = 1 + (\bar{t}) + (\bar{t})^2 + (\bar{t})^3 + \dots + (\bar{t})^{n-1}$$

$n$ : có thể là số năm hoặc số lượng mức độ của từng năm.

*Ví dụ:* Có tài liệu về tình hình sản xuất một loại sản phẩm của xí nghiệp A như sau:

<div>Quý j Năm j</div>	I	II	III	IV	$y_i$
2004	20	21,5	22	23,5	87
2005	20,04	20,63	21,7	24,28	86,65
2006	21,04	22,83	23,5	25,63	93
$y_j$	61,08	64,96	67,2	73,41	266,65

Từ bảng số liệu trên ta có:

$$\bar{t} = \sqrt[4-1]{\frac{93}{87}} = 1,034$$

$$S_r = 1 + (\bar{t}) + (\bar{t})^2 = 1 + 1,034 + (1,034)^2 = 3,103$$

- Dự đoán sản lượng cho các quý của năm 2007 ( $i = 4$ )

$$\frac{(\bar{t})^{4-1}}{S_r} = \frac{1,034^3}{3,103} = 0,356$$

Tỷ lệ này dùng để điều tiết các khoảng thời gian của năm.



$$\hat{y}_{4,I} = y_I \frac{(t)^3}{S_r} = 61,08 \times 0,356 = 21,744 \text{ (nghìn tấn)}$$

$$\hat{y}_{4,II} = y_{II} \frac{(t)^3}{S_r} = 64,96 \times 0,356 = 23,126 \text{ (nghìn tấn)}$$

$$\hat{y}_{4,III} = y_{III} \frac{(t)^3}{S_r} = 67,2 \times 0,356 = 23,923 \text{ (nghìn tấn)}$$

$$\hat{y}_{4,IV} = y_{IV} \frac{(t)^3}{S_r} = 73,41 \times 0,356 = 26,134 \text{ (nghìn tấn)}$$

## 2.2. Mô hình dự báo theo phương trình hồi quy (dự báo dựa vào xu thế)

Từ xu hướng phát triển của hiện tượng nghiên cứu ta xác định được phương trình hồi quy lý thuyết, đó là phương trình phù hợp với xu hướng và đặc điểm biến động của hiện tượng nghiên cứu, từ đó có thể ngoại suy hàm xu thế để xác định mức độ phát triển trong tương lai.

### 2.2.1. Mô hình hồi quy theo thời gian

Mô hình dự báo theo phương trình hồi quy đường thẳng:

$$\hat{y} = a + bt$$

Trong đó: a, b là những tham số quy định vị trí của đường hồi quy.

Từ phương trình này, bằng phương pháp bình phương nhỏ nhất hoặc thông qua việc đặt thứ tự thời gian (t) trong dãy số để tính các tham số a, b.

Đặt thứ tự thời gian t sao cho  $\sum t = 0$ , ta có các công thức tính tham số sau:

$$a = \frac{\sum y}{n} = \bar{y}$$

$$b = \frac{\sum y.t}{\sum t^2}$$

Ví dụ: Hãy dự báo về doanh thu tiêu thụ của cửa hàng thương mại B trong những năm tiếp theo trên cơ sở bảng số liệu sau:

<div>Thời gian</div> <div>Chỉ tiêu</div>	2001	2002	2003	2004	2005	2006
Doanh thu tiêu thụ (tỷ đồng)	70	98	115	120	136	180

Từ nguồn tài liệu, ta có bảng số liệu sau (đặt thứ tự thời gian sao cho  $(\sum t = 0)$ )

Năm (n)	Doanh thu (tỷ đồng) $y_i$	Điều kiện đặt $= 0$			
		t	$t^2$	y.t	$\hat{y}$
2001	70	-5	25	-350	72,045
2002	98	-3	9	-294	91,159
2003	115	-1	1	-115	120,27
2004	120	1	1	120	129,387
2005	136	3	9	408	148,58
2006	180	5	25	900	167,61
N = 6	719	=0	= 70	= 669	

Tính các tham số a và b theo điều kiện đặt  $\sum t = 0$ :

$$a = \frac{\sum y}{n} = \bar{y} = \frac{719}{6} = 119,83$$

$$b = \frac{\sum y \cdot t}{\sum t^2} = \frac{669}{70} = 9,557$$

Hàm xu thế có dạng:  $\hat{y} = 119,83 + 9,55t$

Từ hàm xu thế này ta có thể dự báo doanh thu của cửa hàng B trong những năm tiếp theo như sau:

Doanh thu của năm 2007 ( $t=7$ ):

$$\hat{y}_{2007} = 119,83 + 9,557 \times 7 = 186,729$$

Doanh thu của năm 2008 ( $t=9$ ):

$$\hat{y}_{2008} = 119,83 + 9,557 \times 9 = 205,843$$

Doanh thu của năm 2009 ( $t=11$ ):

$$\hat{y}_{2009} = 119,83 + 9,557 \times 11 = 224,957$$

Doanh thu của năm 2010 ( $t=13$ ):

$$\hat{y}_{2010} = 119,83 + 9,557 \times 13 = 244,071$$

Số liệu dự báo ( $\hat{Y}$ ) và số liệu thực tế  $y_i$  có sự chênh lệch là do có sai số trong dự đoán.

+ Sai số dự báo là sự chênh lệch giữa mức độ thực tế và mức độ tính toán theo mô hình dự báo.

+ Sai số dự báo phụ thuộc vào ba yếu tố: độ biến thiên của tiêu thức trong thời kỳ trước, độ dài của thời gian của thời kỳ trước và độ dài của thời kỳ dự đoán.

+ Vấn đề quan trọng nhất trong dự báo bằng ngoại suy hàm xu thế là lựa chọn hàm xu thế, xác định sai số dự đoán và khoảng dự đoán:

- Công thức tính sai số chuẩn ( $\delta_y$ )

$$\delta_y = \sqrt{\frac{\sum (y_i - \hat{y})^2}{n - p}}$$

Trong đó:

$\hat{y}$ : Giá trị tính toán theo hàm xu thế.

$n$ : Số các mức độ trong dãy số.

$p$ : Số các tham số cần tìm trong mô hình xu thế.

$n - p$ : Số bậc tự do.

Công thức này được dùng để lựa chọn dạng hàm xu thế (so sánh các sai số chuẩn tính được), sai số nào nhỏ nhất chứng tỏ rằng hàm tương ứng với sai số sẽ xấp xỉ tốt nhất và được lựa chọn làm hàm xu thế để dự đoán. Thông thường, để việc dự đoán được tiến hành đơn giản ta vẫn chọn hàm xu thế làm hàm tuyến tính.

Công thức tính sai số dự báo:

$$\tilde{s}_p = \delta_y \sqrt{1 + \frac{1}{n} + \frac{3(n + 2^L - 1)^2}{n(n^2 - 1)}}$$

Trong đó:

n: Số lượng các mức độ.

L: Tầm xa của dự báo.

Sau đó xác định khoảng dự đoán theo công thức sau:

$$\hat{y}_{n+L} \pm t_{\alpha} \hat{S}_p$$

$t_{\alpha}$  : Giá trị theo bảng của tiêu chuẩn t - Student với (n-2) bậc tự do và xác suất tin cậy  $(1-\alpha)$ .

Trở lại ví dụ trên ta đi tính  $\delta_y$ ,

$$\delta_y = \sqrt{\frac{(70-72,045)^2 + (98-91,159)^2 + (115-110,27)^2 + (120-129,387)^2 + (136-148)^2 + (180-167,61)^2}{6-2}} = 10,876$$

Sai số dự báo:

+ Đối với năm 2007 (L=1):

$$\hat{S}_{p2007} = 10,876 \sqrt{1 + \frac{1}{6} + \frac{3(6+2^1-1)^2}{6(6^2-1)}} = 14,856$$

+ Đối với năm 2008 (L=2):

$$\hat{S}_{p2008} = 10,876 \sqrt{1 + \frac{1}{6} + \frac{3(6+2^2-1)^2}{6(6^2-1)}} = 16,93$$

Với xác suất tin cậy là 0,95 và số bậc tự do  $n=4$ , khi đó  $t_{\alpha}=2,132$

Ta có dự báo của năm 2007 là  $186,729 \pm 2,132 \times 14,856 = 186,729 \pm 31,67$

Ta có dự báo của năm 2008 là  $205,843 \pm 2,132 \times 16,93 = 205,843 \pm 36 \dots$

Như vậy, ta đã chuyển từ dự báo điểm sang dự báo khoảng.

### 2.2.2. Mô hình hồi quy giữa các tiêu thức

Mô hình hồi quy tuyến tính giữa hai tiêu thức

Từ việc xây dựng phương trình hồi quy tuyến tính giữa các tiêu thức đã nêu ở phần trên, ta có thể dự đoán các giá trị của Y trong tương lai khi các biến trong hàm hồi quy thay đổi, cụ thể:

Đối với phương trình tuyến tính giản đơn:  $y_x = a + bx$

Trong đó: a, b là những tham số quy định vị trí của đường hồi quy. Hằng số a là điểm cắt trục tung (biểu hiện của tiêu thức kết quả) khi tiêu thức nguyên nhân x bằng 0. Độ dốc b chính là lượng tăng giảm của tiêu thức kết quả khi tiêu thức nguyên nhân thay đổi.

Từ phương trình này, ta sẽ dự đoán được giá trị của tiêu thức kết quả trong tương lai khi có sự thay đổi của tiêu thức nguyên nhân.

Tương tự như trong hồi quy giản đơn, trong hồi quy bội, giá trị dự đoán của y có được tương ứng với các giá trị cho trước của k biến x bằng cách thay các giá trị của k biến x vào phương trình hồi quy bội.

Các giá trị cho trước của biến X lần lượt là  $x_{1,n+1}, x_{2,n+1}, \dots, x_{k,n+1}$ , thì giá trị dự đoán  $y_{n+1}$  sẽ là:

$$y_{n+1} = a + b_1 x_{1,n+1} + b_2 x_{2,n+1} + \dots + b_k x_{k,n+1}$$

### 2.3. Dự báo dựa vào hàm xu thế và biến động thời vụ

Phương pháp dự báo này áp dụng đối với hiện tượng nghiên cứu chịu tác động của nhiều nhân tố biến động. Như biến động thời vụ, biến động xu hướng và biến động bất thường.

- Mô hình dự báo sẽ có thể dựa vào hàm xu thế kết hợp với biến động thời vụ:

$$y_t = \hat{y} + tv + bt \quad (2.7)$$

- Hoặc dự báo dựa vào hàm xu thế kết hợp nhân tố với biến động thời vụ:

$$y_t = \hat{y} \times tv \times bt \quad (2.8)$$

Trong đó:

$\hat{y}$ : Mức độ lý thuyết xác định từ hàm xu thế (hoặc các phương pháp nêu trên)

$tv$ : Ảnh hưởng của nhân tố thời vụ.

$bt$ : Ảnh hưởng của nhân tố bất thường.

Nhìn chung, hàm xu thế, chỉ số thời vụ được xác định từng mô hình còn những nhân tố biến động bất thường thường không dự báo được, do vậy mô hình chỉ còn lại hai nhân tố: biến động xu hướng và biến động thời vụ.

#### 2.3.1. Dự báo dựa vào mô hình cộng

*Vi dụ:* Có tài liệu về sản lượng của doanh nghiệp A như sau:

Năm	Sản lượng (nghìn tấn)	Công	Mức	Chỉ số
-----	-----------------------	------	-----	--------

<div style="display: inline-block; transform: rotate(-45deg);"> (t) Quý </div>	2002	2003	2004	2005	2006	theo cùng quý $\sum y_j$	độ bình quân từng quý $\bar{y}_j$	thời vụ $I_{tv} = \frac{\bar{y}_j}{\bar{y}_i}$
I	20	25	27	31	29	132	26,4	0,678
II	25	32	30	37	36	160	32	0,82
III	38	38	45	44	47	212	42,4	1,14
IV	40	60	55	62	58	275	55	1,41
Cộng theo cùng năm $\sum y_i$	123	155	157	174	170	779	38,95	
$\sum t.y$	123	310	471	696	850			

\* Trước tiên xác định hàm xu thế tuyến tính sản lượng doanh nghiệp có dạng là:

$$\hat{y} = a + bt$$

Trong đó: a, b là các tham số quy định vị trí của hàm xu thế tuyến tính, được tính theo công thức sau:

$$b = \frac{\frac{\sum t.y}{m} - \frac{n+1}{2m}}{\sum y_j} = \frac{\frac{2450}{4} - \frac{5+1}{2 \times 4}}{779} = 0,785$$

$$a = \frac{\sum y_j}{m.n} - b \frac{m.n+1}{2} = \frac{779}{4 \times 5} - 0,785 \frac{4 \times 5 + 1}{2} = 30,708$$

Trong đó:



n: Số năm.

m: Khoảng cách thời gian trong một năm ( $m = 4$  đối với quý,  $m=12$  đối với tháng).

t: Thứ tự thời gian trong dãy số (năm).

Do vậy, hàm xu thế có dạng:  $\hat{y} = 30,708 + 0,785t$

Mức độ bình quân một quý tính chung cho 5 năm:  $\overline{y_i} = 38,95$

\* Tính các mức độ mang tính thời vụ theo công thức sau:

$$tv = \bar{y}_i - \bar{y}_j - b \left( i - \frac{m+1}{2} \right) \text{ với } i = 1, 2, 3, 4 \dots$$

Do vậy, mức độ dự báo về thời vụ cho các quý của năm 2007 như sau:

$$\text{- Quý I: } (26,4 - 38,95) - 0,785 \cdot \left( 1 - \frac{4+1}{2} \right) = -11,373$$

$$\text{- Quý II: } (32 - 38,95) - 0,785 \cdot \left( 2 - \frac{4+1}{2} \right) = -6,557$$

$$\text{- Quý III: } (42,4 - 38,95) - 0,785 \cdot \left( 3 - \frac{4+1}{2} \right) = 3,058$$

$$\text{- Quý IV: } (55 - 38,95) - 0,785 \cdot \left( 4 - \frac{4+1}{2} \right) = 14,873$$

Sau khi xác định xong hàm xu thế và biến động thời vụ thì mô hình dự báo kết hợp cộng giữa xu thế biến động và tính thời vụ có dạng:  $y_t = \hat{y} + tv$

Dự báo sản lượng quý I năm 2007 ( $t = 21$ )

$$y_1 = 30,708 + 0,785 \times 21 - 11,373 = 35,82$$

$$\text{Quý II (t = 22): } y_2 = 30,708 + 0,785 \times 22 - 6,557 = 41,421$$

Cứ tiếp tục như vậy cho đến các quý tiếp theo.

### 2.3.2. Dự báo dựa vào mô hình nhân

Mô hình dự báo theo kết hợp nhân có dạng:

$$y_t = \hat{y}_t \times tv_t \quad (2.9)$$

Để dự báo theo mô hình này, trước hết phải tính được hàm xu thế, hàm xu thế trong trường hợp này phải được loại trừ biến động thời vụ bằng cách xây dựng dãy số bình quân trượt ( $\overline{y_t}$ ) với số lượng mức độ bằng 4 với tài liệu quý và 12 với tài liệu tháng.

Từ đó ta tính được  $\frac{y_t}{\overline{y_t}}$ , tiếp theo xác định thành phần thời vụ

( $tv_t$ ) bằng cách tính các số bình quân  $\overline{tv_t}$ , sau đó tính hệ số điều chỉnh H:

$$H = \frac{m}{\sum \overline{tv_t}} \quad (\text{với } m = 4 \text{ đối với tài liệu quý, } 12 \text{ đối với tài liệu tháng}).$$

Từ đó tính chỉ số thời vụ  $I_{tv} = \overline{tv_t} \times H$

Sau khi xác định được  $tv_t$  thì xác định dãy số  $f_t$  là dãy số đã loại bỏ thành phần thời vụ như sau:  $f_t = \frac{y_t}{tv_t}$

Theo ví dụ trên ta có thể lập bảng tính toán sau đây:

STT	$y_t$	$\overline{y_t}$	$\frac{y_t}{\overline{y_t}}$	$tv_t$	$f_t = \frac{y_t}{tv_t}$
-----	-------	------------------	------------------------------	--------	--------------------------

STT	$y_t$	$\bar{y}_t$	$\frac{y_t}{\bar{y}_t}$	$tv_t$	$f_t = \frac{y_t}{tv_t}$
1	20	-	-	0,7	28,57
2	25	-	-	0,838	29,83
3	38	30,75	1,236	1,08	35,19
4	40	32	1,25	1,376	29,07
5	25	33,75	0,74	0,7	35,71
6	32	33,75	0,948	0,838	38,19
7	38	38,75	0,98	1,08	35,19
8	60	39,25	1,529	1,376	43,6
9	27	38,75	0,697	0,7	38,57
10	30	40,5	0,74	0,838	35,8
11	45	39,25	1,146	1,08	41,67
12	55	40,25	1,366	1,376	39,97
13	31	42	0,738	0,7	44,28
14	37	41,75	0,866	0,838	44,15
15	44	43,5	1,011	1,08	40,74
16	62	43	1,441	1,376	45,06
17	29	42,75	0,678	0,7	41,13
18	36	43,5	0,827	0,838	42,96
19	47	42,5	1,105	1,08	43,5
20	58	-	-	1,376	42,15

Từ  $f_t$  ta lập bảng sau:

Quý Năm	I	II	III	IV
2002	-	-	1,236	1,25

2003	0,74	0,948	0,98	1,529
2004	0,697	0,74	1,146	1,366
2005	0,738	0,886	1,011	1,441
2006	0,678	0,827	1,105	-
Bình quân quý ( $\overline{tv_t}$ )	0,713	0,85	1,096	1,396

Với tài liệu trong bảng tính ta tính được các đại lượng trên như sau:

$$H = \frac{m}{\sum tv_t} = \frac{4}{0,713 + 0,85 + 1,096 + 1,396} = 0,986$$

\* Trước tiên, tính các chỉ số thời vụ:  $I_{tv} = \overline{tv_t} . H$

$$\text{Quý I} = 0,713 \times 0,986 = 0,7$$

$$\text{Quý II} = 0,85 \times 0,986 = 0,838$$

$$\text{Quý III} = 1,096 \times 0,986 = 1,08$$

$$\text{Quý IV} = 1,396 \times 0,986 = 1,376$$

\* Xây dựng hàm xu thế:  $\hat{y} = a + bt$

Để tiện theo dõi, từ ( $f_t$ ) ta lập bảng sau:

Quý Năm	I	II	III	IV	Tổng năm (n)	t.y
2002	28,57	29,83	35,19	29,07	122,66	122,66
2003	35,71	38,19	35,19	43,6	152,69	305,38
2004	38,57	35,8	41,67	39,97	156,01	468,03

2005	44,28	44,15	40,74	45,06	174,23	696,92
2006	41,43	42,96	43,5	42,15	170,04	850,2
Tổng quý (Q)	188,56	190,93	196,29	199,85	775,63	$\sum t.y = 2443,19$
Bình quân quý	37,71	38,186	39,26	39,97		

Các tham số của hàm xu thế được tính như sau:

$$b = \frac{\frac{\sum t.y}{m} - \frac{n+1}{2m}}{\sum y_i} = \frac{\frac{2443,19}{4} - \frac{5+1}{2 \times 4}}{775,63} = 0,787$$

$$a = \frac{\sum y_i}{m.n} - b \frac{m.n+1}{2} = \frac{775,63}{4 \times 5} - 0,787 \frac{4 \times 5 + 1}{2} = 30,518$$

Hàm xu thế có dạng:  $\hat{y} = 30,518 + 0,787t$

Do đó, mô hình nhân có dạng:  $y_t = (30,518 + 0,787t) \cdot I_{tv}$

Dự báo sản lượng của doanh nghiệp năm 2007 theo các quý là:

- Quý I (t=21):  $y_{t1} = (30,518 + 0,787.21) \times 0,7 = 32,932$
- Quý II (t=22):  $y_{t2} = (30,518 + 0,787t \times 22) \times 0,838 = 40,083$
- Quý III (t=23):  $y_{t3} = (30,518 + 0,787t \times 23) \times 1,08 = 52,509$
- Quý IV (t=24):  $y_{t4} = (30,518 + 0,787t \times 24) \times 1,376 = 67,983$

Với hàm kết hợp nhân ta có thể dự báo cho những năm tiếp theo

## **2.4. Dự báo theo phương pháp san bằng mũ**

Phương pháp san bằng mũ (hay còn gọi là phương pháp dự đoán bình quân mũ) là một phương pháp dự đoán thống kê ngắn hạn hiện được sử dụng nhiều trong công tác dự đoán thực tế trên thế giới.

Nếu như một số phương pháp dự đoán thống kê đã đề cập ở trên coi giá trị thông tin của các mức độ trong dãy số thời gian là như nhau, phương pháp san bằng mũ lại coi giá trị thông tin của mỗi mức độ là tăng dần kể từ đầu dãy số cho đến cuối dãy số. Vì trên thực tế, ở những thời gian khác nhau thì hiện tượng nghiên cứu chịu sự tác động của những nhân tố khác nhau và cường độ không giống nhau. Các mức độ ngày càng mới (ở cuối dãy số thời gian) càng cần phải được chú ý đến nhiều hơn so với các mức độ cũ (ở đầu dãy số). Hay nói cách khác, mức độ càng xa so với thời điểm hiện tại thì càng ít giá trị thông tin, do đó càng ít ảnh hưởng đến mức độ dự đoán.

Tuỳ thuộc vào đặc điểm dãy số thời gian (chuỗi thời gian) có biến động xu thế, biến động thời vụ hay không mà phương pháp san bằng mũ có thể sử dụng một trong các phương pháp cơ bản sau:

### **2.4.1. Mô hình đơn giản (phương pháp san bằng mũ đơn giản)**

Điều kiện áp dụng: Đối với dãy số thời gian không có xu thế và không có biến động thời vụ rõ rệt.

Trước hết, dãy số thời gian được san bằng nhờ có sự tham gia của các số bình quân mũ, tức là các số bình quân di động gia quyền theo quy luật hàm số mũ. Theo phương pháp này, ở thời gian  $t$  nào đó dựa vào các giá trị thực tế đã biết để ước lượng giá trị hiện tại (thời gian  $t$ ) của hiện tượng và giá trị hiện tại này để dự toán giá trị tương

lai (thời gian  $t+1$ ). Mô hình san bằng mũ giản đơn được Brown xây dựng năm 1954 dựa trên hai nguyên tắc:

- Trọng số của các quan sát trong dãy số thời gian càng giảm đi khi nó càng cách xa hiện tại.

Sai số dự báo hiện tại (ký hiệu  $e_t = y_t - \hat{y}_t$ ) Phải được tính đến trong những dự báo kế tiếp.

Giả sử ở thời gian  $t$ , có mức độ thực tế là  $y_t$ , mức độ dự đoán là  $\hat{y}_t$ .

Mức độ dự đoán của hiện tượng ở thời gian  $(t+1)$  có thể viết:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \quad (2.10)$$

$$\text{Đặt } 1 - \alpha = \lambda, \text{ ta có: } \hat{y}_{t+1} = \alpha y_t + \lambda \hat{y}_t \quad (2.11)$$

$\alpha$  và  $\lambda$  được gọi là các tham số san bằng với  $\alpha + \lambda = 1$  và  $\alpha, \lambda \in [0;1]$ .

Như vậy, mức độ dự đoán  $\hat{y}_{t+1}$  là trung bình cộng gia quyền của  $y_t$  và  $\hat{y}_t$  với quyền số tương ứng là  $\alpha$  và  $\lambda$

- Mức độ dự đoán của hiện tượng ở thời gian  $t$  là:

$$\hat{y}_t = \alpha y_{t-1} + \lambda \hat{y}_{t-1} \text{ thay vào (2.11) ta có:}$$

$$\hat{y}_{t+1} = \alpha y_t + \lambda \alpha y_{t-1} + \lambda^2 \hat{y}_{t-1} \quad (2.12)$$

- Mức độ dự đoán của hiện tượng ở thời gian  $(t-1)$  là:

$$\hat{y}_{t-1} = \alpha y_{t-2} + \lambda \hat{y}_{t-2} \text{ thay vào (2.12)}$$

Ta có:  $\hat{y}_{t+1} = \alpha y_t + \lambda \alpha y_{t-1} + \lambda^2 \alpha y_{t-2} + \lambda^3 \hat{y}_{t-2}$  (2.13)

- Mức độ dự đoán của hiện tượng ở thời gian (t-2) là:

$\hat{y}_{t-2} = \alpha y_{t-3} + \lambda \hat{y}_{t-3}$  thay vào (2.13)

Ta có:  $\hat{y}_{t+1} = \alpha y_t + \lambda \alpha y_{t-1} + \lambda^2 \alpha y_{t-2} + \lambda^3 \hat{y}_{t-3} + \lambda^4 \hat{y}_{t-4}$  (2.14)

Bằng cách tiếp tục tương tự thay vào các mức độ dự đoán  $y_{t-3}, y_{t-4}, \dots$  ta sẽ có công thức tổng quát.

$$\hat{y}_{t+1} = \alpha \sum_{i=1}^n \lambda^i y_{t-i} + \lambda^{i+1} \hat{y}_{t-i} \quad (*)$$

Trong đó:

$\hat{y}_{t+1}$ : Số bình quân mũ tại thời điểm t+1.

$y_{t-i}$ : Các mức độ thực tế của hiện tượng tại thời điểm

(t-i) ( $i=0 \rightarrow n$ )

$\hat{y}_{t-i}$ : Số bình quân mũ tại thời điểm (t-i) ( $i=0 \rightarrow n$ ).

$\alpha$  và  $\lambda$  được gọi là các tham số san bằng.

( $\alpha$  và  $\lambda$  là hằng số với  $\alpha + \lambda = 1$  và  $\alpha, \lambda \in [0; 1]$ )

n: Số lượng các mức độ của dãy số thời gian.

Vì  $\lambda \in [0; 1]$  nên khi  $i \rightarrow \infty$

thì 
$$\begin{cases} \lambda^{i+1} \rightarrow 0 \Rightarrow \lambda^{i+1} y_{t-i} \rightarrow 0 \\ \alpha \sum_{i=1}^{\infty} \lambda^i \rightarrow 1 \end{cases}$$



Khi đó công thức (\*) trở thành:

$$\hat{y}_{t+1} = \alpha \sum_{i=1}^n \lambda^i y_{t-i}$$

Như vậy: mức độ dự đoán  $\hat{y}_{t+1}$  là trung bình cộng gia quyền của các mức độ của dãy số thời gian mà trong đó quyền số giảm dần theo dạng mũ (khi  $i=0 \rightarrow n$ ) tùy thuộc vào mức độ cũ của dãy số. Vì thế, phương pháp này được gọi là phương pháp san bằng mũ.

Có hai vấn đề quan trọng nhất trong phương pháp san bằng mũ.

- *Thứ nhất*: hệ số san bằng mũ  $\alpha$ .

$\alpha$  là hệ số san bằng để điều chỉnh trọng số của các quan sát riêng biệt của dãy số thời gian. Vì vậy, khi lựa chọn  $\alpha$  phải vừa đảm bảo kết quả dự báo sẽ gần với quan sát thực tế, vừa phải đảm bảo tính linh hoạt (nhạy nhạy với các thay đổi ở gần hiện tại).

Với  $\alpha=1$  thì theo phương trình dự báo (1). Giá trị dự báo  $\hat{y}_{t+1}$  bằng giá trị thực tế ở thời kỳ ngay liền trước ( $y_{t+1}$ ) và các mức độ trước đó không được tính đến.

Với  $\alpha=0$  theo phương trình dự báo (1). Giá trị dự báo  $\hat{y}_{t+1}$  bằng giá trị dự báo ở thời kỳ trước ( $\hat{y}_t$ ) và giá trị thực tế ở thời kỳ ngay liền trước không được tính đến.

Nếu  $\alpha$  được chọn càng lớn thì các mức độ càng mới sẽ càng được chú ý, thích hợp với chuỗi thời gian không có tính ổn định cao.

Ngược lại, nếu  $\alpha$  được chọn càng nhỏ thì các mức độ càng cũ sẽ càng được chú ý, thích hợp với chuỗi thời gian có tính ổn định cao.

Do đó, phải dựa vào đặc điểm biến động của hiện tượng qua thời gian và kinh nghiệm nghiên cứu để lựa chọn  $\alpha$  cho phù hợp. Nói chung, giá trị  $\alpha$  tốt nhất là giá trị làm cho tổng bình phương sai số dự đoán nhỏ nhất.

$$SSE = \sum (y_t - \hat{y}^*)^2 \rightarrow \min$$

Đặt  $e_t = y_t - \hat{y}_t$  là các sai số dự đoán ở thời gian  $t$  hay còn gọi là phần dư ở thời gian  $t$ .

Theo kinh nghiệm của các nhà dự báo thì  $\alpha$  thích hợp cho vận dụng phương pháp san bằng mũ có thể được chọn bằng:

$$\alpha = \frac{2}{n+1} \quad (n: \text{độ dài chuỗi thời gian})$$

- *Thứ hai:* Xác định giá trị ban đầu (điều kiện ban đầu) ký hiệu  $y_0$

Phương pháp san bằng mũ được thực hiện theo phép đệ quy, để tính  $\hat{y}_{t+1}$  thì phải có  $\hat{y}_t$ , để có  $\hat{y}_t$  thì phải có  $\hat{y}_{t-1}$ . Do đó, để tính toán cần phải phải xác định giá trị ban đầu ( $y_0$ ) dựa vào một số phương pháp.

+ Có thể lấy mức độ đầu tiên của dãy số.

+ Trung bình của một số các mức độ của dãy số.

*Ví dụ:* Có hai tài liệu về doanh thu ở một cửa hàng thương mại X qua một số năm như sau:

<div>Năm</div> <div>Chỉ tiêu</div>	2002	2003	2004	2005	2006
Doanh thu (tỷ đồng)	15 $y_1$	15,3 $y_2$	14,8 $y_3$	15,5 $y_4$	15,2 $y_5$

Yêu cầu: Dự đoán doanh thu cho năm 2007 của cửa hàng.

$$\text{Với } n = 5 \rightarrow \alpha = \frac{2}{n+1} = \frac{2}{5+1} \approx 0,3$$

$$y_0 = \frac{1}{5} \sum_{i=1}^5 y_i = \frac{15 + 15,3 + 15,8 + 15,5 + 15,2}{5} = 15,16 \text{ (tỷ đồng)}$$

Công thức tổng quát với  $n=5$ :

$$l = 0 \rightarrow 5: \hat{y}_{t+1} = \alpha \sum_{i=1}^5 (1-\alpha)^i y_{t-i} + (1-\alpha)^{i+1} \hat{y}_{t-i}$$

$$\lambda = 1 - \alpha \Rightarrow$$

$$\hat{y}_{t+1} = \alpha(y_t + \lambda y_{t-1} + \lambda^2 y_{t-2} + \lambda^3 y_{t-3} + \lambda^4 y_{t-4} + \lambda^5 y_{t-5}) + \lambda^6 \hat{y}_{t-5}$$

Với  $t = 5$ , dự báo doanh thu 2007 là:

$$\hat{y}_6 = \alpha(y_5 + \lambda y_4 + \lambda^2 y_3 + \lambda^3 y_2 + \lambda^4 y_1 + \lambda^5 y_0) + \lambda^6 \hat{y}_0$$

$$\hat{y}_6 = 0,3(15,2 + 0,7 \times 15,5 + 0,7^2 \times 14,8 + 0,7^3 \times 15,3 + 0,7^4 \times 15 + 0,7^5 \times 15,16) + 0,7^6 \times 15,16 = 15,19$$

\* Hoặc thay vào công thức (1) ta có thể dự báo doanh thu hàng năm (tỷ đồng) như sau:

Với  $t=0$ , ta có:

$$\hat{y}_1 = \alpha y_0 + (1-\alpha) y_0 = 0,3 \times 15,16 + (1-0,3) \times 15,16 = 15,16$$

Với  $t=1$ , ta có:

$$\hat{y}_2 = \alpha y_1 + (1-\alpha) \hat{y}_1 = 0,3 \times 15 + (1-0,3) \times 15,16 = 15,112$$

Với  $t=2$ , ta có:

$$\hat{y}_3 = \alpha y_2 + (1 - \alpha) \hat{y}_2 = 0,3 \times 15,3 + (1-0,3) \times 15,112 = 15,1684$$

Với  $t=3$  ta có:

$$\hat{y}_4 = \alpha y_3 + (1 - \alpha) \hat{y}_3 = 0,3 \times 14,8 + (1-0,3) \times 15,1684 = 15,05788$$

Với  $t=4$  ta có:

$$\hat{y}_5 = \alpha y_4 + (1 - \alpha) \hat{y}_4 = 0,3 \times 15,5 + (1-0,3) \times 15,05788 = 15,19$$

Với  $t=5$ , ta có:

$$\hat{y}_6 = \alpha y_5 + (1 - \alpha) \hat{y}_5 = 0,3 \times 15,2 + (1-0,3) \times 15,19 = 15,193$$

Đây là giá trị dự đoán cho doanh thu của Công ty năm 2007

#### ***2.4.2. Mô hình xu thế tuyến tính và không có biến động thời vụ (Mô hình san mũ Holt – Winters)***

Mô hình này thường áp dụng đối với sự biến động của hiện tượng qua thời gian có xu thế là tuyến tính và không có biến động thời vụ.

- Giả sử chúng ta có dãy số thời gian  $y_1, y_2, y_3, \dots, y_n$  với biến động có tính xu thế.

*Bước 1:* Chọn các hệ số  $\alpha, \beta$  ( $0 < \alpha, \beta < 1$ )

Nếu chọn hằng số san nhỏ tức là chúng ta coi các mức độ hiện thời của dãy số ít ảnh hưởng đến mức độ dự báo. Ngược lại, nếu chọn hằng số san lớn tức là chúng ta muốn dãy số san số mũ phản ứng mạnh với những thay đổi hiện tại.

**Bước 2:** Tiến hành san mũ cho giá trị ước lượng và xu thế của dãy số:

Coi giá trị của dãy số thời gian là tổng của hai thành phần: Thành phần trung bình có trọng số của các giá trị thực tế (ký hiệu là  $S_t$  – giá trị ước lượng của hiện tượng ở thời điểm  $t$ ) và thành phần xu thế (ký hiệu là  $T_t$ ). Ta có mô hình san số mũ:

$$\hat{y}_{t+1} = S_t + T_t \quad (2.15)$$

Trong đó:

$$S_t = \alpha y_t + (1 - \alpha) [S_{t-1} + T_{(t-1)}] \quad (2.16)$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{(t-1)} \quad (2.17)$$

$$\text{Đặt } S_2 = Y_2 \quad T_2 = Y_2 - Y_1$$

Tiến hành san số mũ từ thời điểm thứ 3 trở đi, ta có:

$$S_3 = \alpha y_3 + (1 - \alpha)(S_2 + T_2)$$

$$T_3 = \beta(S_3 - S_2) + (1 - \beta)T_2$$

$$S_4 = \alpha y_4 + (1 - \alpha)(S_3 + T_3)$$

$$T_4 = \beta(S_4 - S_3) + (1 - \beta)T_3$$

...

**Bước 3:** Sử dụng mức và xu thế đã được san số mũ tại thời điểm để dự đoán cho các thời điểm trong tương lai để dự đoán giá trị của hiện tượng ở thời điểm tương lai  $t + 1$ :

$$\hat{y}_{t+1} = S_t + T_t \quad (2.18)$$

Ở thời điểm tương lai ( $t + h$ ) ( $h=2, 3 \dots$ )

$$\hat{y}_{t+h} = S_t + hT_t \quad (2.19)$$

*Ví dụ:* Theo số liệu của tổng cục thống kê về GDP, giá thực tế của Việt Nam qua thời gian như sau:

<b>Năm</b> <b>Chỉ tiêu</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>
GDP (tỷ đồng)	421295	535762	613443	715307	839211

Áp dụng san mũ Holt – Winters để dự đoán cho 5 năm tới:

*Bước 1:* Chọn hệ số san:  $\alpha = 0,7$ ;  $\beta = 0,6$

*Bước 2:* Tiến hành san số mũ cho mức và cho xu thế của dãy số thời gian.

$$S_2 = y_2 = 535762$$

$$T_2 = y_2 - y_1 = 535762 - 421295 = 114467$$

$$\begin{aligned} S_3 &= \alpha y_3 + (1 - \alpha)(S_2 + T_2) \\ &= 0,7 \times 613443 + (1 - 0,7)(535.762 + 114467) \\ &= 624478,8 \end{aligned}$$

$$\begin{aligned} T_3 &= \beta (S_3 - S_2) + (1 - \beta)T_2 \\ &= 0,6(624478,8 - 535762) + (1 - 0,6) \times 114467 = 99016,88 \end{aligned}$$

$$\begin{aligned} S_4 &= \alpha y_4 + (1 - \alpha)(S_3 + T_3) \\ &= 0,7 \times 715307 + (1 - 0,7)(624478,8 + 99016,88) \\ &= 717763,6 \end{aligned}$$

$$T_4 = \beta (S_4 - S_3) + (1 - \beta)T_3$$

$$= 0,6(717763,6 - 24478,8) + 0,4 \times 99016,88 = 95577,63$$

$$S_5 = \alpha y_5 + (1 - \alpha)(S_4 + T_4)$$

$$= 0,7.839211 + 0,3(717763,6 + 95577,63) = 831450,07$$

$$T_5 = \beta(S_5 - S_4) + (1 - \beta)T_4$$

$$= 0,6(831450,07 - 717763,6) + 0,4 \times 95577,63 = 106442,93$$

Như vậy, mức độ dự báo GDP của những năm tiếp theo sẽ là:

$$\hat{y}_6 = S_5 + T_5 = 831450,07 + 106442,93 = 937893$$

$$\hat{y}_7 = S_5 + 2T_5 = 831450,07 + 2 \times 106442,93 = 1044335,93 \text{ (tỷ đồng)}$$

$$\hat{y}_8 = S_5 + 3T_5 = 831450,07 + 3 \times 106442,93 = 1150778,86 \text{ (tỷ đồng)}$$

$$\hat{y}_9 = S_5 + 4T_5 = 831450,07 + 4 \times 106442,93 = 1257221,7 \text{ (tỷ đồng)}$$

$$\hat{y}_{10} = S_5 + 5T_5 = 831450,07 + 5 \times 106442,93 = 1363664,63 \text{ (tỷ đồng)}$$

#### 2.4.3. Mô hình xu thế tuyến tính và biến động thời vụ

Mô hình này thường áp dụng đối với dự báo thời gian mà các mức độ của nó là tài liệu tháng hoặc quý của một số năm mà các mức độ trong dãy số được lập lại sau một khoảng thời gian  $h$  ( $h = 4$  đối với quý,  $h = 12$  đối với năm). Việc dự đoán có thể được thực hiện theo một trong hai mô hình sau:

$$+ \text{Mô hình cộng: } \hat{y}_{t+1} = S_t + T_t + V_{t+1} \quad (2.19)$$

Trong đó:

$$S_t = \alpha [y_t - V(t-h)] + (1-\alpha) [S_{t-1} + T_{(t-1)}] \quad (2.20)$$

$$T_t = \beta (S_t - S_{t-1}) + (1-\beta) T_{(t-1)} \quad (2.21)$$

$$V_t = \lambda(y_t - S_t) + (1 - \lambda)V_{(t-h)} \quad (2.22)$$

$$+ \text{ Mô hình nhân: } \hat{y}_{t+1} = (S_t + T_t).V_{t+1} \quad (2.23)$$

Trong đó:

$$S_t = \alpha \frac{y_t}{V_{(t-h)}} + (1 - \alpha)(S_{t-1} + T_{(t-1)})$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

$$V_t = \lambda \frac{y_t}{S_t} + (1 - \lambda).V_{(t-h)}$$

Với  $\alpha, \beta, \lambda$  là các tham số san bằng nhận giá trị trong đoạn  $[0;1]$ .

$\alpha, \beta, \lambda$  nhận giá trị tốt nhất khi tổng bình phương sai số là nhỏ nhất.

$$SSE = \sum (y_t - \hat{y}_t)^2 \Rightarrow \min$$

Tham số  $\alpha, \beta, \lambda$  không được xét một cách khách quan mà ít nhiều thông qua trực giác chủ quan, kết quả dự báo sẽ phụ thuộc vào sự lựa chọn các tham số này.

Với  $a_0(0)$  có thể là mức độ đầu tiên trong dãy số.

$a_1(0)$  có thể là lượng tăng (giảm) tuyệt đối trung bình.

$S_j(0)$ : Các chỉ số thời vụ ban đầu ( $j=1, 2, 3, \dots, k$ );  $k=4$  đối với quý;  $k=12$  đối với tháng.

Nếu  $t = 1, 2, 3, 4, 5, \dots, n$ :



Là thứ tự thời gian hay tương ứng với thứ tự các mức độ theo thời kỳ trong chuỗi thời gian thì yếu tố thời vụ  $V_j(0)$  của các mức độ trong chuỗi thời gian được tính sẽ tương ứng với các giá trị  $t \leq k$ .

$$V_j(0) = \overline{V}_j \cdot H; \quad \overline{V}_j = \sum_{j=1}^k \frac{V_j}{k}$$

$\overline{V}_j$  : Chỉ số bình quân thời vụ cho một quý hay một tháng của mỗi năm trong chuỗi thời gian.

$$V_j = \frac{y_t}{\overline{y}_t}$$

$Y_t$  : Mức độ trong chuỗi thời gian ở thời gian  $t$ .

$V_j$  : Chỉ số thời vụ của từng quý hoặc tháng trong từng năm ở thời gian  $t$ :

$\overline{y}_t$  : Số bình quân trượt để loại trừ thành phần thời vụ và thành phần ngẫu nhiên với số lượng mức độ bằng 4 đối với tài liệu quý và bằng 12 đối với tài liệu tháng.

$$H = \frac{k}{\sum \overline{V}_j}$$

*Ví dụ:* Trở lại ví dụ ở mục (2.3.1), dự đoán doanh thu của các quý theo mô hình nhân như sau:

Có tài liệu về sản lượng của doanh nghiệp A như sau:

Năm (t) Quý	Sản lượng (nghìn tấn)					Cộng theo cùng quý ( $\sum y_j$ )
	2002	2003	2004	2005	2006	
I	20	25	27	31	29	132

II	25	32	30	37	36	160
III	38	38	45	44	47	212
IV	40	60	55	62	58	275
Cộng theo cùng năm ( $\sum y_j$ )	123	155	157	174	170	779
Mức độ bình quân năm	30,75	38,75	39,25	43,5	42,5	

$S_{(0)}$ : Bình quân của 4 mức độ đầu tiên (bình quân năm):

$$S_{(0)} = \frac{20 + 25 + 38 + 40}{4} = 30,75$$

$T_0$ : Lượng tăng tuyệt đối bình quân của quý:

$$T_0 = \frac{58 - 20}{20 - 1} = 2$$

Các chỉ số thời vụ  $I_{tv}$  (Đã tính trong phần 2.3.2.)

$$\text{Quý I} = 0,713 \times 0,986 = 0,7$$

$$\text{Quý II} = 0,85 \times 0,986 = 0,838$$

$$\text{Quý III} = 1,096 \times 0,986 = 1,08$$

$$\text{Quý IV} = 1,396 \times 0,986 = 1,376$$

Với các tham số đã cho  $\alpha, \beta, \lambda$  lần lượt là: 0,4; 0,4; 0,8.

Nếu phải lựa chọn một trong hai mô hình để dự đoán thì tùy thuộc vào đặc điểm biến động của hiện tượng.

Đối với hiện tượng ít biến đổi qua thời gian thì dùng mô hình cộng.

Đối với hiện tượng biến đổi nhiều qua thời gian thì dùng mô hình nhân.

\* Ưu, nhược điểm của phương pháp san bằng mũ:

*Ưu điểm:*

- Đơn giản và có kết quả tương đối chính xác, phù hợp với dự đoán ngắn hạn cho các nhà kinh doanh cũng như lập kế hoạch ngắn hạn ở cấp vĩ mô.

- Hệ thống dự báo có thể được điều chỉnh thông qua một tham số duy nhất (tham số san bằng mũ).

Dễ dàng chương trình hoá vì chỉ phải thực hiện một số phép toán sơ cấp để xác định giá trị dự báo.

*Hạn chế:*

- Phương pháp san mũ chỉ bó hẹp trong phạm vi dự báo ngắn hạn vì không tính đến sự thay đổi cấu trúc của chuỗi thời gian mà phải tuân thủ tính ổn định theo thời gian của các quá trình kinh tế - xác hội.

## **2.5. Sử dụng chương trình SPSS để dự báo theo các mô hình**

### **2.5.1. Dự đoán bằng hàm xu thế**

\* *Nhập tài liệu*

+ Một cột là biến theo thứ tự các năm, một cột là thời gian (Years – năm; Years, quarters – năm, quý; Years, months - năm, tháng) (nếu là năm ta nhấp chuột vào Years, ô nhỏ sẽ hiện số 1900, ta xoá đi và đánh số năm đầu tiên trong dãy số).

\* *Thăm dò bằng đồ thị*

### Analyze/ Regression/ Curve Estimation

- Đưa y vào Dependent và Years vào Variable

- Time/ Linear/ Display ANOVA table/ Save/ Predicted values/  
Predict through/ đánh số năm cần dự báo vào hình chữ nhật đứng sau  
year/ continue/ OK.

*\* Một số kết quả*

Constant – tham số a

Time – tham số b

### 2.5.2. Dự đoán bằng san bằng mũ

*\* Mô hình đơn giản*

- Nhập tài liệu.

- Analyze/ Time Series/ Exponential Smoothing.

- Save/ Do not create/ Continue/ OK.

- Đưa Y vào hình vuông bên phải.

- Simple/ Parameters/ Grid search (nằm trong hình vuông thứ  
nhất General)/ Continue/ OK.

*\* Mô hình xu thế tuyến tính không biến động thời vụ*

- Chọn Holt/ Parameters/ Grid Search (có chữ General hình  
vuông bên trái)/ Grid Search (hình vuông bên phải có chữ Trend).

- Continue/ OK.

- Parameters.

- Nhấp chuột vào Value (trái) – đánh số 0.9.

- Nhấp chuột vào Value (phải) – đánh số 0.0.
- Continue/ Save/ Predict through/ đánh số năm cần dự báo vào ô Year/ Continue/ OK/ Đóng cửa màn hình Output sẽ có kết quả dự báo
- \* Mô hình xu thế tuyến tính có biến động thời vụ*
- Nhập tài liệu
- Define Dates/ Year Quarters/ đánh số năm đầu tiên trong dãy số vào hình chữ nhật thứ nhất.
- Analyze/ Time Serier/ Exponential Smoothing/ Winters.
- Đưa Y vào hình vuông dưới chữ Variables.
- Đưa Quarters vào hình chữ nhật dưới chữ Seasonal.
- Parameters/ Grid Search ở trong các hình vuông của General (Alpha), Trend (Gramma), Seasonal (Delta)/ Continue/ OK.

### *Chương 3*

## **PHƯƠNG PHÁP HỒI QUY ĐƠN, HỒI QUY BỘI VÀ THỐNG KÊ HỒI QUY**

### **♦ Phương pháp hồi quy**

Hồi quy - nói theo cách đơn giản, là đi ngược lại về quá khứ (regression) để nghiên cứu những dữ liệu (data) đã diễn ra theo thời gian (dữ liệu chuỗi thời gian - time series) hoặc diễn ra tại cùng một thời điểm (dữ liệu thời điểm hoặc dữ liệu chéo - cross section) nhằm tìm đến một quy luật về mối quan hệ giữa chúng. Mối quan hệ đó được biểu diễn thành một phương trình (hay mô hình) gọi là: phương trình hồi quy, mà dựa vào đó có thể giải thích bằng các kết quả lượng hoá về bản chất, hỗ trợ củng cố các lý thuyết và dự báo tương lai.

Theo thuật ngữ toán, phân tích hồi quy là sự nghiên cứu mức độ ảnh hưởng của một hay nhiều biến số (biến giải thích hay biến độc lập - independent variable), đến một biến số (biến kết quả hay biến phụ thuộc - dependent variable), nhằm dự báo biến kết quả dựa vào các giá trị được biết trước của các biến giải thích.

Trong phân tích hoạt động kinh doanh cũng như trong nhiều lĩnh vực khác, hồi quy là công cụ phân tích đầy sức mạnh không thể thay thế, là phương pháp thống kê toán dùng để ước lượng, dự báo những sự kiện xảy ra trong tương lai dựa vào quy luật quá khứ.

### 3.1. Phương pháp hồi quy đơn

Còn gọi là hồi quy đơn biến, dùng xét mối quan hệ tuyến tính giữa một biến kết quả và một biến giải thích hay là biến nguyên nhân (nếu giữa chúng có mối quan hệ nhân - quả). Trong phương trình hồi quy tuyến tính, một biến gọi là: biến phụ thuộc, một biến kia là tác nhân gây ra sự biến đổi, gọi là biến độc lập.

Phương trình hồi quy đơn biến (đường thẳng) có dạng tổng quát:

$$Y = a + bX \quad (3.1)$$

Trong đó:

Y: Biến số phụ thuộc (dependent variable);

X: Biến số độc lập (independent variable);

a: Tung độ gốc hay nút chặn (intercept);

b: Độ dốc hay hệ số góc (slope).

y trong phương trình trên được hiểu là y ước lượng, người ta thường viết dưới hình thức  $\hat{Y}$

*Ví dụ:*

Phương trình tổng chi phí của doanh nghiệp có dạng:

$$Y = a + bX$$

Trong đó:

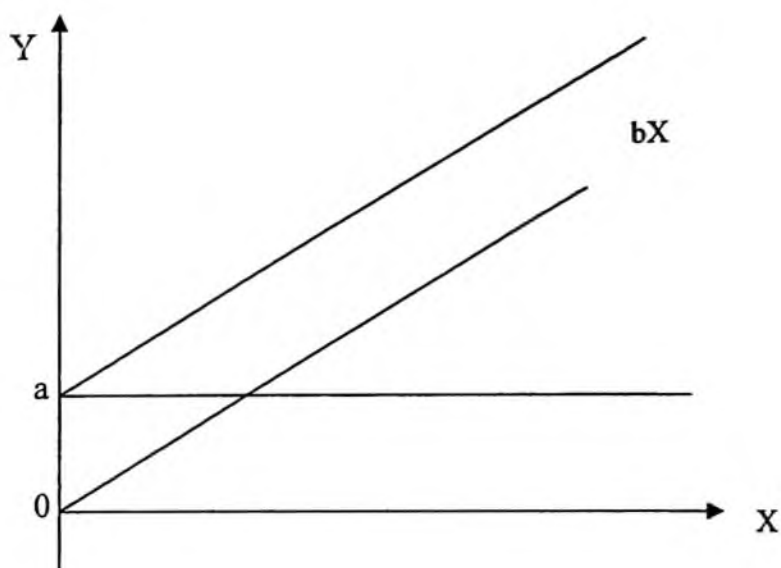
Y: Tổng chi phí phát sinh trong kỳ;

X: Khối lượng sản phẩm tiêu thụ;

a: Tổng chi phí bất biến;

b: chi phí khả biến đơn vị sản phẩm;

bX: Tổng chi phí khả biến.



*Đồ thị 3.1. Ứng xử của các loại chi phí*

### ***Nhận xét***

Với phương trình trên, tổng chi phí  $Y$  chịu ảnh hưởng trực tiếp của khối lượng hoạt động  $X$  theo quan hệ tỷ lệ thuận. Khi  $X$  tăng dẫn đến  $Y$  tăng; khi  $X$  giảm dẫn đến  $Y$  giảm.

Khi  $X = 0$  thì  $Y = a$ : Các chi phí như tiền thuê nhà, chi phí khấu hao, tiền lương thời gian và các khoản chi phí hành chính khác là những chi phí bất biến, không chịu ảnh hưởng từ thay đổi của khối lượng hoạt động.

Đường biểu diễn  $a$  song song với trục hoành. Trị số  $a$  là hệ số cố định, thể hiện “chi phí tối thiểu” trong kỳ của doanh nghiệp (nút chặn trên đồ thị).

Trị số  $b$  quyết định độ dốc (tức độ nghiêng của đường biểu diễn chi phí trên đồ thị).



Đường tổng chi phí  $Y = a + bX$  và đường chi phí khả biến  $bX$  song song với nhau vì giữa chúng có cùng chung một độ dốc  $b$  (slope). Xuất phát điểm của đường tổng chi phí bắt đầu từ nút chặn  $a$  (intercept =  $a$ ) trên trục tung; trong khi đó, đường chi phí khả biến lại bắt đầu từ gốc trục tọa độ vì có nút chặn bằng 0 (intercept = 0). Hay nói một cách khác, theo nội dung kinh tế, khi khối lượng hoạt động bằng 0 ( $X=0$ ) thì chi phí khả biến cũng sẽ bằng 0 ( $bX=0$ ).

*Ví dụ chi tiết:*

Có tình hình về chi phí hoạt động (tài khoản 641 và tài khoản 642: chi phí bán hàng và chi phí quản lý doanh nghiệp) và doanh thu (tài khoản 511) tại một doanh nghiệp được quan sát qua các dữ liệu của 6 kỳ kinh doanh như sau (đơn vị tính: triệu đồng).

**Bảng 3.1. Tình hình thực hiện chi phí của 6 kỳ kinh doanh**

Kỳ kinh doanh	Doanh thu bán hàng	Chi phí hoạt động
1	1.510	323
2	1.820	365
3	2.104	412
4	2.087	410
5	1.750	354
6	2.021	403

*Yêu cầu:* Phân tích cơ cấu chi phí hoạt động (bất biến, khả biến) của doanh nghiệp.

*Hướng dẫn:*

Yêu cầu của vấn đề là thiết lập phương trình chi phí hoạt động của doanh nghiệp, tức đi tìm giá trị các thông số  $a$ ,  $b$  với mục đích phát hiện quy luật biến đổi của chi phí này trước sự thay đổi của

doanh thu, nhằm đến việc dự báo chi phí cho các quy mô hoạt động khác nhau hoặc cho các kỳ kinh doanh tiếp theo.

Phương trình chi phí hoạt động có dạng:

$$Y = a + bX$$

Trong đó:

a: Tổng chi phí bất biến.

b: chi phí khả biến 1 đơn vị doanh thu.

X: Doanh thu bán hàng.

Y: Tổng chi phí hoạt động.

Có nhiều phương pháp thống kê tính a, b như:

*Phương pháp cực trị:*

Còn gọi là phương pháp cận trên - cận dưới (high - low method).  
Cụ thể để tìm trị số a, b của phương trình theo ví dụ trên bằng cách sử dụng công thức và cách tính toán như sau:

$$b = \frac{\text{Hiệu số của chi phí cao nhất và thấp nhất}}{\text{Hiệu số của doanh thu cao nhất và thấp nhất}}$$

$$b = \frac{412 - 323}{2.104 - 1.510} = 0,15$$

Trong đó:

Chi phí cực đại: 412

Chi phí cực tiểu: 323

Doanh thu cực đại: 2.104

Doanh thu cực tiểu: 1.510

Từ phương trình:  $Y = a + bX$ , suy ra:  $a = Y - bX$

Tại điểm đạt doanh thu cao nhất (high), ta có:

$$a = 412 - 0,15 \times 2.104 = 96,4$$

Tại điểm đạt doanh thu thấp nhất (low), ta có:

$$a = 323 - 0,15 \times 1.510 = 96,4$$

Phương trình chi phí kinh doanh đã được thiết lập:

$$Y = 96,4 + 0,15X$$

**Bảng 3.2. Kết quả các đại lượng đặc trưng thống kê**

Column1 (doanh thu)		Column2 (chi phí)		Giải thích
Mean	1.882,00	Mean	377,83	Giá trị trung bình
Standard Error	94,92	Standard Error	14,80	Sai số chuẩn
Median	1.920,50	Median	384,00	Trung vị
Mode	#N/A	Mode	#N/A	Yếu vị
Standard Deviation	232,50	Standard Deviation	36,26	Độ lệch chuẩn
Sample Variance	54.056,40	Sample Variance	1.314,97	Phương sai (mẫu)
Kurtosis	-0,49	Kurtosis	-1,30	Độ nhọn
Skewness	-0,76	Skewness	-0,58	Độ nghiêng
Range	594,00	Range	89,00	Khoảng (miền)
Minimum	1.510,00	Minimum	323,00	Giá trị tối thiểu
Maximum	2.104,00	Maximum	412,00	Giá trị tối đa
Sum	11.292,00	Sum	2.267,00	Tổng cộng giá trị
Count	6,00	Count	6,00	Số lần quan sát

*Lưu ý:*

- Phương pháp cực trị rất đơn giản, dễ tính toán nhưng thiếu chính xác trong những trường hợp dữ liệu biến động bất thường.

- Trường hợp tập dữ liệu có số quan sát lớn, việc tìm thấy những giá trị cực trị gặp khó khăn và dễ nhầm lẫn, Microsoft Excel sẽ cung cấp một cách nhanh chóng và chính xác các giá trị thống kê: Max, Min, Range (= Max-Min) như sau:

Lệnh sử dụng trong Microsoft Excel: Tools / Data Analysis.../ Descriptive Statistics/ OK / Summary Statistics / OK.

Nếu trong Tools không hiện hành sẵn Data Analysis, ta dùng lệnh: Tools / Add - Ins / Analysis ToolPak / OK.

Giải thích các thông số tính được cụ thể tại cột chi phí:

**Mean** (giá trị trung bình): Bình quân số học (Average) của tất cả các giá trị quan sát. Được tính bằng cách lấy tổng giá trị các quan sát (Sum) chia cho số quan sát (Count).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{2267}{6} = 377,83$$

**Standard Error** (sai số chuẩn): Dùng để đo độ tin cậy của giá trị trung bình mẫu. Được tính bằng cách lấy độ lệch chuẩn (Standard Deviation) chia cho căn bậc 2 của số quan sát.

$$S_x = \frac{\delta}{\sqrt{n}} = \frac{36,26}{\sqrt{6}} = 14,80$$

Ta có thể nói: có khả năng 95% là giá trị trung bình nằm trong khoảng cộng trừ (+/-) 2 lần sai số chuẩn so với giá trị trung bình. Theo ví dụ trên, đó là khoảng:

$[377,83 - (2 \times 14,8); 377,83 + (2 \times 14,8)]$  tức là khoảng:  $[348,23; 407,43]$

Dựa vào công thức trên ta cũng thấy rằng: với độ lệch chuẩn  $\delta$  không đổi,  $n$  càng lớn thì  $S$  càng nhỏ. Tức khoảng dao động sẽ hẹp hơn và độ chính xác sẽ cao hơn. Người ta cũng dựa vào công thức này để tính số quan sát cần thiết  $n$ .

**Median** (trung vị): Giá trị nằm ở vị trí trung tâm (khác với giá trị trung bình Mean). Được tính bằng cách:

- Nếu số quan sát  $n$  là số lẻ: Sắp xếp các giá trị quan sát từ nhỏ đến lớn, giá trị đứng vị trí chính giữa là số trung vị.

- Nếu số quan sát  $n$  là số chẵn: Sắp xếp các giá trị quan sát từ nhỏ đến lớn, trung bình cộng của 2 giá trị đứng ở vị trí chính giữa là số trung vị.

Theo ví dụ trên, ta sắp xếp các quan sát có giá trị từ nhỏ đến lớn: 323, 354, 365, 403, 410, 412.

$$\text{Median} = \frac{365 + 403}{2} = 384$$

**Mode** (yếu vị): Giá trị xuất hiện nhiều lần nhất. Theo ví dụ trên, ta không có yếu vị nào cả (#N/A)

**Standard Deviation** (độ lệch chuẩn): Được xem như là độ lệch trung bình, đại diện cho các độ lệch (hiệu số) giữa các giá trị quan sát thực và giá trị trung bình (Mean). Độ lệch chuẩn là đại lượng dùng để đo mức độ phân tán (xa hay gần) của các giá trị quan sát xung quanh giá trị trung bình. Được tính bằng cách lấy căn bậc hai của phương sai  $\sigma^2$  (trung bình của phương sai các độ lệch: độ lệch âm - negative deviation và độ lệch dương - positive deviation)

$$\delta = \sqrt{\delta^2} = \sqrt{1.314,97} = 36,26 \text{ (}\sigma \text{ đọc là sigma)}$$

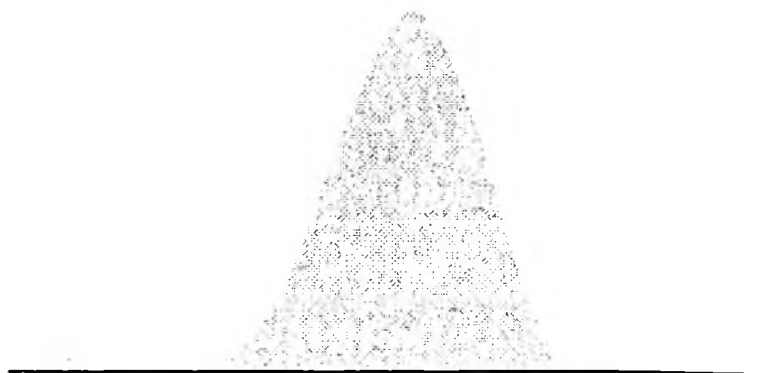
**Sample Variance** (phương sai mẫu): Là trung bình của bình phương các độ lệch. Giống như độ lệch chuẩn, nó cũng dùng để xem mức độ phân tán các giá trị quan sát thực xung quanh giá trị trung bình. Được tính bằng cách lấy tổng các bình phương các độ lệch (tổng các hiệu số giữa giá trị quan sát thực và giá trị trung bình) chia cho số quan sát trừ 1 ( $n - 1$ ). Theo ví dụ trên ta có:

$$\delta^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = 1.314,97$$

( $\sigma^2$  đọc là sigma bình phương)

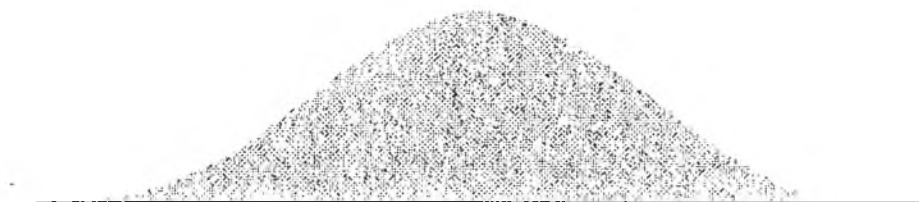
**Kurtosis** (độ nhọn): Hệ số đặc trưng thống kê dùng để đo mức độ “đồng nhất” của các giá trị quan sát.

- Đường cong rất chóp (very peaked): Nhọn đứng, kurtosis  $> 3$ . Nếu đường biểu diễn dưới đây mô tả phân phối các giá trị doanh thu, ta có thể nói rằng đa số các giá trị doanh thu rất gần với nhau (the same revenue) dù có một số ít mang giá trị rất nhỏ hoặc rất lớn.



- Đường cong rất bẹt (very flat): Phẳng nằm, kurtosis  $< 3$ . Nếu đường biểu diễn dưới đây mô tả phân phối các giá trị doanh thu, ta

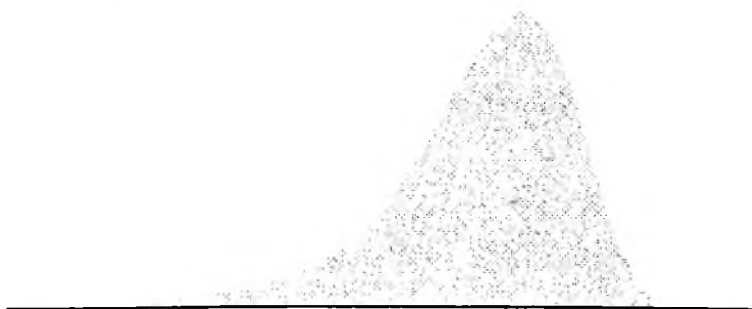
có thể nói rằng đa số các giá trị doanh thu được trải đều từ nhỏ đến lớn trong một khoảng rộng hơn.



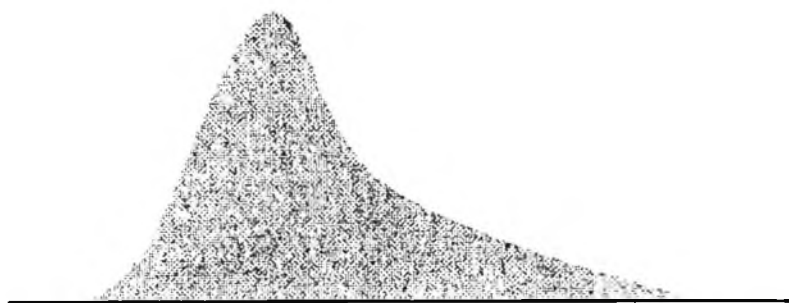
Theo ví dụ trên, độ chóp bằng: - 1,30

**Skewness** (độ nghiêng): Hệ số dùng để đo “độ nghiêng” khi phân phối xác suất không cân xứng theo hình chuông đều.

- Nghiêng về trái ta còn gọi là “nghiêng âm” (Skewed to the left), skewness  $< -1$ : nghiêng nhiều,  $> 0,5$ : nghiêng ít. Nếu đường biểu diễn dưới đây mô tả phân phối các giá trị doanh thu, ta có thể nói rằng đa số các giá trị doanh thu gần với doanh thu lớn nhất dù có một số ít mang giá trị nhỏ hơn hoặc rất nhỏ (ở bên trái).



- Nghiêng về phải ta còn gọi là “nghiêng dương” (Skewed to the right), skewness  $> 1$ : nghiêng nhiều,  $< 0,5$ : nghiêng ít. Nếu đường biểu diễn dưới đây mô tả phân phối các giá trị doanh thu, ta có thể nói rằng đa số các giá trị doanh thu gần với doanh thu nhỏ nhất dù có một số ít mang giá trị lớn hơn hoặc rất lớn (ở bên phải).



Theo ví dụ trên, độ nghiêng bằng: -0,58.

**Range** (khoảng)(also range width hay bề rộng của khoảng): Độ dài của khoảng quan sát (khoảng biến thiên), được tính bằng lấy giá trị quan sát cực đại Max trừ đi giá trị quan sát cực tiểu Min.

$$\text{Range} = \text{Max} - \text{Min} = 412 - 323 = 89$$

**Minimum** (giá trị quan sát cực tiểu): Giá trị nhỏ nhất trong các quan sát.

$$\text{Min} = 323$$

**Maximum** (giá trị quan sát cực đại): Giá trị lớn nhất trong các quan sát.

$$\text{Max} = 412$$

**Sum** (tổng cộng giá trị của các quan sát): Tổng cộng tất cả các giá trị của tất cả các quan sát trong tập dữ liệu.

Theo ví dụ trên, ta có:

$$\text{Sum} = \sum_{i=1}^n X_i = 2.267$$

**Count** (số quan sát): Số đếm của số lần quan sát (n). Theo tập dữ liệu ở ví dụ trên, ta có:  $n = 6$



### 3.2. Phương pháp hồi quy đa biến

Còn gọi là phương pháp hồi quy đa biến, dùng phân tích mối quan hệ giữa nhiều biến số độc lập (tức biến giải thích hay biến nguyên nhân) ảnh hưởng đến một biến phụ thuộc (tức biến phân tích hay biến kết quả).

Trong thực tế, có rất nhiều bài toán kinh tế - cả lĩnh vực kinh doanh và kinh tế học, phải cần đến phương pháp hồi quy đa biến. Chẳng hạn như phân tích những nhân tố ảnh hưởng đến thu nhập quốc dân, sự biến động của tỷ giá ngoại hối; xét doanh thu trong trường hợp có nhiều mặt hàng; phân tích tổng chi phí với nhiều nhân tố tác động; phân tích giá thành chi tiết; những nguyên nhân ảnh hưởng đến khối lượng tiêu thụ...

Một chỉ tiêu kinh tế chịu sự tác động cùng lúc của rất nhiều nhân tố thuận chiều hoặc trái chiều nhau. Chẳng hạn như doanh thu lệ thuộc vào giá cả, thu nhập bình quân xã hội, lãi suất tiền gửi, mùa vụ, thời tiết, quảng cáo tiếp thị... Mặt khác, giữa những nhân tố lại cũng có sự tương quan tuyến tính nội tại với nhau. Phân tích hồi quy giúp ta vừa kiểm định lại giả thiết về những nhân tố tác động và mức độ ảnh hưởng, vừa định lượng được các quan hệ kinh tế giữa chúng. Từ đó, làm nền tảng cho phân tích dự báo và có những quyết sách phù hợp, hiệu quả, thúc đẩy tăng trưởng.

Phương trình hồi quy đa biến dưới dạng tuyến tính:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_iX_i + b_nX_n + e \quad (3.2)$$

Trong đó:

Y: Biến số phụ thuộc (kết quả phân tích);

$b_0$ : Tung độ gốc;

$b_i$ : Các độ dốc của phương trình theo các biến;

$X_i$ ;  $X_i$ : Các biến số (các nhân tố ảnh hưởng);

$e$ : Các sai số;

*Lưu ý*:  $Y$  trong phương trình trên được biểu hiện là  $Y$  ước lượng, người ta thường viết dưới hình thức có nón ( $\hat{Y}$ ).

Mục tiêu của phương pháp hồi quy đa biến là dựa vào dữ liệu lịch sử các biến số  $Y_i$ ,  $X_i$ , dùng thuật toán để đi tìm các thông số  $b_0$  và  $b_i$  xây dựng phương trình hồi quy để dự báo cho ước lượng trung bình của biến  $Y_i$ .

### 3.3. Phương pháp thống kê hồi quy

Còn gọi là thống kê hồi quy đơn giản (simple regression statistical) dùng phương pháp thống kê toán để tính các hệ số  $a$ ,  $b$  của phương trình hồi quy dựa trên toàn bộ quan sát của tập dữ liệu. Đây là phương pháp đáng tin cậy nhất và vì vậy đòi hỏi công phu hơn.

Vẫn dùng số liệu ở ví dụ trên, lập bảng tính các trị số cơ sở rồi căn cứ vào công thức để tính các thông số của phương trình.

Ta có công thức trong thống kê toán  $a = \bar{Y} - b\bar{X}$

$$b = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

*Chứng minh công thức:*

Công thức trên được chứng minh từ phương pháp hồi quy các bình phương tối thiểu của các hiệu số (độ lệch: Deviation) giữa

các giá trị quan sát và giá trị ước lượng của biến số phụ thuộc ( $\hat{Y} = a + bX_i$ )

Với phương pháp tổng các bình phương tối thiểu, gọi  $\hat{e}_i^2$  là bình phương các độ lệch, ta có:

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y - \hat{y})^2 = \sum_{i=1}^n (y_i - a - bX_i)^2 \quad (3.3)$$

$$\text{Min} \sum_{i=1}^n \hat{e}_i^2 \quad (3.4)$$

Giải hệ phương trình vi phân để tìm giá trị các thông số.

Lấy đạo hàm riêng phần theo a và cho bằng 0:

$$\frac{\partial}{\partial a} \sum_{i=1}^n (Y_i - a - bX_i)^2 = 0 \quad (3.5)$$

Lấy đạo hàm riêng phần theo b và cho bằng 0:

$$\frac{\partial}{\partial b} \sum_{i=1}^n (Y_i - a - bX_i)^2 = 0 \quad (3.6)$$

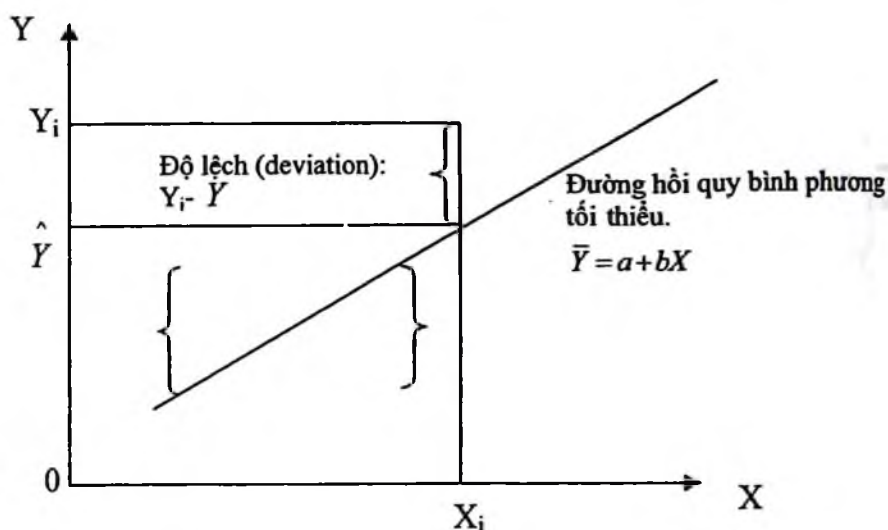
Lấy đạo hàm rồi cùng chia cho -2 (hay nhân với 1/2) ta có hệ phương trình chuẩn với n quan sát:

$$\sum XY = a \sum X + b \sum X^2 \quad (3.7)$$

$$\sum Y = na + b \sum X \quad (3.8)$$

Dùng phương pháp khử, giải hệ phương trình có 2 ẩn số, ta lần lượt có được giá trị các thông số a, b như các công thức (1.3) và (1.4) nêu trên.

Dễ dàng thấy được ý nghĩa các độ lệch tối thiểu qua đồ thị sau:



*Đồ thị 3.2. Độ lệch của các trị quan sát so với giá trị ước lượng*

Giải thích đồ thị:

Đường hồi quy  $Y = a + bX$  là đường ước lượng tốt nhất, chứa các giá trị ước lượng của  $Y$  mà độ lệch trung bình giữa chúng và giá trị quan sát thực là nhỏ nhất (tối thiểu).

Các độ lệch nằm phía trên đường ước lượng nhìn từ gốc của trục tọa độ, gọi là độ lệch dương (Positive deviation); các độ lệch nằm phía dưới đường ước lượng nhìn từ gốc của trục tọa độ, gọi là độ lệch âm (Negative deviation).

Tại sao là bình phương tối thiểu?

Mục đích cuối cùng của phương pháp hồi quy là dùng để giải thích hoặc dự báo một đối tượng cần nghiên cứu. Cụ thể là đi tìm giá trị các thông số  $a$ ,  $b$  để xây dựng phương trình hồi quy tuyến tính (đường thẳng) có dạng tổng quát:

$$\hat{Y} = a + bX$$

**Bảng 3.3. Các trị số cơ sở thống kê**

N	$x_i$	$y_i$	$x_i^2$	$y_i^2$	$x_i y_i$	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1.510	323	2.280.100	104.329	487.730	-372	-55	20.398	138.384	3.007
2	1.820	365	3.312.400	133.225	664.300	-62	-13	796	3.844	165
3	2.104	412	4.426.816	169.744	866.848	222	34	7.585	49.284	1.167
4	2.087	410	4.355.569	168.100	855.670	205	32	6.594	42.025	1.035
5	1.750	354	3.062.500	125.316	619.500	-132	-24	3.146	17.424	568
6	2.021	403	4.084.441	162.409	814.463	139	25	3.498	19.321	633
$\Sigma$	11.292	2.267	21.521.826	863.123	4.308.511	0	1	42.017	270.282	6.575

Mỗi giá trị ước lượng (ước lượng điểm) là giá trị ước lượng trung bình điểm của biến kết quả  $y_i$ . Khả năng chỉ có thể xảy ra các giá trị trong một “khoảng ước lượng” với một “độ tin cậy” nhất định mà thôi. Vì xác suất để giá trị thực  $y_i$  bằng với giá trị ước lượng điểm  $\hat{y}$  là bằng 0, hay nói cách khác là rất khó có khả năng xảy ra.

Ý nghĩa của phương pháp bình phương tối thiểu là làm sao cho độ lệch trung bình giữa  $\hat{y}$  và  $y_i$  nhỏ nhất  $(y_i - \hat{y}) \rightarrow 0$

Trong đó,  $y_i$  là các giá trị quan sát thực và  $\hat{y} = a + bx$  là các giá trị ước lượng (giá trị trung bình) của  $y_i$ .

Khi ấy, giá trị ước lượng “gần với” giá trị quan sát thực và phương trình hồi quy dùng để dự báo sẽ trở nên khả thi, thích hợp nhất và chính xác nhất trong điều kiện có thể.

$$\bar{y} = \frac{2.267}{6} = 377,83 \approx 378$$

Trước hết, xét mức độ tương quan (correlation) giữa biến số phụ thuộc và biến số độc lập bằng công thức:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$R = +1$ : Tương quan hoàn toàn và đồng biến;

$R = -1$ : Tương quan hoàn toàn và nghịch biến;

$|R| = \text{càng gần } 1$ : Tương quan càng mạnh ( $0,8 < |R| < 1$ );

$|R| = \text{từ } 0,4 \text{ đến } 0,8$ : Tương quan trung bình;

$|R| = \text{nhỏ hơn } 0,4$ : Tương quan yếu;

$$R = \frac{42.017}{\sqrt{(270.282)(6.575)}} = 0,993$$

Ý nghĩa của độ tương quan nói lên cường độ của mối quan hệ tuyến tính của hai biến x và y.

Trở lại, thay các giá trị đã tính ở bảng 3.7 vào công thức (3.3) và (3.4) ở trên, ta có:

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{42.017}{270.282} = 0,115$$

$$a = \bar{y} - b\bar{x} = 377.83 - (0,115 \times 1882) = 86,12$$

Vậy phương trình hồi quy có dạng  $y = a + bx$  sẽ là:

$$y = 86,12 + 0,155x$$

**\* *Tính trên phần mềm Microsoft Excel:***

Có hai cách thực hiện trên Excel:

Cách 1: dùng hàm Fx: Paste function

Tìm trị số b (slope), sử dụng lệnh: Insert / Fx / Statistical (select a category: chọn loại hàm) / slope (select a function: lựa chọn tên hàm) / OK / quét đánh dấu khối cột dữ liệu y và cột dữ liệu x/OK.

Tìm trị số a (intercept), sử dụng lệnh giống như tìm trị số a, chỉ thay đổi tên hàng Slope bằng tên hàm Intercept (function name)

Tìm trị số R (correlation), dùng lệnh: Insert / Fx / Statistical (select a category: lựa chọn loại hàm)/ Correl (select a function: lựa chọn tên hàm) / OK / quét đánh dấu khối cột dữ liệu x và cột dữ liệu y / OK.

Cách 2: Dùng Regression (thường dùng để chạy hồi quy đa biến). Khi thao tác trên Microsoft Excel, ta sử dụng lệnh: Tools/ Data Analysis/ Regression/ OK.

Trong phần Input (nhập đầu vào):

Nhập dữ liệu y vào ô: Input Y Range;

Nhập dữ liệu x vào ô: Input X Range;

Trong phần Output options (vị trí đầu ra) có hai lựa chọn:

Chọn sheet mới: dùng New worksheet ply;

Chọn sheet hiện hành: dùng Output Range.

Chương trình Microsoft Excel sẽ cho bảng kết quả sau

Giải thích bảng 3.4:

- Multiple R = 0,9967 là độ tương quan giữa y và x (tương quan mạnh);

- R square ( $R_2$ ) = 0,9935: Hệ số xác định (determination), biểu hiện khả năng giải thích của các biến độc lập x đến biến phụ thuộc y (khả năng giải thích cao);

- Đọc trị số a, b ở cột Coefficients - các hệ số: Intercept - tung độ gốc ( $a=85,265$ ); X Variable 1 - độ dốc với biến độc lập x ( $b = 0,155$ ).

- Trị số thống kê t-stat: 7,136 và 24,641 > 1,96, thể hiện sự “có ý nghĩa về mặt thống kê” ở mức ý nghĩa 5% trong khoảng: cận trên - Upper, cận dưới - Lower. Cận trên và cận dưới của Intercept là (118,44 ; 52,09) và của Slope là (0,17 ; 0,14).

- Một số chỉ tiêu dùng để kiểm định, như ANOVA trong bảng kết quả hồi quy không đề cập hết trong phạm vi môn học này.



**Bảng 3.4. Kết quả hồi quy đơn biến, cho bởi Microsoft Excel**

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.996722138								
R Square	0.993455021								
Adjusted R Square	0.991818777								
Standard Error	3.27994452								
Observations	6								
ANOVA									
	df	SS	MS	F	Significance F				
Regression	1	6531.801	6531.801	607.1555	1.6099E-05				
Residual	4	43.03214	10.75804						
Total	5	6574.833							
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
Intercept	85.26485301	11.94873	7.135891	0.00204	52.08985086	118.4399	52.08985	118.4399	
X Variable 1	155.4561532	6.308962	24.64053	1.61E-05	137.9396656	172.9726	137.9397	172.9726	

## *Chương 4*

# **PHƯƠNG PHÁP BOX - JENKINS (ARIMA)**

### **4.1. Tính ổn định của một chuỗi**

Trước khi xử lý một chuỗi thời gian, nghiên cứu các tính ngẫu nhiên của nó là bước cần thiết cho phép ta đánh giá một cách tổng quát về số liệu nghiên cứu. Nếu kỳ vọng toán vào phương sai của nó thay đổi theo thời gian, chuỗi được xem như là không ổn định. Trong trường hợp ngược lại, ta nói chuỗi ổn định. Xét chuỗi  $y_t$ , về mặt toán học một chuỗi ổn định phải thỏa mãn các điều kiện sau:

$$E(y_t) = E(y_{t+m}) = \text{cte } y_k \text{ và } m$$

$$\text{Var}(y_t) < r_k$$

$$\text{Cov}(y_t ; y_{t+k}) = E ((y_t - ) ( y_{t+k} - ) ) = r_k = \text{hằng số}$$

Với tính chất như vậy ta có thể thấy một nhiễu trắng (giới thiệu sau) là một chuỗi ổn định vì nó thỏa mãn tính chất nêu trên. Một chuỗi thời gian là ổn định khi nó là đại diện của một quá trình nghiên cứu ổn định. Nói một cách cụ thể hơn đó là chuỗi không có tính xu thế, không có tính chu kỳ.

### **4.2. Hàm số tự tương quan đơn và tự tương quan riêng phần**

**Hệ số tương quan riêng phần** là hệ số dùng để đánh giá quan hệ giữa hai biến khi ảnh hưởng của biến thứ ba được loại trừ.

Hàm số tự tương quan  $\hat{P}_k$  nhằm xác định sự tương quan của chuỗi và chính nó nhưng lệch đi một chu kỳ  $k$  bất kỳ (xem bảng sau). Công thức xác định hàm số tương quan  $\hat{P}_k$  như sau:

$$\hat{P}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum (y_t - \bar{y})^2}$$

Tính chất:

$$\hat{P}_0 = 1 \text{ và } \hat{P}_k = \hat{P}_{-k}$$

Bảng sau đây giới thiệu cách tính hàm tự tương quan.

Khảo sát chuỗi quan trắc  $y_t$ . Các chuỗi lệch  $y_{t-k}$  tương ứng cũng được giới thiệu:

**Bảng 4.1. Xác định các chuỗi lệch  $y_{t-k}$**

k	0	1	2	3	4
t	$y_t$	$y_{t-1}$	$y_t$	$y_t$	$y_{t-2}$
1	123				
2	130	123			
3	125	130	123		
4	138	125	130	123	
5	145	138	125	130	123
6	142	145	138	125	130
7	141	142	145	138	125
8	146	141	142	145	138
9	147	146	141	142	145
10	157	147	146	141	142
11	150	157	147	146	141
12	145	150	157	147	146

Kết quả tính giá trị trung bình vô phương sai của các chuỗi và hàm số tự tương quan  $\hat{P}_k$  được trình bày trong bảng sau:

**Bảng 4.2**

Trung bình $y_t$	140,7	142,3	143,6	145,6	146,6
Trung bình $y_{t-k}$	140,7	140,3	139,4	137,4	136,2
Phương sai $y_t$	95	72,4	62,8	27,1	22,2
Phương sai $y_{t-k}$	95	101,8	101,8	74,9	71,4
$\hat{P}_k$	1	0,77	0,62	0,59	0,55

Với định nghĩa của hàm số tự tương quan trên, ta thấy không tiện lợi trong việc tính toán vì nó đòi hỏi phải lùi lại khi tính mỗi số hạng  $R_k$ . Do đó trong thực tế áp dụng ta thường tính hàm tự tương quan cho mẫu bằng một công thức đơn giản hơn như sau:

$$\hat{P}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum (y_t - \bar{y})^2}$$

với giá trị trung bình của chuỗi tính trên  $n$  chu kỳ.

Khi số lượng quan trắc đủ lớn, hai cách tính giá trị hàm tự tương quan trên cho kết quả rất gần nhau ( $\hat{P}_k \sim \hat{P}_{-k}$ ). Hàm số tự tương quan riêng phần bắt nguồn từ khái niệm tương quan riêng phần. Với khái niệm này cho phép ta đánh giá, ví dụ, ảnh hưởng của  $x_1$  lên  $x_2$  trong bối cảnh loại hết các ảnh hưởng của các biến khác  $x_3, x_4, \dots, x_k$ .

Tương tự như vậy ta, định nghĩa hàm tự tương quan riêng phần có mức độ trễ  $k$  như là hệ số tương quan riêng phần giữa  $y_t$  và  $y_{t-k}$ , có nghĩa là trong đó các ảnh hưởng của các biến  $y_{t-1}, y_{t-2} \dots y_{k+1}$  được loại bỏ.

### 4.3. Kiểm định nhiễu trắng

#### 4.3.1. Phân tích hàm tự tương quan

Mục đích của phân tích hàm tự tương quan nhằm xác định khả năng có tính tự tương quan trong chuỗi khảo sát (thường là chuỗi sai số) hay không. Khi chúng ta phân tích hàm tự tương quan của một chuỗi thời gian, một câu hỏi luôn luôn đặt ra là các hệ số  $\hat{P}_k$  nào khác 0. Thật vậy, nếu ta hoàn toàn không có giá trị nào của  $\hat{P}_k$  khác 0 ta nói quá trình nghiên cứu không có “bộ nhớ”. Nó hoàn toàn không có tính xu thế cũng như không có tính chu kỳ. Ví dụ, trong trường hợp nếu chuỗi có tính chu kỳ theo tháng ta sẽ thấy giá trị của  $P_{12}$  sẽ lớn (tương quan giữa  $y_t$  và  $y_{t-12}$ ) Chuỗi chắc chắn có tính chu kỳ. Kiểm định cho  $\hat{P}_k$  có giá trị khác 0 được thực hiện dựa vào nguyên tắc kiểm định giả thiết như sau:

$$H_0: \hat{P}_k = 0$$

$$H_1: \hat{P}_k \neq 0$$

Trong thực hành, tác giả Quenouille đã chứng minh được rằng với một mẫu có kích thước tương đối lớn, hệ số  $\hat{P}_k$  tiến một cách tiệm cận về một phân phối chuẩn có giá trị trung bình bằng 0

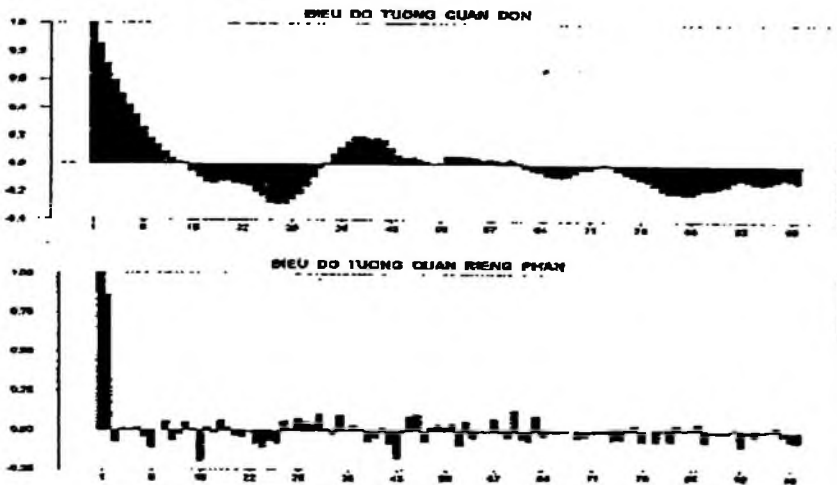
Khoảng tin cậy  $\Delta$  của hệ số  $\hat{P}_k$  như sau:  $\alpha$  với  $n$  là số lượng quan trắc  $\alpha$

Nếu hệ số  $\hat{P}_k$  tính được nằm ngoài khoảng trên, ta kết luận  $\hat{P}_k$  khác 0 với rủi ro  $\alpha\%$  (thường ta lấy  $\alpha = 5\%$ ).

#### 4.3.2. Tham số thống kê của Box-Pierce và Ljung-Box

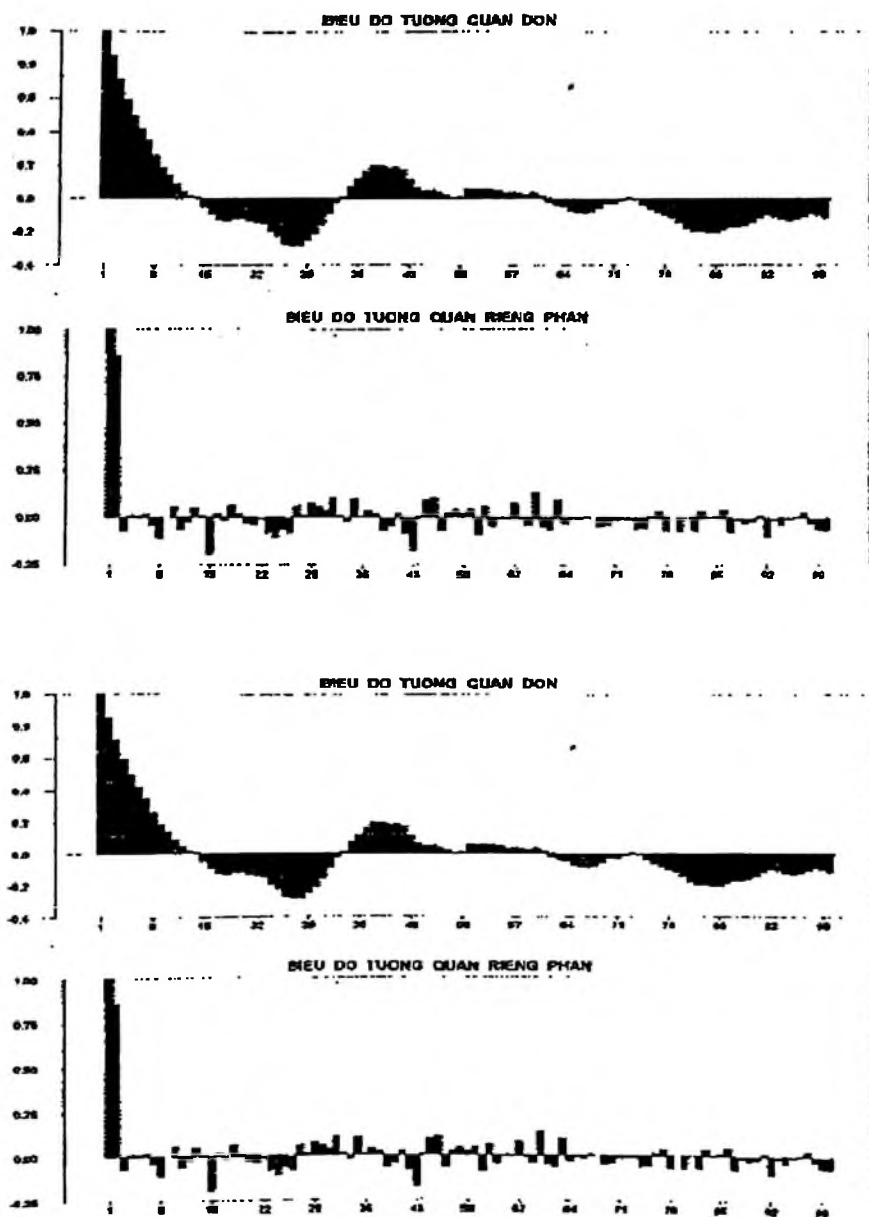
Kiểm định của Box-Pierce cho phép nhận biết đó là nhiễu trắng hay không. Chúng ta phải kiểm định  $\text{Cov}(y_t, y_{t-k}) = 0$  và  $\hat{P}_k = 0$  với  $k > q$ . Một quá trình nhiễu trắng bắt buộc phải có:  $\hat{P}_1 = \hat{P}_2 = \hat{P}_3 = \dots = \hat{P}_h$  chúng ta có thể kiểm định riêng lẻ các giá trị của  $p$ , tuy nhiên thường ta hay sử dụng giá trị thống kê Q định nghĩa bởi Box-Pierce như sau:  $Q = n$  với  $h$  số lượng của sự trễ,  $\hat{P}_k$  giá trị tự tương quan kinh nghiệm bậc  $k$  và  $n$  chỉ số quan trắc. Giá trị thống kê Q tuân theo gần như một phân phối  $\chi^2$  có bậc tự do  $h$ . Với mức độ rủi ro  $\alpha\%$  và bậc tự do  $h$ , ta có giá trị cho từ bảng tra. Nếu  $Q > \chi^2_{\alpha, h}$  ta sẽ chấp nhận giả thiết  $H_1$ : đó không phải là một nhiễu trắng. Và ngược lại, ta sẽ kết luận đó là một nhiễu trắng.

Đồ thị sau đây cho ta thấy biến đổi của một nhiễu trắng Hình.4.1



Hình 4.1

Biểu đồ tương quan đơn và biểu đồ tương quan riêng phần tương ứng của chuỗi này như sau:



Hình 4.2

Trong thực hành để khảo sát đó là một nhiễu trắng hay không ta sẽ sử dụng các kiểm định Bartleu và Quenouille. Kiểm định liên quan đến độ lớn của các giá trị hệ số tương quan và tương quan riêng phần.

Khi ta thấy cường độ của nhiễu toàn bộ nằm trong giới hạn cho phép, ta kết luận đó là một nhiễu trắng. Đối với trường hợp hình trên, ta nhận thấy ở kiểm định Quenouille còn có giá trị vượt quá giới hạn, đây chưa phải là một nhiễu trắng hoàn toàn.

#### 4.4. Mô hình AR(p) (Auto Regression)

Trong một quá trình tự hồi quy bậc p, số liệu quan trắc tại thời điểm hiện tại  $y_t$  được tạo ra bởi một tổng trung bình có trọng số của các giá trị quan trắc trong quá khứ tính cho đến giá trị quan trắc quá khứ thứ p. Công thức định nghĩa như sau:

$$AR(1): y_t = q_1 \times y_{t-1} + e_t$$

$$AR(2): y_t = q_1 \times y_{t-1} + q_2 \times y_{t-2} + e_t$$

---


$$AR(P): y_t = q_1 \times y_{t-1} + q_2 \times y_{t-2} + \dots + q_p \times y_{t-p} + e_t$$

Trong đó:  $q_1; q_2; \dots; q_p$  là các thông số cần phải xác định;  $e_t$  là một nhiễu trắng ngẫu nhiên có dạng Gaussien. Chúng ta cũng có thể thêm vào quá trình này một hằng số mà nó vẫn không ảnh hưởng đến tính chất ngẫu nhiên của chuỗi. Phương trình trên có thể viết dưới dạng đơn giản hơn nhờ vào định nghĩa toán tử lệch pha D như sau:

$$(1 - q_1 \times D - q_2 \times D^2 - \dots - q_p \times D^p) \times y_t = e_t$$

*Tính chất:*



- Người ta đã chứng minh biểu đồ tương quan đơn của một quá trình AR(p) được mô tả bởi một cấp số nhân có công bội nhỏ hơn 1 (chuỗi giảm) có dạng:

$$\hat{P}_k = \hat{P}_{-k}$$

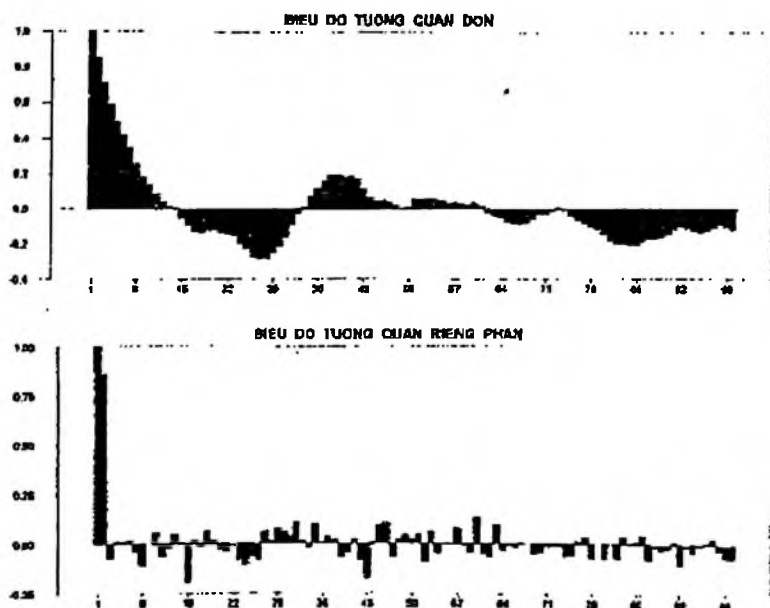
- Biểu đồ tương quan riêng phần chỉ có p số hạng đầu tiên là khác 0.

Các ví dụ sau đây cho phép chúng ta nhận biết mô hình dạng AR dựa trên phân tích biểu đồ tương quan đơn là tương quan riêng phần. Xét một mô hình AR(1) có dạng:

$$y_t = 1 + 0,9y_{t-1} + e_t$$

với  $e_t$  là giá trị thặng dư.

Các biểu đồ tương quan của mô hình trên có dạng sau:



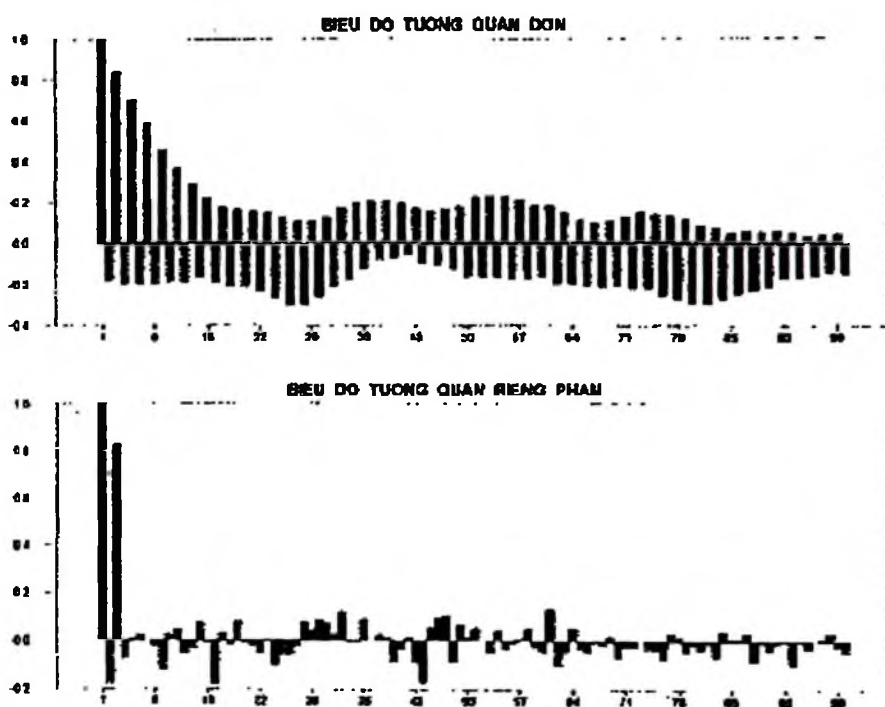
Hình 4.3

Ta thấy giá trị đầu tiên của biểu đồ tương quan riêng phần rất lớn so với các giá trị còn lại và biểu đồ tương quan đơn có giá trị giảm dần. Đó là biểu thị đặc thù cho phép chúng ta nhận dạng đó là một mô hình AR(L).

Xét một mô hình AR(2) có dạng:

$$y_t = 0,9y_{t-2} + 1 + e_t$$

Các biểu đồ tương quan của mô hình trên có dạng sau:



Hình 4.4

So với trường hợp trước ta thấy có sự khác nhau. Thay vì giá trị thứ 1 như ví dụ trước, trường hợp này ta thấy giá trị thứ 2 trong biểu đồ tương quan riêng phần lớn trội hơn hẳn so với các giá trị còn lại. Trong khi đó tính chất của biểu đồ tương quan đơn cũng giống như

trước. Điều này cho phép ta biết đây là một mô hình AR(2). Ta cũng lưu ý thêm với số hạng AR(1) là không đáng kể.

#### 4.5. Mô hình MA(q) (Moving Average)

Trong một quá trình trung bình động bậc q, số liệu quan trắc tại thời điểm hiện tại  $y_t$  được tính bởi tổng trung bình có trọng số giá trị của các nhiễu ngẫu nhiên cho đến nhiễu thứ q. Công thức định nghĩa như sau:

$$\text{MA}(1): \quad y_t = e_t - a_1 * e_{t-1}$$

$$\text{MA}(2): \quad y_t = e_t - a_1 * e_{t-1} - a_2 * e_{t-2}$$

---


$$\text{MA}(q): \quad y_t = e_t - a_1 * e_{t-1} - a_2 * e_{t-2} - \dots - a_q * e_{t-q}$$

Trong đó:  $a_1, a_2, \dots, a_p$  là các thông số cần phải xác định;  $e_t$  là một nhiễu trắng ngẫu nhiên có dạng Gaussien. Phương trình trên có thể viết dưới dạng đơn giản hơn nhờ vào định nghĩa một toán tử lệch pha D như sau:

$$(1 - a_1 D - a_2 D^2 - \dots - a_q D^q) e_t = y_t$$

Trong quá trình dạng này cũng như tất cả các mô hình tự hồi quy, các nhiễu ngẫu nhiên được giả thiết là được tạo ra bởi một “nhiễu trắng” Chúng ta có thể hiểu quá trình trung bình động là một chuỗi thời gian dao động ngẫu nhiên chung quanh giá trị trung bình của chúng.

*Tính chất:*

- Chuỗi trung bình động bậc 1 chính là một quá trình tự hồi quy bậc p vô hạn.

- Biểu đồ tương quan đơn của một quá trình trung bình động bậc q, MA(q), được xác định bởi:

$$\hat{P}_k \neq 0 \text{ khi } k \leq q$$

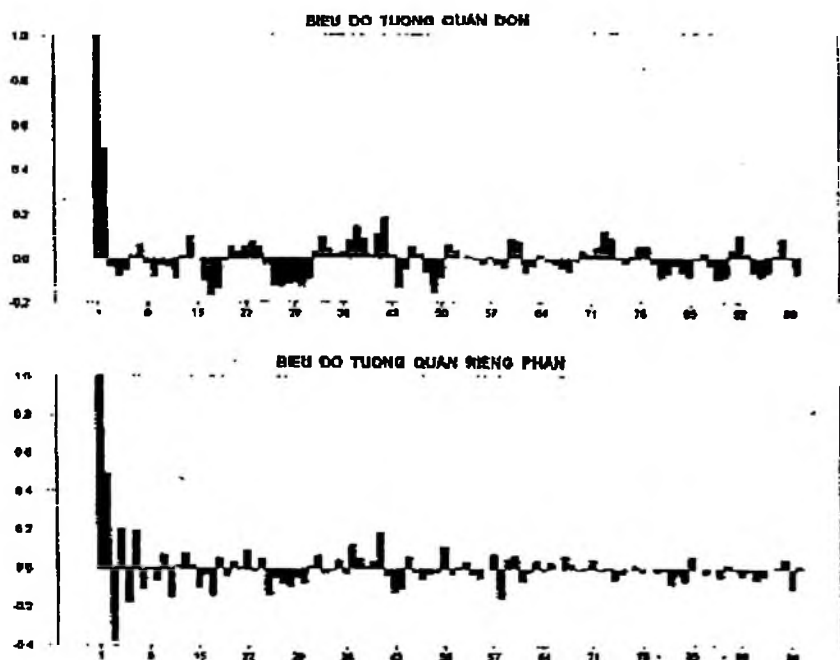
$$\hat{P}_k = 0 \text{ khi } k > q$$

Điều này có nghĩa là chỉ có  $q$  số hạng đầu tiên của biểu đồ tương quan là khác 0. Đối với biểu đồ tương quan riêng phần sẽ được mô tả bởi một chuỗi cấp số giảm theo hướng các chậm pha trong quá khứ. Các ví dụ sau đây cho phép chúng ta nhận biết theo kinh nghiệm, hình dạng MA dựa trên cơ sở phân tích biểu đồ tương quan đơn và tương quan riêng phần. Xét một mô hình MA(L) có dạng:

$$y_t = 5 + e_t + 0,9e_{t-1}$$

với  $e_t$  là giá trị trắng dư ở thời điểm  $t$

Các biểu đồ tương quan của mô hình trên có dạng sau:



Hình 4.5

Ta thấy giá trị đầu tiên của biểu đồ tương quan đơn vượt trội so với các giá trị còn lại và biểu đồ tương quan riêng phần giảm dần dần. Đó là dạng đặc thù của một mô hình MA có bậc là 1.

Xét trường hợp cho một mô hình MA(2) có dạng:

$$Y_t = 5 + e_t + 1 \cdot e_{t-2}$$

Các biểu đồ tương quan của mô hình trên có dạng sau:

Trong trường hợp này, thay vì giá trị đầu tiên trên biểu đồ tương quan có giá trị lớn trội như trước, ta thấy giá trị thứ 2 trên biểu đồ này lớn trội hơn so với các giá trị còn lại và giá trị của biểu đồ tương quan riêng phần giảm dần dần, đó là biểu thị đặc thù của một mô hình MA(2).

#### 4.6. Mô hình ARMA(p, q)

Mô hình ARMA(p, q) là một quá trình được tạo ra bởi từ tổ hợp giữa các giá trị của chuỗi trong quá khứ và các giá trị của nhiễu trong quá khứ. Nó được xác định bởi phương trình sau đây:

$$\begin{aligned} & (1 - \theta_1 \times D - \theta_2 \times D^2 - \dots - \theta_p \times D^p) \times y_t \times E_t \\ & = (1 - \alpha_1 \times D^1 - \alpha^2 \times D^2 - \dots - \alpha_p \times D^p) \times E_t \end{aligned}$$

Ta có thể nói đây là một mô hình có được từ sự tổng hợp của hai loại mô hình AR và MA.

*Tính chất:*

$$ARMA(1, 0) = AR(1) ; ARMA(0, 1) = MA(1)$$

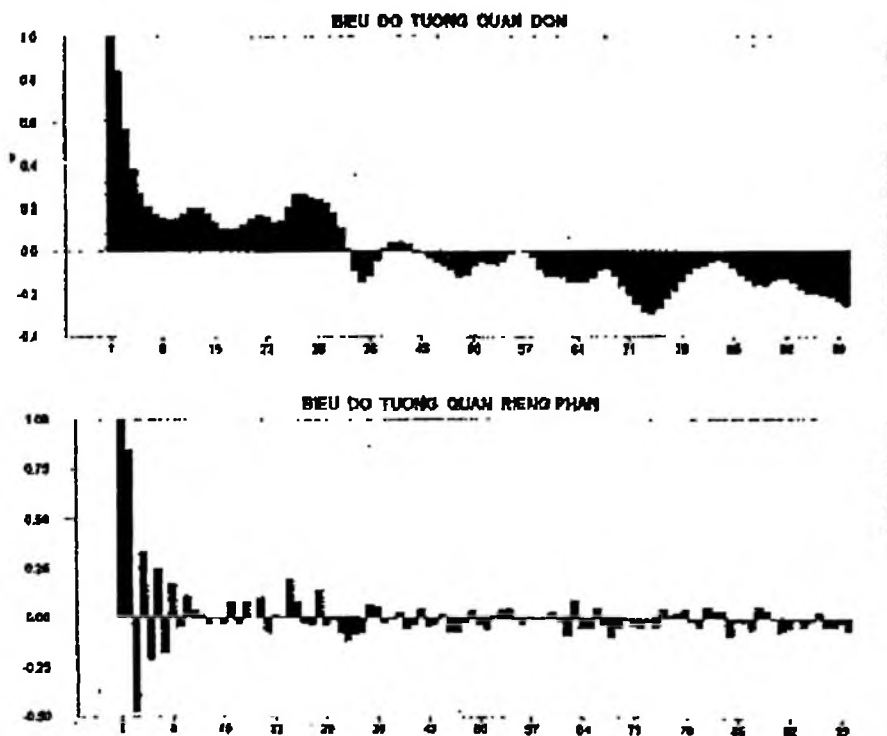
Ta chú ý trong trường hợp này, biểu đồ tương quan đơn và biểu đồ tương quan riêng phần sẽ phức tạp hơn so với hai trường hợp trên. Do vậy, chúng ta phải lưu ý khi xác định các thông số p, q của mô hình ARMA từ các biểu đồ này.

### Ví dụ 5

Xét mô hình ARMA(1, 1) sau đây:

$$y = 5 + 0,8y_{t-1} + 1,1$$

Các biểu đồ tương quan của mô hình trên có dạng sau:



Hình 4.6

Với biểu đồ trên ta thấy đây là một sự pha lẫn giữa hai loại mô hình AR và MA. Ta thấy đều có giá trị đầu tiên vượt trội trong các biểu đồ tương quan. Cường độ trong các biểu đồ cũng tắt dần.

Dự đoán bậc của mô hình đòi hỏi phải có một kinh nghiệm nhất định.

#### 4.7. Mô hình ARMA mở rộng: ARIMA, SARIMA

Trong trường hợp chuỗi quan trắc có xu thế không ổn định (có xu thế tăng hoặc giảm theo thời gian), ta định nghĩa một mô hình có dạng ARMA(p, d, q) với d là bậc của đường xu thế. Nói một cách khác, d biểu thị cho số lần lấy “sai biệt” cần thiết lên chuỗi quan trắc để ta có thể nhận được một chuỗi nghiên cứu có tính ổn định theo xu thế. Ví dụ trong trường hợp chuỗi có xu thế tuyến tính ta có d=1; trong trường hợp đường xu thế là một hàm bậc 2 ta có d=2.

Thật vậy, giả sử chuỗi có một xu thế tuyến tính biểu thị bởi phương trình sau đây:

$$y = a + bt$$

Định nghĩa sai biệt bậc 1  $D(y_t)$  ta có:

$$D(y_t) = y_t - y_{t-1} = (a + bt) - (a + b[t-1]) = b = cte$$

Ta thấy chuỗi sai biệt bậc 1 có xu thế ổn định.

Trong trường hợp có xu thế bậc 2 phương trình có dạng:

$$Y_t = a + bt + ct^2$$

Tính sai biệt bậc 1 ta có:

$$D(y_t) = y_t - y_{t-1} = (a + bt + ct^2) - (a + b[t-1] + c[t-1]^2) = b - c + 2tc$$

Ta thấy chuỗi  $Dy_t$  có xu thế bậc 1. Để có xu thế ổn định ta chỉ cần tính thêm một lần nữa cho sự khác biệt như trường hợp ta đã có trong trường hợp xu thế là tuyến tính ở trên.

Như vậy, ta có hai lần lấy sai biệt cho trường hợp bậc 2 này để chuỗi quan trắc trở nên ổn định về xu thế. Tóm lại ta có thể viết chuỗi  $(1-D)^d y_t$  là một ARMA(p,q) khi  $y_t$  là một ARIMA(p,d,q); với D được định nghĩa là toán tử sai biệt:

$$D(y_t) = y_t - y_{t-1}$$

Mô hình SARIMA cho phép giải quyết vấn đề sai biệt liên quan đến biến đổi mùa. Sự biến đổi được định nghĩa như sau:

$$(1 - Ds) \times y_t = y_t - y_{t-s}$$

với  $s$  biểu thị tính chu kỳ của số liệu ( $s=4$  cho một chuỗi biến đổi theo quý,  $s=12$  cho chuỗi biến đổi theo tháng).

*Chú ý: Chúng ta chỉ áp dụng mô hình ARMA để nghiên cứu cho các chuỗi không có xu thế.*

#### 4.8. Phương pháp Box - Jenkins

Dưới đây nghiên cứu một cách có hệ thống các dạng khác nhau của chuỗi thời gian dựa vào các tính chất của nó. Mục tiêu là tìm trong số tất cả các mô hình ARIMA (AR: tự hồi quy, MA: trung bình động, I: thông số cho biết bậc cần thiết để có thể tạo một chuỗi ổn định) một mô hình thích hợp nhất với số liệu của hiện tượng nghiên cứu.

Phương pháp bao gồm ba bước chính sau đây:

*Bước 1: Tìm các mô hình thích hợp nhất*

Đây là bước quan trọng và khó nhất. Nó cho phép nhận biết được trong họ tất cả các mô hình ARIMA mô hình nào là có khả năng thích hợp nhất. Phương pháp dựa vào nghiên cứu các biểu đồ tương quan đơn và các biểu đồ tương quan riêng phần. Một vài nguyên tắc sau đây cho phép tìm các thông số  $p, d, q$  của mô hình ARIMA.

\* Khử tính chu kỳ

Để đơn giản trong trường hợp chuỗi nghiên cứu có chứa yếu tố biến đổi có tính chu kỳ ta nên “khử” yếu tố này trước khi đi vào các xử lý thống kê nhằm đơn giản hóa cho các bước tính sau.



**\* Khảo sát và xác định bậc của xu thế nếu có**

Trong trường hợp biểu đồ tương quan đơn giảm chậm hoặc hoàn toàn không giảm, chuỗi có chứa một xu thế. Trong trường hợp này ta sẽ loại tính xu thế đó nhờ vào áp dụng của toán tử sai biệt lên chuỗi. Trong thực tế, ta có thể gặp trường hợp  $d=1$  hoặc 2. Giá trị thích hợp của  $d$  sẽ cho ta một biểu đồ tương quan đơn có xu thế giảm nhanh.

**\* Xác định  $p, q$  của mô hình ARMA nhờ vào biểu đồ tương quan**

- Nếu biểu đồ tương quan đơn chỉ có  $q$  giá trị đầu tiên là khác 0 ( $q=3$  là lớn nhất) và các giá trị của biểu đồ tương quan riêng phần giảm từ từ, ta có thể tiên đoán có một MA( $q$ ).

- Nếu biểu đồ tương quan riêng phần chỉ có  $p$  giá trị đầu tiên là khác 0 ( $p=3$  là lớn nhất) và các giá trị của biểu đồ tương quan đơn giảm từ từ ta có thể tiên đoán có một AR( $p$ ).

- Nếu biểu đồ tương quan đơn và biểu đồ tương quan riêng phần không có sự cắt ngắn như hai trường hợp trên, ta sẽ có một quá trình ARMA và các thông số của nó tùy thuộc vào dạng cụ thể của các biểu đồ tương quan.

Trong thực hành, phương pháp phân tích đồ thị chỉ cho ta tìm được  $p, q$  trong các trường hợp đơn giản mà thôi. Trong trường hợp tổng quát, ta có thể áp dụng các tiêu chuẩn sau đây để xác định các thông số  $p, q$  trong một mô hình ARMA. Thực chất chung của các tiêu chuẩn này là dựa vào sự khảo sát các giá trị liên quan đến phương sai của chuỗi sai số cho bởi mô hình với thông số đề nghị.

Có ba tiêu chuẩn thông dụng được sử dụng như sau:

*Tiêu chuẩn Akaike:*

$$Akaike = \text{Log}(\%rss) + 2 \frac{p+q}{\%nobs}$$

*Tiêu chuẩn BIC:*

$$BIC = \text{Log}(\%rss) + (p + q) \frac{\text{Log}(\%nobs)}{\%nobs}$$

*Tiêu chuẩn HQ:*

$$HQ = \text{Log}(\%rss) + 2(p + q) \frac{\text{Log}(\text{Log}(\%nobs))}{\%nobs} 270$$

Trong đó: %rss: tổng các thặng dư bình phương của mô hình đề nghị; %nobs: số lượng quan trắc.

Trong trường hợp lý tưởng, giá trị chọn của p, q tương ứng với trường hợp cho ta các giá trị Akaike, BIC, HQ cực tiểu. Trong áp dụng ta có thể có trường hợp ở đó giá trị p, q đề nghị không làm cho ba tiêu chuẩn này đồng thời cực tiểu. Tuy vậy thường các tiêu chuẩn này cho giá trị p,q tối ưu không khác nhau lớn. Trong trường hợp này ta sẽ khảo sát từng tổ hợp (p, q) cụ thể để quyết định chọn mô hình hợp lý nhất.

*Bước 2: Ước lượng các hệ số của mô hình*

Trong trường hợp mô hình AR(p), tác giả áp dụng phương pháp bình phương tối thiểu hay sử dụng quan hệ giữa tính tự tương quan và các hệ số của mô hình (phương trình Yule Walker). Ước lượng các hệ số cho mô hình MA(Q) tương đối phức tạp hơn. Các tác giả đề nghị sử dụng một phương pháp lặp dưới dạng quét mà chúng ta có thể hiểu một cách đơn giản như sau.

Giả sử ta có một mô hình ARMA(2,2) xác định bởi:

$$(1 - q_1 D - q_2 D^2)y_t = (1 - a_1 D - a_2 D^2)e_t$$

và chúng ta có thể viết dưới dạng:

$$y_t = \frac{1}{(1 - \theta_1 D - \theta_2 D^2)} (1 - \alpha_1 D - \alpha_2 D^2) \times e_t$$

Ta đặt:

$$V_t = \frac{1}{(1 - \theta_1 D - \theta_2 D^2)} \times e_t \rightarrow V_t - \theta_1 V_{t-1} + V_2 - \theta_2 V_{t-2} = e_t$$

Do đó:

$$y_t = V_t - \theta_1 V_{t-1} + V_2 - \theta_2 V_{t-2} \rightarrow V_t = y_t + \alpha_1 V_{t-1} + \alpha_2 V_{t-2}$$

Từ đó, chúng ta có thể khởi đầu bằng cách tính quét với hai khoảng giá trị chấp nhận được cho  $a_1$  và  $a_2$  và với một gia số cho trước. Tiếp theo, cho mỗi cặp giá trị của  $a_1$  và  $a_2$  ta đặt  $n_0 = 0$ ,  $n_1 = 0$  và chúng ta sẽ ước lượng giá trị của  $v_t$  theo các bước sau:

$$n_2 = y_2$$

$$n_3 = y_3 + a_2 n_2$$

$$n_4 = y_4 + a_1 n_1 + a_2 n_2$$

....

sau khi tính tất cả các giá trị của  $n_t$  ta sẽ ước lượng các thông số  $q_1$  và  $q_2$  bởi phương pháp bình phương tối thiểu áp dụng vào phương trình sau:

$$n_t = q_1 n_{t-1} + q_2 n_{t-2} + e_t$$

và chúng ta sẽ lấy giá trị  $a_1$ ,  $a_2$  sao cho các tổng bình phương của các thặng dư từ phương trình hồi quy trên tối thiểu. Chú ý phương pháp này chỉ có giá trị trong trường hợp số lượng các thông số cần xác định

không nhiều lắm. Ngoài phương pháp bình phương tối thiểu ta còn có thể áp dụng phương pháp cực đại hóa các hàm tương thích.

### *Bước 3: Kiểm tra giá trị của mô hình và dự báo*

Sau khi các thông số của mô hình được xác định, chúng ta sẽ kiểm định các kết quả của ước lượng này.

*Các hệ số của mô hình phải khác 0 (kiểm định Student cổ điển).*

Nếu có một hay nhiều hệ số không thỏa mãn, ta sẽ loại bỏ nó ra khỏi mô hình AR hoặc MA đang xét.

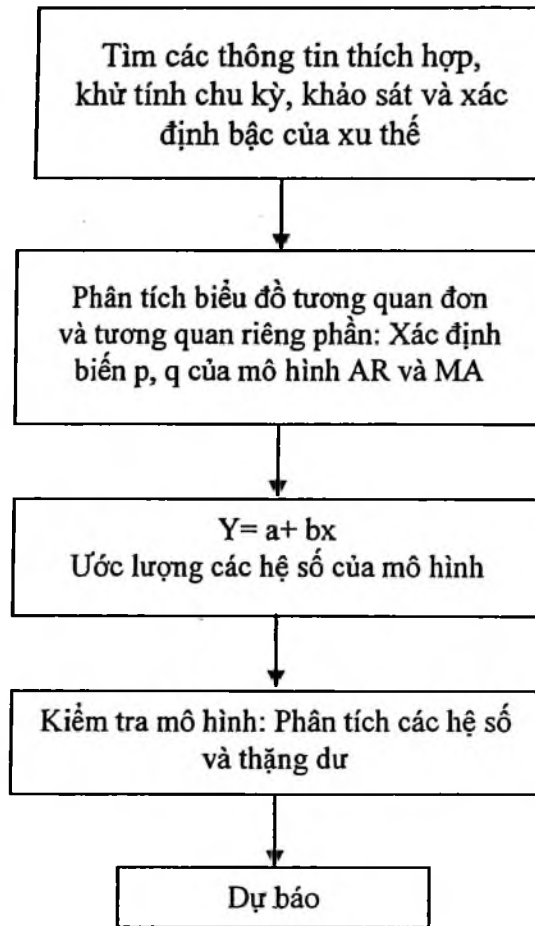
*Phân tích các giá trị thặng dư được thực hiện từ hai tiêu chuẩn sau:*

- Giá trị trung bình số học triệt tiêu, trong trường hợp ngược lại ta nên thêm một hằng số vào mô hình.

- Chuỗi giá trị thặng dư là một nhiễu trắng. Các giá trị thống kê của Box-Pierce và của Ljung-Box cho phép kiểm định tính chất này. Nếu nó không phải là một nhiễu trắng, ta kết luận mô hình là không hoàn chỉnh và ta phải thêm vào mô hình các bậc bổ sung cần thiết.

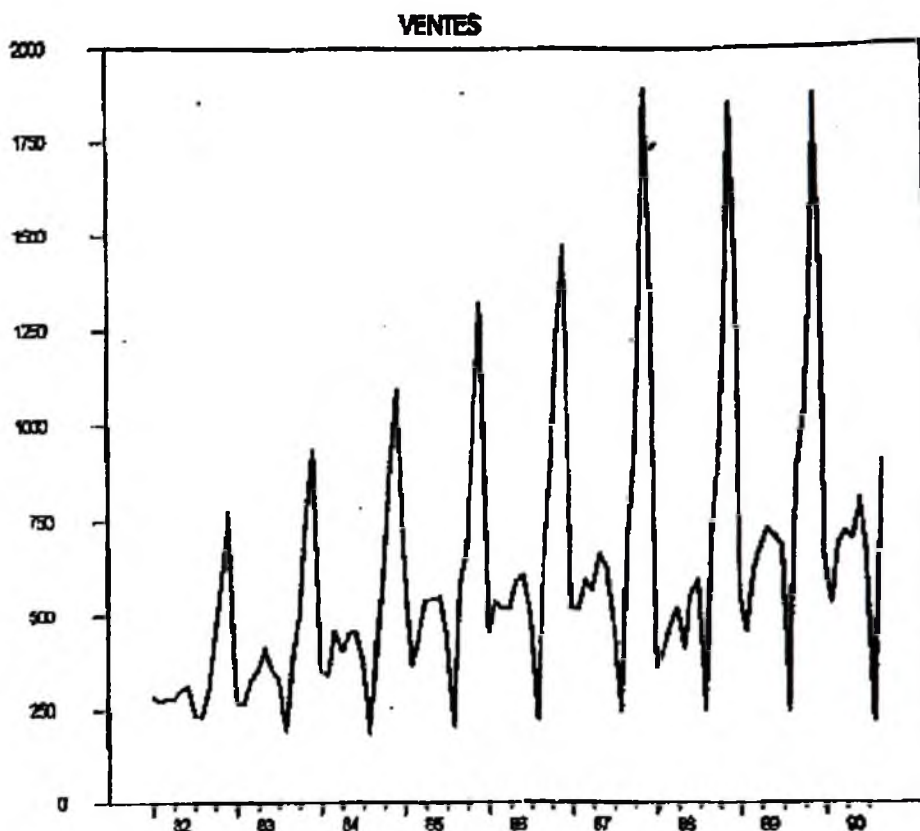
- Bước kiểm định mô hình rất quan trọng và có thể ta phải trở lại bước 1 nếu mô hình đề nghị không thích hợp. Một khi mô hình đã được kiểm định, ta có thể tiến hành dự báo giới hạn trong một vài chu kỳ. Phân tích chuỗi thời gian với mô hình SARIMA chỉ cho phép tiến hành các dự báo ngắn hạn. Nó không cho phép một dự báo trung hạn và dài hạn với độ chính xác cần có vì biên độ của sai số gia tăng rất nhanh trong trường hợp này.

Chúng ta có thể tóm tắt các bước cơ bản của phương pháp Box-Jenkins như sau:



*Ví dụ:*

Áp dụng phương pháp Box-Jenkins, doanh thu của một công ty trong chu kỳ 01/82 đến 09/90 được trình bày bởi đồ thị sau đây:

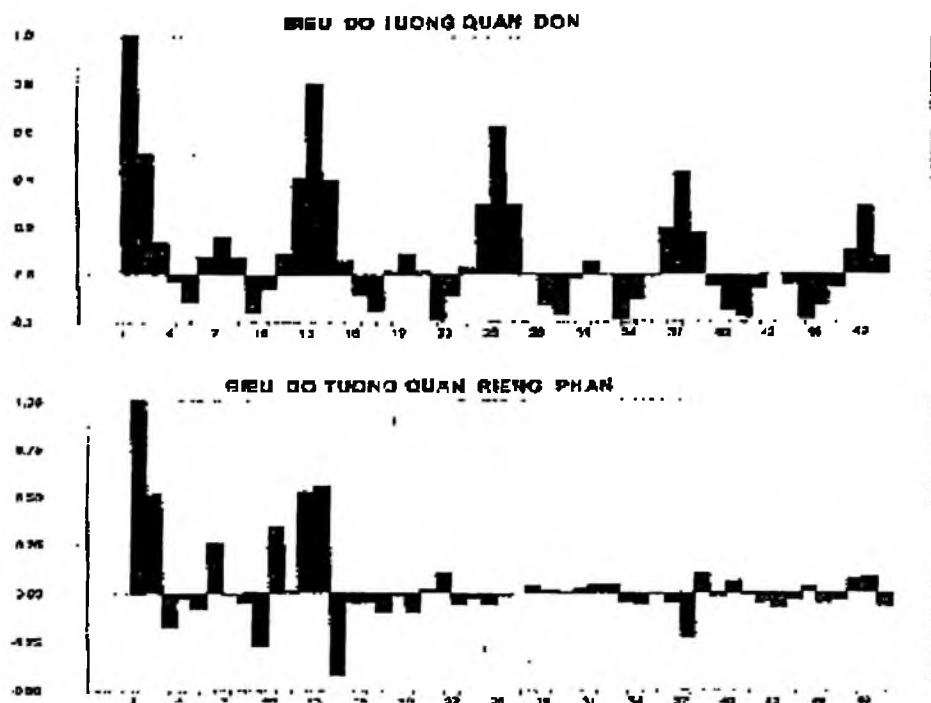


Hình 4.7

Hãy phân tích chuỗi trên bằng phương pháp Box-Jenkins và dự báo cho doanh số trong 6 tháng tiếp theo (10/90 - 3/91).

*Hướng dẫn (kết quả tính toán được thực hiện với logiciel RATS)*

Biểu đồ tương quan đơn và biểu đồ tương quan riêng phần của chuỗi trên như sau:



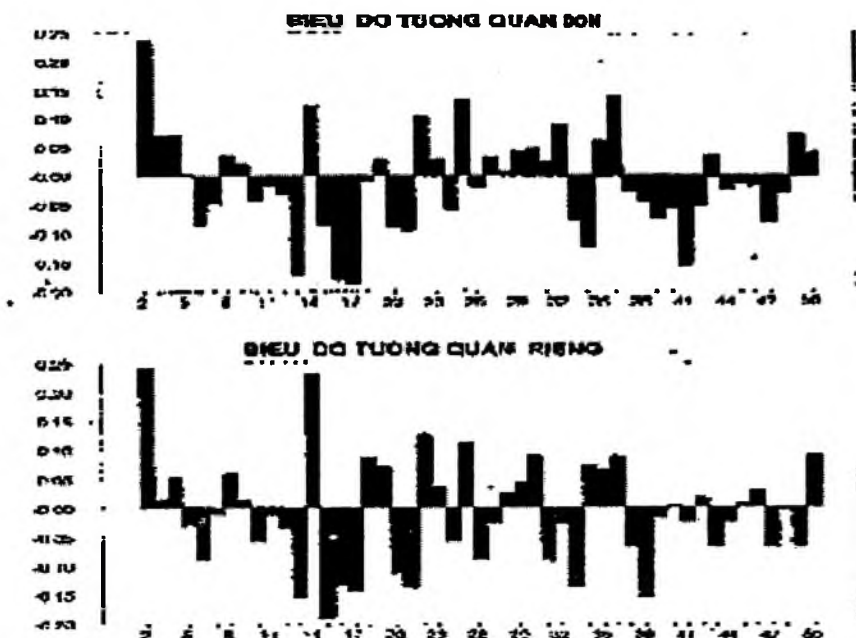
Hình 4.8

Ta thấy trên các biểu đồ tương quan xuất hiện 1 “pic” rất rõ khi  $k=12$ . Nhận xét này cho ta kết luận số liệu có tính chu kỳ ( $T=12$  tháng). Để khử tính chu kỳ trong chuỗi, ta sẽ định nghĩa chuỗi  $Y_t$  nhờ vào một biến đổi như sau:

$$Y_t = y_t - y_{t-12}; \forall_t$$

Để khử xu thế ta áp dụng biến đổi sau:  $D(Y_t) = Y_t - Y_{t-1}$

Biểu đồ tương quan của  $D(Y_t)$  như sau:



Hình 4.9

Ta thấy giá trị đầu tiên của biểu đồ tương quan đơn lớn hơn hẳn sơ với các giá trị tiếp theo, trong khi đó giá trị của biểu đồ tương quan riêng phần giảm từ từ; ta có thể dự đoán đây là một mô hình có dạng MA(1). Tóm lại mô hình đề nghị cho chuỗi số liệu trên như sau SARIMA(0,1,1) với  $s=12$ . Kết quả cho từ logiciel RATS như sau:

*Biến nghiên cứu VENTE -ước lượng bởi Box-Jenkins*

*Số lần lặp 21*

*Chuỗi số liệu 83:02 đến 90:09*

*Số quan trắc hiệu dụng 92 Bộc tự do 90*

*Hệ số xác định  $R^2$  0,921215, Hệ số xác định hiệu chỉnh 0,920340*



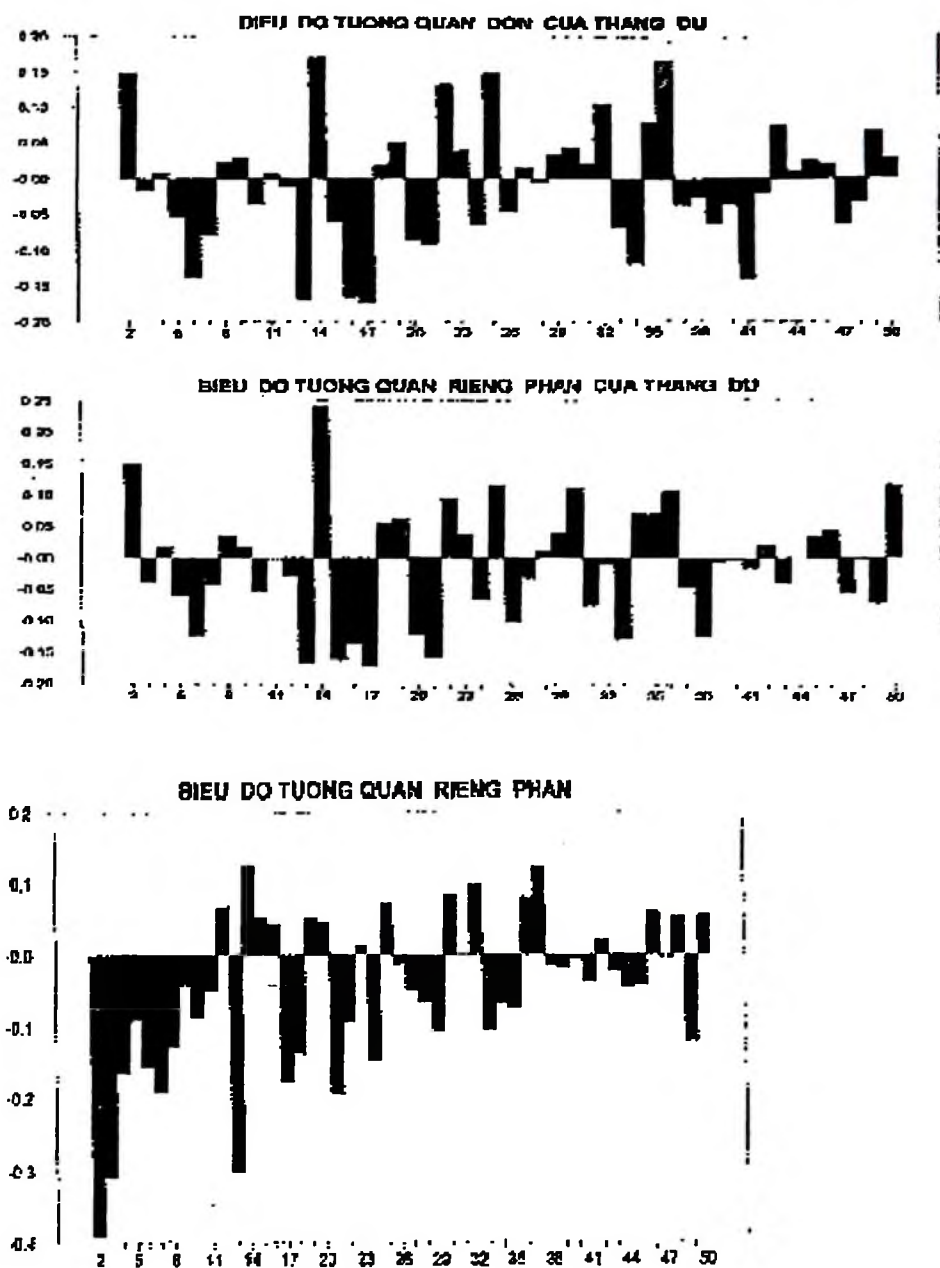
<i>Giá trị biến nghiên cứu</i>	<i>646. 71640217</i>
<i>Độ lệch chuẩn của biến nghiên cứu</i>	<i>365.92740004</i>
<i>Sai số chuẩn hóa của ước lượng</i>	<i>103.28000630</i>
<i>Tổng các thặng dư bình phương</i>	<i>960008.37314</i>
<i>Giá trị thống kê Durbin-Watson</i>	<i>1.751202</i>
<i>Giá trị thống kê của Ljung-Box <math>Q(23-2)</math></i>	<i>29.883511</i>
<i>a tương ứng của <math>Q</math></i>	<i>0.09435394</i>

<b>Biến</b>	<b>Hệ số</b>	<b>Độ lệch chuẩn</b>	<b>T-student</b>	<b>a</b>
1.AR(12)	1.058169	0.032803	32.25804	0.000
2.Ma(1)	0.820817	0.060968	-13.46307	0.000

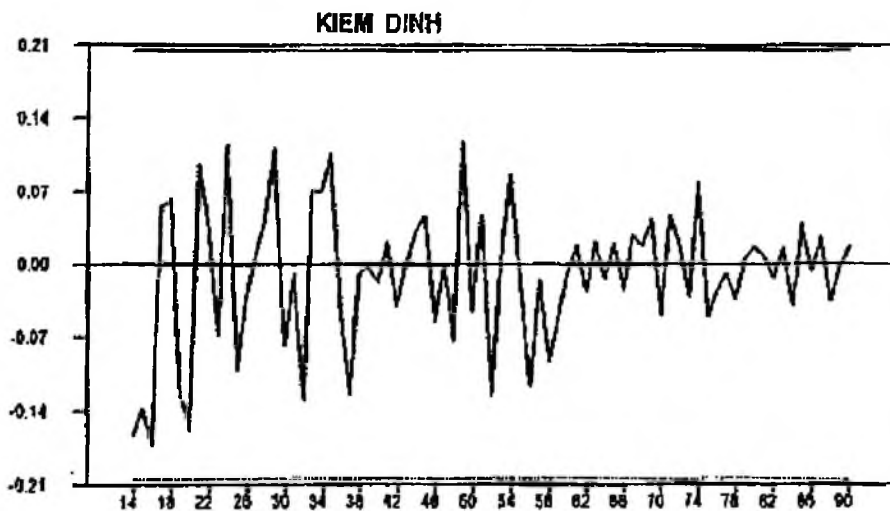
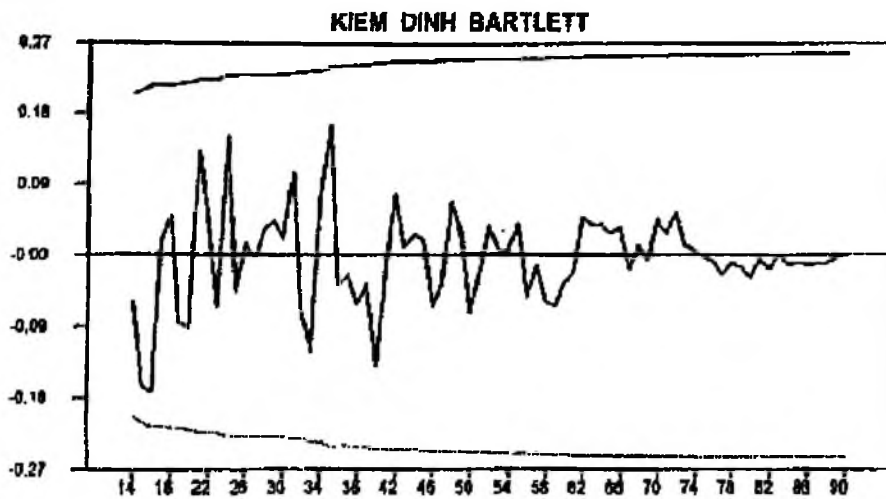
Biểu đồ tương quan đơn và biểu đồ tương quan riêng phần của thặng dư cho bởi mô hình được chọn từ phương pháp Box-Jenkins như sau:

Để đánh giá chất lượng của mô hình ta phải kiểm tra xem giá trị thặng dư trên có phải là một nhiễu trắng hay không. Sau đây là kết quả của kiểm định Bartlett và Quenouille:

Ta thấy cường độ của hệ số tương quan đơn và tương quan riêng phần hoàn toàn nằm trong giới hạn cho phép trong cả hai loại kiểm định. Do đó, chuỗi giá trị thặng dư cho bởi mô hình chọn là một nhiễu trắng như mong đợi.



Hình 4.10



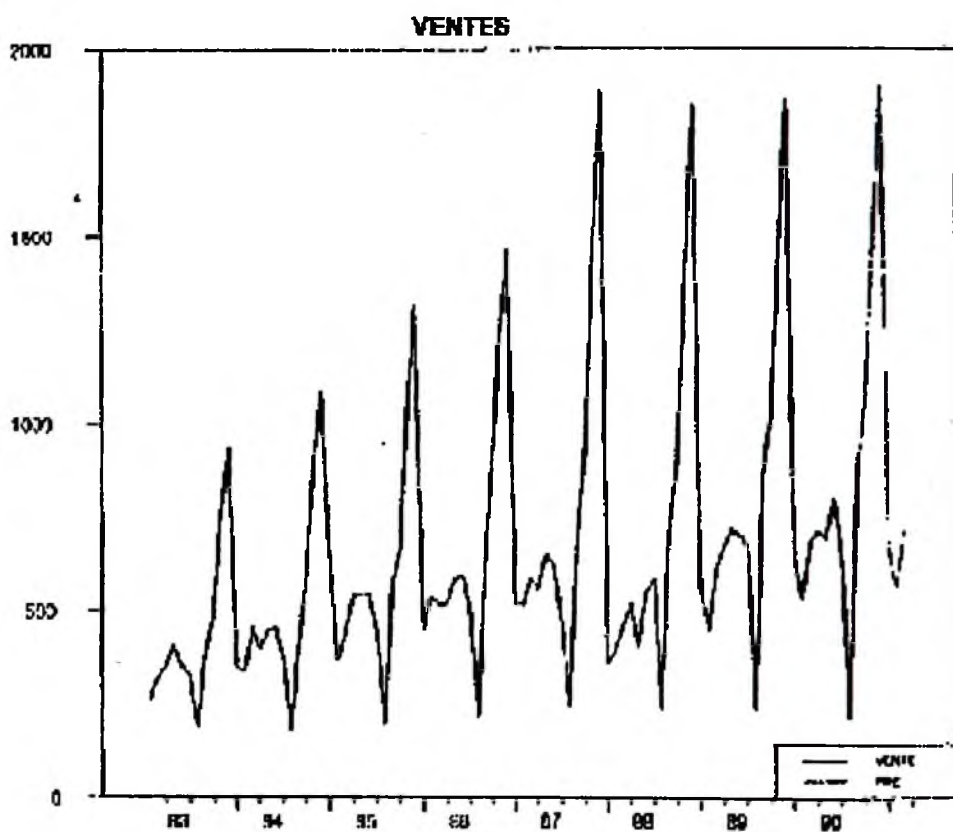
Hình 4.11

Dự báo ngắn hạn:

Tiến hành dự báo ngắn hạn về doanh số của công ty cho bởi mô hình Box-Jenkins được trình bày trong bảng sau:

Thời gian	90:10	90:11	90:12	91:01	91:02	91:03	91:04
Dự báo	1055.3	1480.7	1901.4	676.1	561.8	561.8	714.6

Đồ thị sau biểu diễn tổng hợp giữa doanh thu trong quá khứ và dự báo ngắn hạn của công ty như sau:



Hình 4.12

## *Chương 5*

# DÃY SỐ THỜI GIAN

### 5.1. Khái niệm

Một lượng của hiện tượng thường xuyên biến động qua thời gian. Trong thống kê, để nghiên cứu sự biến động này ta thường dựa vào dãy số thời gian.

Dãy số thời gian là dãy số các trị số của chỉ tiêu thống kê được sắp xếp theo thứ tự thời gian.

*Ví dụ:* Có số liệu về doanh thu của bưu điện X từ năm 1999 - 2003 như sau:

**Bảng 5.1**

*ĐVT: tỷ đồng*

Năm	1999	2000	2001	2002	2003
<b>Doanh thu</b>	23,9	28,1	37,3	47,2	67,4

Ví dụ trên đây là một dãy số thời gian về chỉ tiêu doanh thu của đơn vị bưu điện này từ năm 1999- 2003. Qua dãy số thời gian có thể nghiên cứu các đặc điểm về sự biến động của hiện tượng, vạch rõ xu hướng và tính quy luật của sự phát triển, đồng thời để dự đoán các mức độ của hiện tượng trong tương lai.

Mỗi dãy số thời gian có hai thành phần:

- Thời gian: có thể là ngày, tuần, tháng, quý, năm,... Độ dài giữa hai thời gian liên nhau được gọi là khoảng cách thời gian.

- Chỉ tiêu về hiện tượng nghiên cứu: Chỉ tiêu này có thể là số tuyệt đối, số tương đối, số bình quân. Trị số của chỉ tiêu còn gọi là mức độ của dãy số.

Phân loại dãy số thời gian:

Căn cứ vào tính chất thời gian của dãy số, có thể phân biệt thành hai loại:

1. Dãy số thời kỳ: Dãy số biểu hiện mặt lượng của hiện tượng qua từng thời kỳ nhất định
2. Dãy số thời điểm: là loại dãy số biểu hiện mặt lượng của hiện tượng qua các thời điểm nhất định. Dãy số này còn được phân biệt thành 2 loại:

- Dãy số thời điểm có khoảng cách thời gian đều nhau:

*Ví dụ:* Có giá trị và hàng hóa tồn kho của công ty X vào các ngày đầu tháng 1, 2, 3, 4 năm 1995 như sau:

**Bảng 5.2**

Ngày	1-1	2-1	3-1	4-1
Giá trị hàng tồn kho (triệu đồng)	356	364	370	352

- Dãy số thời điểm có khoảng cách thời gian không đều:

*Ví dụ:* Có số liệu về số dư tiền vay ngân hàng của công ty Y như sau:

**Bảng 5.3**

Ngày (thời điểm)	1-1	20-1	15-2	10-3
Số dư tiền vay (triệu đồng)	400	600	500	700

Các yếu tố ảnh hưởng đến biến động thời gian:

1. Biến động có xu hướng (thay đổi dài hạn trong chuỗi dữ liệu thời gian: xu thế tăng, giảm hoặc dừng).
2. Biến động theo thời vụ (thay đổi đều đặn trong chuỗi dữ liệu thời gian tại cùng thời điểm mỗi năm).
3. Biến động theo chu kỳ (là thành phần tăng giảm có dạng sóng xung quanh đường xu thế).
4. Biến động bất thường (thay đổi không phải do các yếu tố kể trên).

## 5.2. Các chỉ tiêu phân tích

Để phản ánh đặc điểm biến động qua thời gian của hiện tượng nghiên cứu, người ta thường tính các chỉ tiêu sau đây:

### 5.2.1. Mức độ trung bình theo thời gian

Chỉ tiêu này phản ánh mức độ đại biểu của các mức độ tuyệt đối trong một dãy số thời gian. Mức độ trung bình theo thời gian được xác định theo các công thức khác nhau, tùy theo tính chất thời gian của dãy số.

#### 5.2.1.1 Đối với dãy số thời kỳ

Muốn tính mức độ bình quân: ta cộng các mức độ trong dãy số rồi chia cho số các mức độ, tức là:

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_i}{n}$$

Trong đó:

$y_i$  ( $i = 1, \dots, n$ ): Các mức độ của dãy số thời kỳ.

n: Số mức độ của dãy số.

Từ ví dụ trên ta tính doanh thu bình quân mỗi năm của đơn vị bưu điện X là:

$$\bar{y} = (23,9 + 28,1 + 37,3 + 47,2 + 67,4)/5 = 40,78 \text{ (tỷ đồng)}$$

Kết quả nói lên trong thời kỳ từ năm 1996 đến 2000, doanh thu trung bình hàng năm của bưu điện X là 40,78 tỷ đồng.

#### 5.2.1.2. Đối với dãy số thời điểm

\* Dãy số có khoảng cách thời gian bằng nhau: Mức độ trung bình được tính theo công thức sau:

$$\bar{y} = (y_1/2 + y_2 + y_3 + \dots + y_{n-1} + y_n/2) / (n - 1)$$

Trong đó:  $y_i$  ( $i=1, 2, \dots, n$ ) là các mức độ của dãy số thời điểm, n: Số mức độ của dãy số

Từ ví dụ bảng 5.2 ta tính  $\bar{y}$  :

$$\bar{y} = (256 / 2 + 364 + 370 + 352 / 2)/3 = 362,666$$

Có nghĩa là hàng hóa tồn kho trung bình của quý I là 362,666 triệu đồng.

\* Dãy số thời điểm có khoảng cách thời gian không bằng nhau; mức độ trung bình được tính theo công thức:

$$\bar{y} = \frac{y_1 t_1 + y_2 t_2 + y_3 t_3 + \dots + y_n t_n}{t_1 + t_2 + t_3 + \dots + t_n} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i}$$

Trong đó:



$y_i$  ( $i=1, 2, 3, \dots, n$ ): Các mức độ của dãy số thời điểm.

$t_i$  ( $i=1, 2, \dots, n$ ): Độ dài của các khoảng cách thời gian.

Từ ví dụ bảng 5.3, để tính  $y$  ta lập bảng sau:

**Bảng 5.4**

$y_i$	$t_i$ (số ngày)	$y_i t_i$
400	19 (1.1 đến 19.1)	7600
600	26 (20.1 đến 14.2)	15600
500	23 (15.2 đến 9.3)	11500
700	22 (10.3 đến 31.3)	15400
Cộng	90 ngày	50100

$$\bar{y} = \frac{50100}{90} \approx 556,7$$

Kết quả trên nói lên số dư tiền vay trung bình của quý I là 556,7 triệu đồng.

### 5.2.2. Lượng tăng hoặc giảm tuyệt đối

Là chỉ tiêu phản ánh sự thay đổi về trị số tuyệt đối của chỉ tiêu giữa hai thời gian nghiên cứu. Tùy theo mục đích nghiên cứu ta có:

#### 5.2.2.1. Lượng tăng (giảm) tuyệt đối từng kỳ (liên hoàn)

Chỉ tiêu này cho thấy lượng tăng (hoặc giảm) tuyệt đối của hiện tượng qua hai kỳ liên nhau.

Công thức tính:  $\delta = y_i - y_{i-1}$

$y_i$  : Mức độ của kỳ nghiên cứu .

$y_{i-1}$  :Mức độ của kỳ đứng liền trước đó.

\* *Nhận xét:* Một dãy số thời gian có  $n$  mức độ thì chỉ có thể tính được nhiều nhất là  $(n-1)$  lượng tăng (giảm) tuyệt đối từng kỳ.

Từ ví dụ bảng 5.1 ta có:

$$\delta_1 = y_2 - y_1$$

$$\delta_2 = y_3 - y_2$$

$$\delta_3 = y_4 - y_3$$

#### 5.2.2.2. *Lượng tăng (hoặc) giảm tuyệt đối định gốc*

Chỉ tiêu này phản ánh lượng tăng (hoặc giảm) của hiện tượng nghiên cứu qua một thời gian dài.

Công thức tính:  $\Delta_y = y_i - y_1$

$y_i$  : Mức độ của kỳ nghiên cứu.

$y_1$  : Mức độ kỳ gốc (thường là mức độ đầu tiên của dãy số).

Mối quan hệ giữa  $\Delta_y$  và  $\delta_y$

Tổng đại số của các lượng tăng (giảm) tuyệt đối từng kỳ bằng lượng tăng (giảm) tuyệt đối định gốc:

$$\Delta_y = \sum \delta_{yi}$$

#### 5.2.2.3. *Lượng tăng (giảm) tuyệt đối trung bình*

Chỉ tiêu này phản ánh lượng tăng (giảm) tuyệt đối điển hình của hiện tượng trong cả thời kỳ nghiên cứu:

$$\bar{\delta}_y = \sum \delta_{yi} / (n-1) = \Delta_y / (n-1) = (y_n - y_1) / (n-1)$$

### 5.2.3. Tốc độ phát triển

Là một số tương đối (thường được biểu hiện bằng lần hoặc %) phản ánh tốc độ và xu hướng biến động của hiện tượng qua thời gian. (tùy theo mục đích nghiên cứu ta có tốc độ phát triển sau đây:)

#### 5.2.3.1. Tốc độ phát triển từng kỳ (liên hoàn)

Chỉ tiêu này phản ánh hiện tượng đã phát triển với tốc độ phát triển cụ thể là bao nhiêu qua hai kỳ liên nhau:

$$k_i = y_i / (y_{i-1}) \text{ (ĐVT: lần hoặc \%)}$$

*Nhận xét:* Dãy số thời gian có  $n$  mức độ, chỉ có thể tính được nhiều nhất là  $(n-1)$  tốc độ phát triển từng kỳ.

#### 5.2.3.2. Tốc độ phát triển định gốc

Chỉ tiêu này đánh giá nhịp độ phát triển của hiện tượng nghiên cứu qua một thời gian dài.

$$K = y_n / y_1 \text{ (lần)}$$

$$\text{hoặc } K = y_n \times 100 / y_1 \text{ (\%)}$$

Trong đó:

$y_i$ : Mức độ từng kỳ nghiên cứu ( $i=2, 3, \dots, n$ )

$y_1$ : Mức độ kỳ gốc (thường là mức độ đầu tiên của dãy số).

Mối quan hệ giữa  $K$  và  $k$ : tích số của các tốc độ phát triển từng kỳ bằng tốc độ phát triển định gốc.

$$k_1 \cdot k_2 \dots k_{n-1} = K$$

### 5.2.3.3. Tốc độ phát triển trung bình

Chỉ tiêu này phản ánh tốc độ phát triển điển hình của hiện tượng trong cả thời kỳ nghiên cứu:

$$\bar{k} = \sqrt[n]{k_1 \cdot k_2 \cdot k_3 \dots k_{n-1}} = \sqrt[n]{\prod_{i=1}^n k_i} = \sqrt[n]{\frac{y_n}{y_1}} \quad (\text{lần hoặc } \%)$$

### 5.2.4. Tốc độ tăng hoặc giảm

Là chỉ tiêu cho thấy nhịp độ tăng trưởng của hiện tượng nghiên cứu qua thời gian.

#### 5.2.4.1. Tốc độ tăng (giảm) liên hoàn (từng kỳ)

Chỉ tiêu này phản ánh hiện tượng đã tăng (hoặc giảm) với tốc độ là bao nhiêu qua hai thời kỳ nghiên cứu liên nhau.

$$a = \frac{\delta_y}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = k - 1$$

hoặc  $a = k - 100$  (%)

#### 5.2.4.2. Tốc độ tăng (giảm) định gốc

Chỉ tiêu này phản ánh hiện tượng đã tăng (hoặc giảm) với tốc độ là bao nhiêu qua một thời gian dài.

$$b = \frac{\Delta_y}{y_1} = \frac{y_i - y_1}{y_1} = K - 1 \quad (\text{lần})$$

hoặc  $b = K - 100$  (%)

#### 5.2.4.3. Tốc độ tăng (giảm) trung bình

Chỉ tiêu này cho thấy nhịp độ tăng (giảm) điển hình của hiện tượng trong cả thời kỳ nghiên cứu.

$$\bar{a} = \bar{k} - 1 \text{ (lần)}$$

$$\text{hoặc } \bar{a} = \bar{k} - 100 \text{ (\%)}$$

#### 5.2.5. Trị tuyệt đối của 1% tăng (hoặc giảm)

Chỉ tiêu này dùng để đánh giá trị số tuyệt đối tương ứng với 1% của tốc độ tăng (hoặc giảm) từng kỳ.

$$c = \frac{\delta_y}{a} = \frac{y_i - y_{i-1}}{k - 100} = \frac{y_{i-1}}{100}$$

(ĐVT trùng với ĐVT của lượng biến)

### 5.3. Các phương pháp biểu hiện xu hướng phát triển của hiện tượng

#### 5.3.1. Phương pháp mở rộng khoảng cách thời gian

Phương pháp này được sử dụng khi một dãy số thời kỳ có khoảng cách thời gian tương đối ngắn và có nhiều mức độ mà qua đó chưa phản ánh được xu hướng biến động của hiện tượng.

*Ví dụ:* Có tài liệu về sản lượng hàng tháng của năm 1999 ở một xí nghiệp như sau:

Dãy số trên cho thấy sản lượng các tháng khi thì tăng, khi thì giảm thất thường, không nói rõ xu hướng biến động. Người ta có thể mở rộng khoảng cách thời gian từ tháng sang quý.

**Bảng 5.5**

Tháng	Sản lượng (1.000 tấn)	Tháng	Sản lượng (1.000 tấn)
1	40,4	7	40,8
2	36,8	8	44,8
3	40,6	9	49,4
4	38,0	10	48,9
5	42,2	11	46,2
6	48,5	12	42,2

Do khoảng cách thời gian được mở rộng (từ tháng sang quý), nên trong mỗi mức độ của dãy số mới chịu sự tác động của các nhân tố ngẫu nhiên (với chiều hướng khác nhau) phần nào đã được bù trừ (triệt tiêu) và do đó cho ta thấy rõ xu hướng biến động cơ bản là: tình hình sản xuất của xí nghiệp tăng dần từ quý 1 đến quý 4 của năm 1999.

**Bảng 5.6**

Quý	Sản lượng (1.000 tấn)
1	117,8
2	128,7
3	135,0
4	137,3

### ***5.3.2. Phương pháp số trung bình trượt***

Số trung bình trượt (còn gọi là số trung bình di động) là số trung bình cộng của một nhóm nhất định các mức độ của dãy số được tính bằng cách lần lượt loại dần các mức độ đầu, đồng thời thêm vào các mức độ tiếp theo sao cho tổng số lượng các mức độ tham gia tính số trung bình không thay đổi.

Giả sử có dãy thời gian  $y_1, y_2, y_3, \dots, y_{n-1}, y_n$

Nếu tính trung bình trượt cho nhóm ba mức độ, ta sẽ có:

$$\bar{y}_2 = (y_1 + y_2 + y_3) / 3$$

$$\bar{y}_3 = (y_2 + y_3 + y_4) / 3$$

$$\bar{y}_4 = (y_3 + y_4 + y_5) / 3$$

$$\bar{y}_{n-1} = (y_{n-2} + y_{n-1} + y_n) / 3$$

Từ đó, ta có một dãy số mới gồm các số trung bình trượt là  $\bar{y}_2, \bar{y}_3, \dots, \bar{y}_{n-1}$

Từ ví dụ (bảng 5.5), tính số trung bình trượt cho nhóm ba mức độ, ta có:

**Bảng 5.7**

Tháng trượt $y_i$	Sản lượng trượt $y_i$	Số trung bình	Tháng trượt	Sản lượng trượt $y_i$	Số trung bình
1	40,4		7	40,8	44,7
2	36,8	39,3	8	44,8	45,0
3	40,6	38,5	9	49,4	47,7
4	38,0	40,3	10	48,9	48,2
5	42,2	42,9	11	46,4	45,8
6	48,5	43,8	12	42,2	

Trung bình trượt càng được tính từ nhiều mức độ thì càng có tác dụng san bằng ảnh hưởng của các nhân tố ngẫu nhiên. Nhưng mặt khác bị làm giảm số lượng các mức độ của dãy trung bình trượt.

### 5.3.3. Phương pháp hồi quy

Trên cơ sở dãy số thời gian, người ta tìm một hàm số (gọi là phương trình hồi quy) phản ánh sự biến động của hiện tượng qua thời gian có dạng tổng quát như sau:

$$Y = a_0 + a_1 t_1 + \dots + a_n t_n + e$$

Trong đó:

Y: Biến số phụ thuộc (kết quả phân tích)

$a_0, a_1, \dots, a_n$ : Các tham số.

t: Thứ tự thời gian.

e: Sai số

Để lựa chọn đúng đắn dạng của phương trình hồi quy đòi hỏi phải dựa vào sự phân tích đặc điểm biến động của hiện tượng qua thời gian, đồng thời kết hợp với một số phương pháp đơn giản khác (như dựa vào đồ thị, dựa vào độ tăng (giảm) tuyệt đối, dựa vào tốc độ phát triển, ...).

Các tham số  $a_i$  ( $i = 1, 2, 3, \dots, n$ ) thường được xác định bằng phương pháp bình phương nhỏ nhất. Tức là:

$$\sum (y_{LT} - y_{TT})^2 = \min$$

Sau đây là một số dạng phương trình hồi quy đơn giản thường được sử dụng: Phương trình đường thẳng:  $y = a_0 + a_1 t$

Phương trình đường thẳng được sử dụng khi các lượng tăng (hoặc giảm) tuyệt đối liên hoàn (còn gọi là sai phân bậc 1) xấp xỉ nhau.

Để xác định  $a_0$  và  $a_1$ : ta áp dụng phương pháp bình phương nhỏ nhất. Từ đó  $a_0$  và  $a_1$  được xác định bởi hệ phương trình sau:



$$\begin{cases} \sum_{i=1}^n y = na_0 + a_1 \sum_{i=1}^n t \\ \sum_{i=1}^n yt = a_0 \sum_{i=1}^n t + a_1 \sum_{i=1}^n t^2 \end{cases} (*)$$

Ví dụ: Có số liệu về doanh thu của một đơn vị sản xuất qua các năm như sau:

Năm	1998	1999	2000	2001	2002
Doanh thu (tỷ đồng)	30	32	31	34	33

Để tính  $a_0$  và  $a_1$  cho ví dụ này, ta lập bảng sau:

Năm	y	t	$t^2$	$T_y$	$y_{Lr}t$
1998	30	1	1	30	30,4
1999	32	2	4	64	31,2
2000	31	3	9	93	32,0
2001	34	4	16	136	32,8
2002	33	5	25	165	33,6
$\Sigma$	160	15	55	488	

Thế các giá trị tương ứng trong bảng vào hệ phương trình trên (\*) ta được:

$$\begin{cases} 160 = 5a_0 + 15a_1 \\ 488 = 15a_0 + 55a_1 \end{cases}$$

Từ đây ta tính được  $a_0 = 29,6$  và  $a_1 = 0,8$ . Thế các giá trị  $t$  lần lượt từ 1 đến 5 tương ứng với thời gian từ năm 1998 đến năm 2002 ta

tính được các giá trị doanh thu theo đường hồi quy lý thuyết  $y = a_0 + a_1 t$ . là các giá trị trong cột  $y_{L\Gamma t}$ .

Ta nhận thấy rằng: biến  $t$  là biến thứ tự thời gian, ta có thể thay  $t$  bằng  $t'$  (nhưng vẫn đảm bảo tính thứ tự), sao cho  $\sum t' = 0$  thì việc tính toán sẽ đơn giản hơn. Có hai trường hợp:

1. Nếu thứ tự thời gian là số lẻ thì lấy thời gian đứng ở giữa bằng 0, các thời gian đứng trước là -1, -2, -3 và thời gian đứng sau là 1, 2, 3.
2. Nếu thứ tự thời gian là số chẵn thì lấy hai thời gian đứng giữa là -1 và 1, các thời gian đứng trước lần lượt là -3, -5, ... và đứng sau lần lượt là 3, 5, ...

Với  $\sum t' = 0$  thì hệ phương trình trên sẽ là:

$$\sum y = na_0 \Rightarrow a_0 = \sum y / n$$

$$\sum t' y = a_1 \sum t'^2 \Rightarrow a_1 = \sum t' y / \sum t'^2$$

Khi đó:  $y_{L\Gamma t'} = a'_0 + a'_1 t'$

Với cách chọn  $\sum t' = 0$ , ta lập bảng sau:

**Bảng 5.8**

Năm	y	t'	t' <sup>2</sup>	t'y	y <sub>LΓt</sub>
1998	30	-2	4	-60	30,4
1999	32	-1	1	-32	31,2
2000	31	0	0	0	32,0
2001	34	1	1	34	32,8
2002	33	2	4	66	33,6
$\sum$	160	0	10	8	

$$a_0 = 160 / 5 = 32 ; \quad a_1 = 8 / 10 = 0,8$$

$$y_{T'} = 32 + 0,8t' (*)$$

Để dự đoán doanh thu cho năm 2003 thế  $t = 3$  vào phương trình (\*) ta được

$$y = 32 + 0,8 \times 3 = 34,4 \text{ (tỷ đồng)}$$

Với hai cách chọn  $\sum t \neq 0$  và  $\sum t = 0$ , ta thấy kết quả vẫn như nhau.

#### 5.3.4. Phương pháp biểu hiện biến động thời vụ

Sự biến động của một số hiện tượng trong kinh tế xã hội thường có tính thời vụ, nghĩa là hàng năm trong từng thời gian nhất định, sự biến động được lặp đi lặp lại.

Nghiên cứu biến động thời vụ nhằm đề ra những chủ trương, biện pháp phù hợp, kịp thời, hạn chế những ảnh hưởng của biến động thời vụ đối với sản xuất và sinh hoạt của xã hội.

Nhiệm vụ của nghiên cứu thống kê là dựa vào số liệu của nhiều năm (ít nhất là ba năm) để xác định tính chất và mức độ của biến động thời vụ. Phương pháp này thường được sử dụng để tính các chỉ số thời vụ.

Chỉ số thời vụ được tính theo công thức:

$$I_i = \left( \overline{y_i} / \overline{y_0} \right) \times 100(\%)$$

Trong đó:

$I_i$ : Số thời vụ của thời gian  $i$ .

$\overline{y}_i$  : Số bình quân của các mức độ cùng thời gian i.

$\overline{y}_0$  : Số bình quân của tất cả các mức độ trong dãy số.

*Ví dụ:* Có số liệu về sản lượng điện thoại đường dài của một đơn vị bưu điện qua các năm như sau:

$$\bar{y} = \frac{9.339.023}{36} = 259.417,3 \text{ (cuộc)}$$

Qua kết quả trình bày ở bảng trên ta thấy sản lượng điện thoại đường dài trong nước tăng cao nhất ở những tháng cuối năm (gần tết) và giảm thấp nhất ở tháng giêng và tháng hai.

Giả sử kế hoạch sản lượng điện thoại cho năm 2000 là 6 000 000 cuộc thì ta dự đoán sản lượng của từng tháng của năm 2000 sẽ là: các sản lượng trong cột (7).

Bảng 5.9

Tháng	Sản lượng điện thoại đường dài (cuộc)			Cộng các tháng cùng tên ( $\bar{y}_i$ )	Bình quân các tháng cùng tên ( $\bar{y}_i$ )	Chỉ số thời vụ	
	1997	1998	1999				
A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	137.139	184.326	241.892	563.357	187.785	72.38	361.937
2	130.009	213.218	270.682	613.909	204.636	78.88	394.415
3	159.241	234.3	350.684	744.255	248.075	95.62	478.158
4	147.674	222.667	338.037	708.378	236.125	91.02	455.108
5	148.589	236.26	353.488	738.337	246.112	94.87	474.356
6	162.643	229.976	368.601	761.22	253.74	97.81	489.058
7	160.598	235.483	376.304	772.385	257.461	99.25	496.231
8	172.235	246.789	383.399	802.423	267.474	103.1	515.529
9	180.119	249.628	410.292	840.039	280.013	107.9	539.696
10	181.161	254.651	421.905	857.717	285.905	110.2	551.054
11	185.552	246.818	415.502	847.872	282.624	108.94	544.729
12	197.785	259.143	632.233	1089.16	363.053	139.95	699.748
<b>Cộng</b>	<b>9.339.023</b>						

**Bảng T-Students**

<b>One Sided</b>	<b>75%</b>	<b>80%</b>	<b>85%</b>	<b>90%</b>	<b>95%</b>	<b>97.5%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.75%</b>	<b>99.9%</b>	<b>99.95%</b>
<b>Two Sided</b>	<b>50%</b>	<b>60%</b>	<b>70%</b>	<b>80%</b>	<b>90%</b>	<b>95%</b>	<b>98%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.8%</b>	<b>99.9%</b>
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318

<b>One Sided</b>	<b>75%</b>	<b>80%</b>	<b>85%</b>	<b>90%</b>	<b>95%</b>	<b>97.5%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.75%</b>	<b>99.9%</b>	<b>99.95%</b>
<b>Two Sided</b>	<b>50%</b>	<b>60%</b>	<b>70%</b>	<b>80%</b>	<b>90%</b>	<b>95%</b>	<b>98%</b>	<b>99%</b>	<b>99.5%</b>	<b>99.8%</b>	<b>99.9%</b>
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



## TÀI LIỆU THAM KHẢO

- 1) Võ Đức Hoàng Vũ, *Slide bài giảng Phân tích dữ liệu và dự báo kinh tế*, Khoa Kinh tế phát triển - Trường Đại học Kinh tế TP HCM.
- 2) Phùng Thanh Bình - Nguyễn Trọng Hoài, *Slide bài giảng Phân tích dữ liệu và dự báo kinh tế*, Khoa Kinh tế phát triển - Trường Đại học Kinh tế TP HCM.
- 3) Financial Statement Analysis - Leopold A. Bernstein, John J. Wild.
- 4) Võ Thành Danh, *Kế toán Phân tích*, NXB Thống kê, 2001.
- 5) Võ Văn Huy và Huỳnh Ngọc Liễu, *Hệ thống Thông tin Quản lý*, Nhà Xuất Bản Khoa học Kỹ thuật, 2001.
- 6) Giáo trình phân tích hoạt động kinh doanh - Trường Đại học Kinh tế quốc dân.
- 7) Phạm Văn Dược, *Phân tích hoạt động kinh doanh*, NXB Thống kê, 1999.
- 8) Nguyễn Tấn Bình, *Phân tích hoạt động kinh doanh nghiệp*, NXB TP.HCM, 2003.
- 9) Huỳnh Đức Lộng, *Phân tích hoạt động kinh tế doanh nghiệp*, NXB Thống kê, 1977.
- 10) Nguyễn Năng Phúc, *Phân tích hoạt động kinh doanh của doanh nghiệp*, NXB Thống kê, 1998.
- 11) Nguyễn Năng Phúc, *Giáo trình phân tích hoạt động kinh doanh*, NXB Thống kê, 2001.
- 12) Nguyễn Quang Quỳnh, *Phân tích kinh tế hoạt động kinh doanh của doanh nghiệp*, NXB Thống kê, 1991.