

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC



BÁO CÁO MÔN HỌC
CHUỖI THỜI GIAN

ĐỀ TÀI:

**Ứng dụng mô hình ARIMA trong dự báo chỉ số
giá chứng khoán theo cách tiếp cận của
phương pháp Box - Jenkins**

GV hướng dẫn: TS. NGUYỄN THỊ NGỌC ANH

Nhóm sinh viên thực hiện: Nhóm 13

Họ và tên	MSSV	Lớp
Nguyễn Minh Tuấn	20185420	Toán - Tin 01 - K63
Nguyễn Thị Bích Ngọc	20185388	Toán - Tin 01 - K63
Hồ Thị Tuyết	20185424	Toán - Tin 01 - K63

Hà Nội, Tháng 02 năm 2022

Mục lục

Lời nói đầu	3
1 Lý luận chung về chứng khoán	4
1.1 Tổng quan về chứng khoán	4
1.1.1 Khái niệm chứng khoán	4
1.1.2 Phân loại chứng khoán	4
1.1.3 Một số loại chứng khoán cơ bản	5
1.2 Chỉ số giá chứng khoán	8
1.2.1 Khái niệm	8
1.2.2 Một số phương pháp ước lượng chỉ số giá chứng khoán .	9
2 Cơ sở lý thuyết của mô hình ARIMA	11
2.1 Tổng quan về bài toán dự báo	11
2.1.1 Khái niệm	11
2.1.2 Quy trình dự báo chung	11
2.2 Các khái niệm cơ bản của chuỗi thời gian	13
2.2.1 Khái niệm và các thành phần của một chuỗi thời gian .	13
2.2.2 Khái niệm dừng yếu	14
2.2.3 Hàm tự tương quan ACF	14
2.2.4 Phương pháp khử thành phần xu hướng và thành phần mùa	15
2.2.5 Quá trình tuyến tính	15
2.3 Một số mô hình chuỗi thời gian dừng	16
2.3.1 Chuỗi iid noise và chuỗi White noise	16
2.3.2 Quá trình trung bình trượt MA(q)	17
2.3.3 Quá trình tự hồi quy AR(p)	18
2.3.4 Quá trình ARMA(p, q)	19
2.4 Chuỗi thời gian không dừng và mô hình ARIMA	20

2.4.1	Mô hình ARIMA cho chuỗi thời gian không dừng	20
2.4.2	Mô hình ARIMA có tính mùa	21
2.4.3	Phương pháp Box - Jenkins	21
3	Ứng dụng mô hình ARIMA trong dự báo chỉ số giá chứng khoán	23
3.1	Chuẩn bị mẫu dữ liệu	23
3.1.1	Giới thiệu bộ dữ liệu	23
3.1.2	Phân tích dữ liệu khám phá	24
3.2	Xây dựng mô hình dự báo chỉ số giá chứng khoán	25
3.2.1	Nhận dạng mô hình ARIMA(p, d, q)	25
3.2.2	Ước lượng các tham số của mô hình	27
3.2.3	Kiểm tra chẩn đoán	28
3.2.4	Ứng dụng mô hình dự báo chỉ số giá chứng khoán	30
	Kết luận	31
	Tài liệu tham khảo	32

Lời nói đầu

Qua hơn mười năm thực hiện đường lối đổi mới của Đảng, nền kinh tế nước ta đã đạt được những thành tựu và bước tiến hết sức quan trọng. Tuy nhiên, để đạt được mục tiêu là đến năm 2020 đưa nước ta về cơ bản trở thành một nước công nghiệp có cơ sở vật chất và kỹ thuật hiện đại, cơ cấu kinh tế hợp lý, đảm bảo nhịp độ tăng trưởng bình quân GDP ở mức 9-10%/năm, thực hiện công nghiệp hoá hiện đại hoá thành công thì đòi hỏi chúng ta phải có vốn rất lớn. Vì vậy, để đáp ứng yêu cầu đó đến tháng 7/2000 Việt Nam đã cho ra đời trung tâm giao dịch chứng khoán tại Thành Phố Hồ Chí Minh. Tuy nhiên trong giai đoạn hiện nay thị trường mới hình thành và còn mang tính chất thử nghiệm nên việc đưa ra những phân tích về xu hướng biến động giá cổ phiếu cho thị trường là rất có ý nghĩa.

Qua một thời gian nghiên cứu các tài liệu về chứng khoán, về mô hình để phân tích về sự biến động của giá chứng khoán trên thị trường và bằng sự hiểu biết của nhóm em trong suốt quá trình học tập, nhóm em đã mạnh dạn chọn đề tài: **Ứng dụng mô hình ARIMA trong dự báo chỉ số giá chứng khoán theo cách tiếp cận của phương pháp Box - Jenkin.**

Ngoài các phần lời nói đầu, mục lục, kết luận, tài liệu tham khảo, nội dung bản báo cáo của nhóm em sẽ được trình bày theo 3 phần:

- **Chương 1: Lý luận chung về chứng khoán**
- **Chương 2: Cơ sở lý thuyết của mô hình ARIMA**
- **Chương 3: Ứng dụng mô hình ARIMA trong dự báo chỉ số giá chứng khoán**

Chúng em xin gửi lời cảm ơn sâu sắc đến: **TS. Nguyễn Thị Ngọc Anh** đã truyền đạt cho chúng em những kiến thức nền tảng để hoàn thành bài báo cáo này. Báo cáo không tránh khỏi những thiếu sót, chúng em kính mong nhận được sự nhận xét, góp ý của cô và các bạn để báo cáo của nhóm chúng em được hoàn thiện hơn.

Chương 1

Lý luận chung về chứng khoán

1.1 Tổng quan về chứng khoán

1.1.1 Khái niệm chứng khoán

Chứng khoán là một bằng chứng xác nhận sự sở hữu hợp pháp dưới hình thức chứng chỉ, bút toán ghi sổ hoặc dữ liệu điện tử của người sở hữu đó (hoặc quyền đòi nợ hợp pháp) với tài sản hoặc phần vốn của công ty hay tổ chức đã phát hành.

1.1.2 Phân loại chứng khoán

- **Theo tính chất**

Phân loại chứng khoán theo tính chất được phân thành chứng khoán vốn, chứng khoán nợ và chứng khoán phát sinh.

- Chứng khoán vốn là một loại chứng khoán xác nhận sự góp vốn và quyền sở hữu hợp pháp đối với tổ chức phát hành.
- Chứng khoán nợ (trái phiếu, tín phiếu) là một loại chứng khoán xác nhận một khoản nợ quy định nghĩa vụ của người phát hành phải trả cho người sở hữu chứng khoán một khoản tiền xác định trong khoảng thời gian cụ thể.
- Chứng khoán phát sinh là các chứng khoán hình thành do nhu cầu giao dịch mua và bán cổ phiếu, trái phiếu. và phát triển phụ thuộc vào độ phát triển của thị trường chứng khoán.
- Chứng khoán phát sinh có thể phân loại thành quyền mua trước, chứng khế, hợp đồng kì hạn, hợp đồng tương lai, quyền chọn.

- **Khả năng chuyển nhượng**

Phân loại chứng khoán theo khả năng chuyển nhượng được phân thành chứng khoán vô danh và chứng khoán ghi danh.

- Chứng khoán vô danh (giấy xác nhận quyền sở hữu không ghi tên chủ sở hữu) là loại chứng khoán có thể chuyển nhượng một cách dễ dàng, không cần những thủ tục xác nhận của công ty hoặc cơ quan công chứng. Người mua có trách nhiệm chi trả cho người bán một khoản tiền được xác định.
- Chứng khoán ghi danh (giấy xác nhận quyền sở hữu có ghi chủ sở hữu) là loại chứng khoán có thể được chuyển nhượng nhưng phải tuân theo những quy định pháp lý cụ thể. Nếu muốn chuyển nhượng, người chủ sở hữu phải chứng minh mình được quyền chuyển nhượng (xác nhận chữ ký, căn cước công dân, tên trong danh sách cổ đông)

• Theo thu nhập

Phân loại chứng khoán theo thu nhập gồm chứng khoán thu nhập cố định, chứng khoán thu nhập biến đổi và chứng khoán hỗn hợp.

- Chứng khoán thu nhập cố định là loại chứng khoán có quyền yêu cầu thu nhập cố định không phụ thuộc vào kết quả hoạt động của nhà phát hành và tuân theo quy định pháp lý cụ thể. Chứng khoán cố định bao gồm trái phiếu nhà nước, trái phiếu địa phương, trái phiếu công ty, trái phiếu kho bạc và trái phiếu ngân hàng và tín phiếu quỹ tiết kiệm.
- Chứng khoán có thu nhập biến đổi là loại chứng khoán được thu nhập do việc sở hữu cổ phiếu mang lại (cổ tức) và nó biến động theo kết quả kinh doanh của công ty.
- Chứng khoán hỗn hợp là loại chứng khoán kết hợp giữa chứng khoán thu nhập cố định và chứng khoán có thu nhập biến đổi. Chứng khoán này ra đời với mục đích thích ứng với nhu cầu thị trường vốn. Chứng khoán hỗn hợp bao gồm trái phiếu công ty có khả năng chuyển đổi, trái phiếu có quyền mua cổ phiếu và trái phiếu có thu nhập bổ sung.

1.1.3 Một số loại chứng khoán cơ bản

• Cổ phiếu

Cổ phiếu là giấy chứng nhận quyền sở hữu hợp pháp đối với thu nhập và tài sản của công ty cổ phần.

Cổ phiếu được chia làm hai loại bao gồm cổ phiếu thông thường và cổ phiếu ưu đãi.

- Cổ phiếu thông thường là loại cổ phiếu biến động tùy theo sự biến động về lợi nhuận của công ty. Cổ phiếu thông thường có thu nhập không cố định, không chỉ phụ thuộc vào lợi nhuận công ty mà còn phụ thuộc vào các nhân tố như môi trường kinh tế, biến đổi lãi suất,...

- Cổ phiếu ưu đãi là loại cổ phiếu không phụ thuộc vào lợi nhuận công ty cho nên thu nhập luôn cố định. Tuy nhiên cổ phiếu ưu đãi lại phụ thuộc vào sự thay đổi lãi suất trái phiếu và tình hình tài chính của công ty.

- **Trái phiếu**

Trái phiếu là giấy chứng nhận việc vay vốn của người phát hành và người sở hữu trái phiếu. Tổ chức nhà phát hành phải có trách nhiệm hoàn trả toàn bộ gốc và lãi cho người sở hữu trái phiếu khi tới hạn.

Trái phiếu được phân loại theo người phát hành và theo lợi tức trái phiếu.

- Phân loại theo người phát hành

- * Trái phiếu chính phủ là loại trái phiếu do chính phủ hoặc những ngành đặc biệt như giao thông, bưu điện, . . . do Chính phủ được ủy quyền phát hành. Trái phiếu này được phát hành nhằm mục đích xây dựng các công trình công cộng hoặc sự thiếu hụt Ngân sách nhà nước giải quyết các vấn đề về tài chính. Trái phiếu chính phủ có khả năng chuyển nhượng, cầm cố và có uy tín nhất thị trường, ít rủi ro nhất.
- * Trái phiếu doanh nghiệp là loại trái phiếu do doanh nghiệp nhà nước, công ty cổ phần, công ty trách nhiệm hữu hạn phát hành nhằm mục đích tăng vốn đầu tư hoặc giải quyết các vấn đề tài chính.
- * Trái phiếu của tổ chức tài chính, ngân hàng là loại trái phiếu do tổ chức tài chính hoặc ngân hàng phát hành nhằm tạo nguồn vốn tín dụng.

- Phân loại theo lợi tức trái phiếu

- * Trái phiếu có lãi suất biến đổi là loại trái phiếu có sự biến đổi về lãi suất.
- * Trái phiếu có lãi suất bằng không là loại trái phiếu mà người mua trái phiếu không nhận được lãi nhưng được mua với mức giá thấp hơn mức giá phát hành.
- * Trái phiếu có lãi suất cố định là loại trái phiếu lợi tức được xác định theo tỷ lệ mệnh giá.

- **Chứng khoán phái sinh**

- **Quyền mua trước**

Quyền mua trước là quyền ưu đãi gắn liền với một cổ phiếu đang lưu hành để huy động thêm vốn cổ phần. Cho phép người sở hữu những cổ phần đang lưu hành được mua một số cổ phiếu nhất định trong đợt phát hành mới của công ty với mức giá thấp hơn giá bán ra thị trường bán cho công chúng trong thời hạn nhất định.

Mức giá đăng ký dạng quyền mua trước thấp hơn mức giá hiện hành của cổ phiếu trong thời điểm phát hành là do hai nguyên nhân chính. Thứ nhất, rủi ro có thể xảy ra khi giá thị trường giảm trong thời điểm phát hành dẫn đến ảnh hưởng tới đợt phát hành. Thứ hai, mức giá chênh lệch giữa người có quyền mua trước và người không có quyền mua trước làm tang tính hấp dẫn cho các cổ đông cũ. Giá trị quyền mua trước được xác định bằng sự chênh lệch giữa thị giá cổ phiếu trước và sau khi tăng vốn.

$$M = \frac{P - F}{N + 1}$$

Trong đó:

P là thị giá cổ phiếu

F là giá trị ưu đãi của chứng quyền

N là số lượng quyền cần để mua một cổ phần mới = Số lượng cổ phần cũ đang lưu hành/số lượng cổ phần mới. trong đó, số cổ phần mới bằng với mức vốn cần huy động/ giá đăng ký mua.

– Chứng khế

Chứng khế (chứng phiếu) là một loại chứng khoán được phát hành cùng với trái phiếu mà người sở hữu nó có quyền mua chứng khoán với một số lượng, mức giá, thời hạn nhất định.

Chứng khế được phát hành nhằm khuyến khích các nhà đầu tư tiềm năng mua cổ phiếu, trái phiếu với giá cao hơn giá bán ra thị trường.

Giá trị được xác định như sau:

Giá trị chứng khế = giá trị nội tại + giá trị thời gian

Trong đó

Giá trị nội tại = (giá cổ phần – giá thực hiện) * số cổ phần được mua của chứng khế.

Do thực hiện quyền chứng khế sẽ làm giảm giá cổ phần ảnh hưởng tới lợi nhuận của cổ phần. Do vậy định giá chứng khế được áp dụng với hợp đồng quyền chọn mua kiểu Mỹ.

Giá hợp lý của chứng khế = $\frac{Pc}{1+q}$ * số cổ phần được mua chứng quyền

– Hợp đồng kỳ hạn

Hợp đồng kỳ hạn là một loại hợp đồng thỏa thuận giữa người mua và người bán để thực hiện một giao dịch hàng hóa với khối lượng và mức giá nhất định trong thời điểm tương lai.

Hợp đồng kỳ hạn không trao đổi trên thị trường, không có định giá hàng ngày và phải tuân theo tiêu chuẩn của thị trường riêng biệt.

Giá trị của hợp đồng kỳ hạn là giá trị giao hàng, chuyển giao để giá trị hai bên mua bán bằng không (có nghĩa không có chi phí mua bán hợp đồng kỳ hạn).

– **Hợp đồng tương lai**

Hợp đồng tương lai là hợp đồng thỏa thuận dùng để mua hoặc bán tài sản với mức giá nhất định trong thời điểm trong tương lai.

Hợp đồng tương lai có chức năng cung cấp cơ hội cho những người tham gia thị trường dùng để phòng ngừa rủi ro về những biến động giá.

Hợp đồng tương lai có những đặc tính linh hoạt giúp khắc phục được những nhược điểm của hợp đồng kỳ hạn và là một phương án tốt để phòng ngừa rủi ro.

– **Quyền chọn**

Hợp đồng quyền chọn là một hợp đồng cho phép người sở hữu có quyền được mua, bán với một khối lượng hàng hóa cơ sở, mức giá, thời hạn nhất định. Các hàng hóa cơ sở có thể là cổ phiếu, chỉ số cổ phiếu, trái phiếu,...

Hợp đồng quyền chọn gồm những đặc điểm như loại quyền, tên hàng hóa cơ sở và khối lượng được mua hoặc bán theo quyền, ngày đáo hạn, giá thực thi.

Đối tượng tham gia trên thị trường quyền chọn bao gồm người mua quyền mua (buyers of call), người bán quyền mua (sellers of calls), người mua quyền bán (buyers of puts), người bán quyền bán (sellers of puts).

Các bộ phận cấu thành nên giá của quyền chọn bao gồm giá trị nội tại của quyền, giá trị thời gian của quyền. Các yếu tố ảnh hưởng tới quyền chọn bao gồm giá thị trường chứng khoán cơ sở, giá thực hiện, thời gian cho tới khi đáo hạn, lãi suất ngắn hạn phi rủi ro trong suốt thời gian tồn tại của quyền, lãi suất Coupon, mức dao động dự đoán của các mức lãi suất trong suốt thời hạn của quyền.

1.2 Chỉ số giá chứng khoán

1.2.1 Khái niệm

Chỉ số giá chứng khoán là thông tin thể hiện giá chứng khoán bình quân hiện tại so với giá bình quân thời kỳ gốc đã chọn dùng chủ yếu để đo lường sự biến động giá chứng khoán trên thị trường. Chỉ số giá chứng khoán biểu hiện tình hình hoạt động của thị trường chứng khoán, là thông tin rất quan trọng đối với hoạt động của thị trường của nhà đầu tư và các nhà phân tích kinh tế.

1.2.2 Một số phương pháp ước lượng chỉ số giá chứng khoán

- **Chỉ số giá bình quân giản đơn**

Đây là chỉ số bình quân số học giản đơn, không có sự tham gia của quyền số.

$$I = \frac{\sum P_t}{\sum P_0}$$

Trong đó:

I là chỉ số giá bình quân giản đơn

P_t là giá thời kỳ t của hàng hóa tham gia tính toán

P_0 là giá thời kỳ gốc chọn trước

Chỉ số này tính toán rất đơn giản vì không phải theo dõi sự biến động của quyền số, Nhưng chỉ số này chứa đựng tính chất của số bình quân giản đơn. Bởi vậy, chỉ nên ứng dụng khi tổng thể khá đồng đều, chênh lệch không quá lớn.

- **Chỉ số giá bình quân gia quyền**

Chỉ số giá bình quân gia quyền là chỉ số giá bình quân được tính có sự tham gia của những nhân tố tỷ trọng khối lượng. Tỷ trọng khối lượng của tổng thể càng lớn thì càng ảnh hưởng đến chỉ số giá chung và ngược lại.

$$I = \sum \frac{q * P_t}{q * P_0}$$

Trong đó:

I là chỉ số giá bình quân gia quyền

P_t là giá thời kỳ báo cáo

q là khối lượng (có thể theo thời kì gốc hoặc thời kỳ báo cáo)

- **Chỉ số giá bình quân Laspeyres** Chỉ số giá bình quân Laspeyres là chỉ số giá bình quân gia quyền với quyền số là khối lượng thời kỳ gốc.

Chỉ số giá bình quân Laspeyres được xác định qua công thức sau:

$$I_L = \frac{\sum q_0 * P_t}{\sum q_0 * P_0}$$

Trong đó:

I_L là chỉ số giá bình quân Laspeyres

P_t là giá thời kỳ báo cáo

P_0 là giá thời kỳ gốc

q_0 là khối lượng thời kỳ gốc hoặc cơ cấu của khối lượng thời kỳ gốc.

- **Chỉ số giá bình quân Paascher**

Chỉ số giá bình quân Paascher là chỉ số giá bình quân gia quyền lấy quyền số là khối lượng thời kỳ báo cáo.

Chỉ số giá bình quân Paascher được xác định qua công thức sau:

$$I_P = \frac{\sum q_t * P_t}{\sum q_t * P_0}$$

Trong đó:

P_t là chỉ số giá Paascher

P_t là giá thời kỳ cần tính

P_0 là giá thời kỳ gốc

Q_t là khối lượng thời kỳ tính hoặc cơ cấu của khối lượng thời kỳ báo cáo.

• Chỉ số giá bình quân Fisher

Chỉ số giá bình quân Fisher là chỉ số giá bình quân giữa tích chỉ số giá Paascher và chỉ số giá Laspayres.

Chỉ số giá bình quân Fisher được xác định qua công thức sau:

$$I_F = \sqrt{I_P * I_L}$$

Trong đó:

I_F là chỉ số giá bình quân Fisher

I_P là chỉ số giá bình quân Paasche

I_L là chỉ số giá bình quân Laspeyres

Một số chỉ số giá chứng khoán quốc tế như của hồng kong, mỹ

- Các chỉ số giá cổ phiếu của thị trường chứng khoán Hong Kong như chỉ số HangSeng (HIS), chỉ số tổng hợp cổ phiếu thường (AOI), chỉ số tham chiếu HangSeng London (HSLRI), chỉ số HangSeng châu Á (HSAI).
- Các chỉ số giá cổ phiếu của thị trường chứng khoán Mỹ như chỉ số Dow Jones (Dow Jones Average), chỉ số DJIA (Dow Jones công nghiệp), chỉ số DJTA (Dow Jones vận tải), chỉ số ngành phục vụ công cộng (DJUA),...

Chỉ số chứng khoán Việt Nam

$$VNIndex = \frac{\sum Q_1 * P_1}{\sum Q_0 * P_0} * I_0$$

Trong đó:

P_1 là giá hiện hành chứng khoán

Q_1 là khối lượng cổ phiếu đang lưu hành

P_0 là giá cổ phiếu thời kỳ gốc

Q_0 là khối lượng chứng khoán thời kỳ gốc

Chương 2

Cơ sở lý thuyết của mô hình ARIMA

2.1 Tổng quan về bài toán dự báo

2.1.1 Khái niệm

Dự báo là quá trình đưa ra dự đoán về tương lai dựa trên dữ liệu quá khứ và hiện tại cùng với việc phân tích các xu hướng. Dự báo liên quan đến việc lấy các mô hình phù hợp với dữ liệu lịch sử và sử dụng chúng để dự đoán các quan sát trong tương lai.[8]

2.1.2 Quy trình dự báo chung

Theo Wilson và Keating, quy trình dự báo gồm 9 bước:[10]

1. Xác định mục tiêu

- Kết quả dự báo sẽ được sử dụng như thế nào trong việc ra quyết định
- Thống nhất giữa người tiến hành dự báo và người sử dụng kết quả dự báo

2. Xác định đối tượng dự báo

- Xác định đối tượng hay biến dự báo cụ thể (đo bằng gì?)
- Phạm vi (một sản phẩm hay nhóm sản phẩm, thị trường trong nước, khu vực hay xuất khẩu)

3. Xác định thời đoạn dự báo

- Dự báo ngắn hạn, trung hạn hay dài hạn
- Ví dụ:

GDP	Năm
Kim ngạch XNH	Năm hoặc Quý
Doanh số	Quý

4. Thu thập, khảo sát dữ liệu

- Chất lượng dự báo phụ thuộc vào chất lượng của dữ liệu
- Dữ liệu có thể thu thập từ các nguồn nội bộ hoặc bên ngoài
- Chuyển đổi dữ liệu tương thích với yêu cầu của mục đích nếu có đầy đủ cơ sở

5. Chọn mô hình

- Việc chọn mô hình tùy thuộc vào các tiêu chí sau:
 - Dạng phân bố của dữ liệu
 - Số lượng quan sát sẵn có
 - Độ dài của thời đoạn dự báo
- Nếu dữ liệu phân bố thể hiện tính xu hướng thì có thể áp dụng các phương pháp bình quân di động, san bằng số mũ
- Nếu số lượng quan sát quá ít: không thể sử dụng mô hình hồi quy

6. Đánh giá mô hình

- Thông qua các thông số: Độ lệch tuyệt đối trung bình (MAD), sai số bình phương trung bình (MSE),...
- Nếu không mô hình nào cho kết quả tới một độ chính xác chấp nhận được, quay về bước 5 để chọn mô hình thay thế

7. Chuẩn bị dự báo

Sau khi qua bước 6, chúng ta tin rằng mô hình được chọn sẽ cho kết quả chuẩn xác. Có trường hợp phải chọn phương pháp kết hợp giữa định tính và định lượng

8. Trình bày dự báo

- Trình bày rõ ràng, dù là văn bản hay truyền đạt
- Không cần phải phức tạp hóa kết quả bằng các công thức hay thuật toán phức tạp

9. Theo dõi kết quả

- Liên tục theo dõi, so sánh kết quả dự báo với giá trị thực tế
- Thay đổi phương pháp khi cần

2.2 Các khái niệm cơ bản của chuỗi thời gian

2.2.1 Khái niệm và các thành phần của một chuỗi thời gian

Chuỗi thời gian là một chuỗi các phép đo được thực hiện theo thời gian, thường thu được ở các khoảng cách đều nhau, có thể là hàng ngày, hàng tháng, hàng quý hoặc hàng năm. Phân tích chuỗi thời gian bao gồm các phương pháp để phân tích dữ liệu chuỗi thời gian để trích xuất số liệu thống kê có ý nghĩa và các đặc điểm khác của dữ liệu. Dự báo chuỗi thời gian là việc sử dụng một mô hình để dự đoán các giá trị trong tương lai dựa trên các giá trị được quan sát trước đó.

Nói cách khác, chuỗi thời gian là một chuỗi các điểm dữ liệu được ghi lại tại các thời điểm cụ thể.

Một số ví dụ về chuỗi thời gian có thể là:

- Nhiệt độ không khí hàng ngày hoặc lượng mưa hàng tháng ở Bangalore, Ấn Độ
- Lưu lượng dòng chảy hàng năm của sông Ganga tại Patna
- Dữ liệu dân số Ấn Độ hàng năm
- Giá cổ phiếu đóng cửa hàng ngày
- Lãi suất hàng tuần

Một chuỗi thời gian bao gồm 4 thành phần:[9]

- **Trend** (Xu hướng): cho thấy xu hướng chung của dữ liệu là tăng hoặc giảm hoặc giảm trong một khoảng thời gian dài.
- **Seasonal Variations** (Biến đổi theo mùa): Các mùa hoặc điều kiện khí hậu khác nhau đóng một vai trò quan trọng trong sự thay đổi theo mùa. Chẳng hạn như sản xuất cây trồng phụ thuộc vào mùa vụ, việc bán ô và áo mưa vào mùa mưa, việc bán quạt điện tăng lên vào mùa hè.
- **Cyclic Variations** (Các biến thể theo chu kỳ): một chu kỳ gồm 4 giai đoạn là thịnh vượng, suy thoái, đình trệ và phục hồi
- **Random or Irregular movements - Residual** (Chuyển động ngẫu nhiên hoặc bất thường - Phần dư): là những biến động không lường trước được, không thể kiểm soát được, không thể đoán trước và rất thất thường như động đất, chiến tranh, lũ lụt, nạn đói và bất kỳ thảm họa nào khác.

2.2.2 Khái niệm dừng yếu

Định nghĩa 2.1. Cho một chuỗi thời gian $\{X_t\}$ có $E(X_t^2) < \infty$. Khi đó:

i) $\{X_t\}$ có hàm trung bình là: $\mu_X(t) = E(X_t)$.

ii) Hàm hiệp phương sai của $\{X_t\}$ là:

$$\gamma_X(r, s) = Cov(X_r, X_s) = E[(X_r - \mu_X(r))(X_s - \mu_X(s))]$$

với mọi số nguyên r, s .

Định nghĩa 2.2. Cho một chuỗi thời gian $\{X_t\}$ có $E(X_t^2) < \infty$. Khi đó, $\{X_t\}$ được gọi là dừng yếu nếu hai điều kiện sau đây đồng thời được thỏa mãn:

i) $\mu_X(t)$ không phụ thuộc vào t .

ii) $\gamma_X(t + h, t)$ không phụ thuộc vào $t \forall h$.

Nhận xét 2.1. Hàm hiệp phương sai trong điều kiện thứ hai của **Định nghĩa 2.2** của một chuỗi thời gian dừng $\{X_t\}$ được xác định bởi hàm một biến:

$$\gamma_X(h) = \gamma(h, 0) = \gamma_X(t + h, t)$$

Khi đó, $\gamma_X(h)$ được gọi là hàm tự hiệp phương sai ACVF với giá trị tại độ trễ h bằng $\gamma_X(h)$.

2.2.3 Hàm tự tương quan ACF

Định nghĩa 2.3 Cho một chuỗi thời gian dừng $\{X_t\}$. Khi đó, tại độ trễ h :

- $\{X_t\}$ có hàm tự hiệp phương sai ACVF được xác định bởi:

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

- $\{X_t\}$ có hàm tự tương quan ACF được xác định bởi:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

Tính chất 2.1. Tính chất tuyến tính của hiệp phương sai

Với các số thực tùy ý a, b, c , nếu $E(X^2) < \infty, E(Y^2) < \infty, E(Z^2) < \infty$ thì:

$$Cov(aX + bY + c, Z) = aCov(X, Z) + bCov(Y, Z)$$

Tính chất 2.2. Một số tính chất cơ bản của $\gamma(\cdot)$

- $\gamma(0) \geq 0$.
- $|\gamma(h)| \leq \gamma(0) \forall h$.
- $\gamma(\cdot)$ là một hàm chẵn, có nghĩa là $\gamma(h) = \gamma(-h) \forall h$.

2.2.4 Phương pháp khử thành phần xu hướng và thành phần mùa

$\{X_t\}$ là một chuỗi thời gian và mô hình cổ điển có dạng:

$$X_t = m_t + s_t + Y_t \quad t = 1, \dots, n$$

trong đó:

m_t là thành phần xu hướng

s_t là thành phần theo mùa với $s_t = s_{t+d}$ (d là chu kỳ đã biết) và $\sum_{j=1}^d s_j = 0$

Y_t là thành phần ngẫu nhiên với $EY_t = 0$

Giả sử có n quan sát x_1, \dots, x_n

***Tách thành phần xu hướng**

TH1: d lẻ, $d = 2q + 1 \Rightarrow q = \frac{d-1}{2}$

$$\hat{m}_t = \frac{1}{d} \sum_{j=-q}^q x_j$$

TH2: d chẵn, $d = 2q \Rightarrow q = \frac{d}{2}$

$$\hat{m}_t = \frac{1}{d} \left(0.5x_{t-q} + \sum_{j=t-q+1}^{t+q-1} x_j + 0.5x_{t+q} \right) \quad q < t \leq n - q$$

\Rightarrow Phân tách xu hướng: $g_t = x_t - \hat{m}_t$

***Tách thành phần mùa**

$l = \left\lfloor \frac{n}{d} \right\rfloor$, $q < t = k + jd \leq n - q$ ($k = 1, \dots, d$)

$$w_k = \frac{1}{l} \sum_{j=0}^{l-1} (x_{k+jd} - \hat{m}_{k+jd})$$

$$\hat{s}_k = w_k - \frac{1}{d} \sum_{i=1}^d w_i$$

$$\rightarrow \hat{s}_t = w_t - \frac{1}{d} \sum_{i=1}^d w_{i+t-k}$$

\Rightarrow Phân tách mùa: $d_t = x_t - \hat{s}_t$

2.2.5 Quá trình tuyến tính

Định nghĩa 2.4 Chuỗi thời gian $\{X_t\}$ là một quá trình tuyến tính nếu nó được biểu diễn dưới dạng:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \quad \forall t$$

trong đó: $\{Z_t\} \sim WN(0, \sigma^2)$ và $\{\psi_j\}$ là chuỗi các hằng số với $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

Về phương diện toán tử dịch chuyển lùi B, ta có:

$$X_t = \psi(B)Z_t$$

ở đó, $\psi(B) = \sum_{j=-\infty}^{+\infty} \psi_j B^j$.

Mệnh đề 2.1.

- Giả sử rằng một chuỗi thời gian dừng $\{Y_t\}$ có trung bình bằng 0 và hàm hiệp phương sai γ_Y . Nếu $\sum_{j=-\infty}^{+\infty} \gamma_Y(j) < \infty$ thì chuỗi thời gian:

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j Y_{t-j} = \psi(B)Y_t$$

là dừng với trung bình bằng 0 và hàm tự hiệp phương sai:

$$\gamma_X(h) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \psi_j \psi_k \gamma_Y(h + k - j)$$

- Đặc biệt, nếu $\{X_t\}$ là một quá trình tuyến tính thì:

$$\gamma_X(h) = \sum_{j=-\infty}^{+\infty} \psi_j \psi_{j+h} \sigma^2$$

Nhận xét 2.2. Xét hai bộ lọc $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$ và $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$.

- Một chuỗi thời gian dừng mới được tạo bởi chuỗi dừng $\{Y_t\}$ có dạng:

$$W_t = \sum_{j=-\infty}^{+\infty} \psi_j Y_{t-j}$$

ở đó, $\psi_j = \sum_{k=-\infty}^{+\infty} \alpha_k \beta_{j-k} = \sum_{k=-\infty}^{+\infty} \beta_k \alpha_{j-k}$

- Về phương diện của toán tử dịch chuyển lùi B, ta có: $W_t = \psi(B)Y_t$, trong đó, $\psi(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B)$.

2.3 Một số mô hình chuỗi thời gian dừng

2.3.1 Chuỗi iid noise và chuỗi White noise

Cho chuỗi $\{X_t\}$ gồm một dãy các biến ngẫu nhiên X_1, X_2, X_3, \dots

*** Chuỗi iid noise**

Định nghĩa 2.5. Chuỗi $\{X_t\}$ là chuỗi iid noise nếu X_1, X_2, X_3, \dots là các biến ngẫu nhiên độc lập, cùng phân phối và có trung bình bằng 0.

Chuỗi iid noise là chuỗi thời gian đơn giản nhất không chứa các thành phần xu hướng và thời vụ.

Với $E(X_t^2) = \sigma^2 < \infty$ và $E(X_t) = 0 \forall t$ nên hàm tự hiệp phương sai ACVF của $\{X_t\}$ là:

$$\gamma_X(t+h, t) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

không phụ thuộc vào t .

“Kí hiệu: $\{X_t\} \sim IID(0, \sigma^2)$ để chỉ các biến ngẫu nhiên X_t độc lập, cùng phân phối, mỗi biến có trung bình bằng 0 và phương sai σ^2 . [4, tr. 16]”

*** Chuỗi White noise**

Định nghĩa 2.6. Chuỗi $\{X_t\}$ được gọi là nhiễu trắng (White noise) nếu X_1, X_2, X_3, \dots là một dãy các biến ngẫu nhiên không tương quan, có trung bình bằng 0 và phương sai σ^2 .

Kí hiệu: $\{X_t\} \sim WN(0, \sigma^2)$

Chuỗi X_t là chuỗi dừng với hàm tự hiệp phương sai giống với chuỗi iid noise.

Chuỗi iid noise là nhiễu trắng nhưng nhiễu trắng chưa chắc đã là iid noise.

2.3.2 Quá trình trung bình trượt MA(q)

Định nghĩa 2.7. $\{X_t\}$ được gọi là một quá trình trung bình trượt với bậc q nếu:

$$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

với $\{Z_t\} \sim WN(0, \sigma^2)$ và $\theta_1, \dots, \theta_q$ là các hằng số.

Theo thuật ngữ của toán tử dịch chuyển lùi, quá trình MA(q) là:

$$X_t = Z_t(1 + \theta_1 B + \dots + \theta_q B^q) = \theta(B)Z_t$$

với $\theta(B) = 1 + \sum_{i=1}^q \theta_i B^i$.

Do $\{Z_t\} \sim WN(0, \sigma^2)$ nên giá trị trung bình của quá trình MA(q) là:

$$E(X_t) = E(Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}) = 0$$

và phương sai là:

$$Var(X_t) = E(X_t^2) - E(X_t)^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2)$$

Suy ra, hàm tự hiệp phương sai sẽ bằng:

$$\begin{aligned} \gamma_X(h) &= Cov(X_{t+h}, X_t) = E(X_{t+h}X_t) \\ &= E[(Z_{t+h} + \theta_1 Z_{t+h-1} + \dots + \theta_q Z_{t+h-q})(Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q})] \\ &= \begin{cases} \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2), & h = 0 \\ \sigma^2(\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q), & h = 1, 2, \dots, q \\ 0, & h > q \end{cases} \end{aligned}$$

Do đó, hàm tự tương quan của MA(q) là:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = \begin{cases} 1, & h = 0 \\ \frac{\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & h = 1, 2, \dots, q \\ 0, & h > q \end{cases}$$

Vì $E(X_t^2) = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2) < \infty$, mà trung bình và hàm tự hiệp phương sai ACVF không phụ thuộc vào t nên MA(q) là quá trình dừng.

2.3.3 Quá trình tự hồi quy AR(p)

Định nghĩa 2.8. Quá trình tự hồi quy bậc p, viết tắt là AR(p), được định nghĩa như sau:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t,$$

với $\{Z_t\} \sim WN(0, \sigma^2)$ và ϕ_1, \dots, ϕ_p là các hằng số, $|\phi_i| < 1$ ($i = \overline{1, p}$) và Z_s không tương quan với X_s ($s < t$).

Theo thuật ngữ của toán tử dịch chuyển lùi, quá trình AR(p) là:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t = Z_t \Leftrightarrow \phi(B)X_t = Z_t$$

ở đó, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 1 - \sum_{i=1}^p \phi_i B^i$

Để tìm hàm tự tương quan của AR(p), thực hiện nhân hai vế của phương trình với X_{t-h} ($h > 0$) và lấy kỳ vọng hai vế ta được:

$$E(X_{t-h}X_t) - E(\phi_1 X_{t-h}X_{t-1}) - \dots - E(\phi_p X_{t-h}X_{t-p}) = E(X_{t-h}Z_t)$$

Vì Z_t không tương quan với $X_s \forall s < t$ nên $E(X_{t-h}Z_t) = 0$, do đó ta có:

$$\gamma_h = \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2} + \dots + \phi_p \gamma_{h-p} \quad (h > 0)$$

Chia cả hai vế cho $\gamma(0)$ ta được hàm tự tương quan thỏa mãn phương trình:

$$\rho_h = \phi_1 \rho_{h-1} + \phi_2 \rho_{h-2} + \dots + \phi_p \rho_{h-p} \quad (h > 0)$$

2.3.4 Quá trình ARMA(p, q)

Định nghĩa 2.9. Chuỗi thời gian $\{X_t\}$ là một quá trình ARMA(p, q) nếu nó là **chuỗi thời gian dừng** và $\forall t$ thì:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \quad (1)$$

với $\{Z_t\} \sim WN(0, \sigma^2)$.

Gọi $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ là đa thức có bậc p và $\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ là đa thức có bậc q (z là một số phức).

Khi đó, trên phương diện của toán tử lùi B, phương trình (1) có dạng:

$$\phi(B)X_t = \theta(B)Z_t$$

Một quá trình ARMA(p, q) là quá trình AR(p) nếu $\theta(z) \equiv 1$, hoặc là quá trình MA(q) nếu $\phi(z) \equiv 1$.

* Sự tồn tại và duy nhất nghiệm

Phương trình (1) tồn tại và có một nghiệm dừng $\{X_t\}$ duy nhất khi và chỉ khi:

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \quad \forall |z| = 1$$

* Tính nhân quả

Định nghĩa 2.10. Một quá trình ARMA(p, q) được gọi là nhân quả hoặc hàm nhân quả của $\{Z_t\}$ nếu chuỗi thời gian $\{X_t\}$ có thể được biểu diễn dưới dạng quá trình tuyến tính một phía như sau:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} = \psi(B)Z_t \quad \forall t$$

ở đó, $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$ và $\sum_{j=0}^{\infty} |\psi_j| < \infty$, đặt $\psi_0 = 1$.

Tính chất 2.3. Tính nhân quả của một quá trình ARMA(p, q)

Một quá trình ARMA(p, q) là nhân quả khi và chỉ khi $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \quad \forall |z| \leq 1$. Các hệ số ψ_j có thể được xác định bằng cách giải:

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

* Tính khả nghịch

Định nghĩa 2.11. Một quá trình ARMA(p, q) là khả nghịch nếu chuỗi thời gian $\{X_t\}$ có thể được biểu diễn dưới dạng:

$$\pi(B)X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = Z_t \quad \forall t$$

ở đó, $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$ và $\sum_{j=0}^{\infty} |\pi_j| < \infty$, đặt $\pi_0 = 1$.

Tính chất 2.4. Tính nghịch đảo của một quá trình ARMA(p, q)

Một quá trình ARMA(p, q) là nghịch đảo khi và chỉ khi $\theta(z) = 1 + \theta_1 z - \dots - \theta_q z^q \neq 0 \quad \forall |z| \leq 1$. Các hệ số π_j có thể được xác định bằng cách giải:

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1.$$

2.4 Chuỗi thời gian không dừng và mô hình ARIMA

2.4.1 Mô hình ARIMA cho chuỗi thời gian không dừng

Định nghĩa 2.12. “Nếu d là một số nguyên không âm, thì $\{X_t\}$ là một quá trình ARIMA(p, d, q) nếu $Y_t = (1 - B)^d X_t$ là một quá trình ARMA(p, q) nhân quả” [4, tr. 180].

Khi đó, $\{X_t\}$ thỏa mãn phương trình sai phân:

$$\phi^*(B) = \phi(B)(1 - B)^d X_t = \theta(B)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2)$$

với $\phi(z)$ là đa thức bậc p, $\theta(z)$ là đa thức bậc q và $\phi(z) \neq 0$ với $|z| \leq 1$.

Đa thức $\phi^*(z)$ có bậc d = 0 tại z = 1. Quá trình $\{X_t\}$ dừng khi và chỉ khi d = 0, lúc này $\{X_t\}$ chính là quá trình ARMA(p, q).

Nếu $d \geq 1$, có thể thêm một xu hướng được biểu diễn bởi một đa thức tùy ý có bậc (d - 1) vào $\{X_t\}$, điều này không bị vi phạm phương trình sai phân ở trên. Quá trình ARIMA(p, d, q) có thể phù hợp cho cả mô hình chuỗi thời gian có xu hướng và không có xu hướng.

* Một số trường hợp đặc biệt quan trọng của mô hình ARIMA:

- Quá trình ARIMA(0, 1, 1):

$$(1 - B)X_t = Z_t + \theta_1 Z_{t-1} = (1 + \theta_1 B)Z_t$$

tương ứng với $p = 0, d = 1, q = 1, \phi(B) = 1, \theta(B) = 1 + \theta_1 B$.

- Quá trình ARIMA(0, 2, 2):

$$(1 - B)^2 X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} = (1 + \theta_1 B + \theta_2 B^2) Z_t$$

tương ứng với $p = 0, d = 2, q = 2, \phi(B) = 1, \theta(B) = 1 + \theta_1 B + \theta_2 B^2$.

- Quá trình ARIMA(1, 1, 1):

$$(1 - B)X_t - \phi_1(1 - B)Z_{t-1} = Z_t + \theta_1 Z_{t-1} \Leftrightarrow (1 - \phi_1 B)(1 - B)X_t = (1 - \theta_1 B)Z_t$$

tương ứng với $p = 1, d = 1, q = 1, \phi(B) = 1 - \phi_1 B, \theta(B) = 1 + \theta_1 B$.

2.4.2 Mô hình ARIMA có tính mùa

Định nghĩa 2.13. “Nếu d và D là các số nguyên không âm, thì $\{X_t\}$ là một quá trình $ARIMA(p, d, q) \times (P, D, Q)_s$ có tính mùa với chu kỳ s nếu chuỗi sai phân $Y_t = (1 - B)^d(1 - B^s)^D X_t$ là một quá trình ARMA nhân quả được xác định bởi:

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t, \quad \{Z_t\} \sim WN(0, \sigma^2)$$

ở đó, $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$

$\Phi(z) = 1 - \Phi_1 z - \dots - \Phi_p z^p$

$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$

$\Theta(z) = 1 + \Theta_1 z + \dots + \Theta_q z^q$ ” [4, tr. 203].

Nhận xét 2.3. Quá trình $\{Y_t\}$ là nhân quả khi và chỉ khi $\phi(z) \neq 0, \Phi(z) \neq 0 \quad \forall |z| \leq 1$. Trong nhiều mô hình, D hiếm khi lớn hơn một và P, Q thường nhỏ hơn 3.

Nhận xét 2.4.

2.4.3 Phương pháp Box - Jenkins

Phương pháp Box - Jenkins bao gồm các khâu chính sau:

- **Bước 1 - Nhận dạng mô hình ARIMA(p, d, q):** Một mô hình ARIMA(p, d, q) có thể được nhận dạng thông qua các dữ liệu lịch sử.

- **Bước 2 - Ước lượng cho các tham số:** Các tham số của một mô hình thử nghiệm sẽ được ước lượng dựa trên dữ liệu quá khứ.

- **Bước 3 - Kiểm tra chẩn đoán:** Mô hình thử nghiệm sẽ được kiểm tra sự phù hợp thông qua các đánh giá khác nhau. Nếu cần thiết, có thể gợi ý một mô hình khác tốt hơn, và mô hình này sẽ được xem xét như một mô hình mới cần thử nghiệm.

- **Bước 4 - Dự báo:** Sau khi đã thực hiện kiểm tra các mô hình và chọn ra được mô hình cuối cùng phù hợp nhất, có thể đưa mô hình này vào dự báo cho các giá trị trong tương lai.

Chương 3

Ứng dụng mô hình ARIMA trong dự báo chỉ số giá chứng khoán

3.1 Chuẩn bị mẫu dữ liệu

3.1.1 Giới thiệu bộ dữ liệu

Bộ dữ liệu về lịch sử chứng khoán của Ngân hàng Thương mại cổ phần Ngoại thương Việt Nam (Vietcombank - VCB) từ 04/01/2010 - 28/01/2022. Link dữ liệu:

<https://vn.investing.com/equities/joint-stock-commercial-bank-historical-data>

Dữ liệu Lịch sử VCB

Khung Thời Gian:

Hàng ngày



Tải Dữ Liệu Xuống

01/01/2010 - 06/02/2022



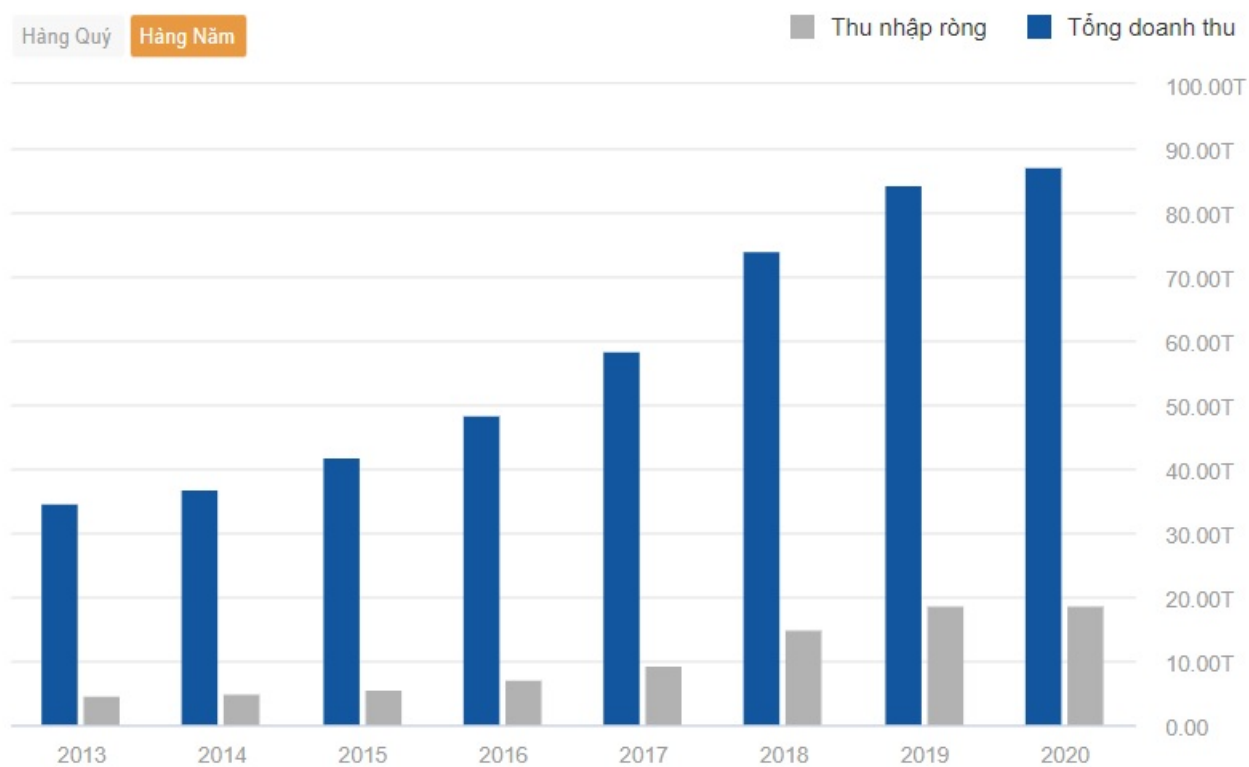
Ngày	Lần cuối	Mở	Cao	Thấp	KL	% Thay đổi
28/01/2022	89,000	91,000	92,000	89,000	2.06M	-2.20%
27/01/2022	91,000	93,000	94,500	90,400	1.73M	-3.70%
26/01/2022	94,500	95,800	95,800	92,600	2.52M	-1.36%
25/01/2022	95,800	92,500	96,000	89,000	2.40M	3.01%
24/01/2022	93,000	89,400	93,500	88,200	2.73M	4.26%
21/01/2022	89,200	86,900	90,000	86,100	2.50M	2.41%
20/01/2022	87,100	85,200	87,500	83,100	2.79M	0.35%
19/01/2022	86,800	87,500	87,500	85,000	914.80K	-0.91%
18/01/2022	87,600	85,500	87,600	84,000	1.70M	1.62%
17/01/2022	86,200	83,000	88,000	82,300	2.63M	3.36%
14/01/2022	83,400	81,900	83,400	81,000	1.52M	1.83%
13/01/2022	81,000	80,500	81,400	80,200	2.52M	2.28%

Hình 1. Bảng dữ liệu lịch sử chứng khoán của Vietcombank

3.1.2 Phân tích dữ liệu khám phá



Hình 2. Chỉ số chứng khoán của Vietcombank trong 5 năm gần đây



Hình 3. Biểu đồ thu nhập của Vietcombank

Trong 5 năm qua chỉ số chứng khoán của Vietcombank đang có xu hướng tăng. Vào khoảng tháng 7/2018 - 6/2019 là thời điểm cổ phiếu ngân hàng bị rớt giá do ảnh hưởng của dịch bệnh COVID-19. Nhưng sau cuộc khủng hoảng, ngành ngân hàng đã vực mình dậy và giá lại tiếp tục leo thang vào khoảng tháng 6/2021 chỉ số chứng khoán đã đạt mức kỷ lục là khoảng 90,000 và đã giảm mạnh sau đó nhưng sau đó đến 1/2022 chỉ số đã tăng trở lại và tương đối ổn định. Chính vì vậy, tổng doanh thu của Vietcombank luôn tăng và năm 2020 đã đạt 86.87 tỷ VNĐ.

3.2 Xây dựng mô hình dự báo chỉ số giá chứng khoán

3.2.1 Nhận dạng mô hình ARIMA(p, d, q)

Khai báo thư viện và trích xuất dữ liệu chứng khoán:

```
In [108]: import warnings
import itertools
import pandas as pd
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
df=pd.read_csv('vietcombank.csv')
df['Date'] = pd.to_datetime(df['Date'])
df = df.set_index('Date')
df
```

Out[108]:

	Index
Date	
2010-01-04	13465
2010-01-05	13602
2010-01-06	13110
2010-01-07	12837
2010-01-08	12837
2010-01-09	12837
2010-01-10	12837
2010-01-11	12837
2010-01-12	12427

Vẽ đồ thị chuỗi thời gian:

```
df.plot(figsize=(15, 6))
plt.show()
```



Tách phần Trend, Seasonal, Residual:

```
from statsmodels.tsa.seasonal import seasonal_decompose
decomp = seasonal_decompose(x=df, model='additive')
est_trend = decomp.trend
est_seasonal = decomp.seasonal
est_residual = decomp.resid
fig, axes = plt.subplots(4, 1)
fig.set_figheight(10)
fig.set_figwidth(15)
axes[0].plot(df, label='Original')
axes[0].legend()
axes[1].plot(est_trend, label='Trend', color="b")
axes[1].legend()
axes[2].plot(est_seasonal, label='Seasonality', color='r')
axes[2].legend()
axes[3].plot(est_residual, label='Residuals', color='g')
axes[3].legend()
```



3.2.2 Ước lượng các tham số của mô hình

Ta sử dụng mô hình ARIMA theo mùa, được ký hiệu là $ARIMA(p,d,q)(P,D,Q,S)$. Ở đây, (p,d,q) là các tham số không theo mùa, còn (P,D,Q) theo cùng một định nghĩa với (p,d,q) nhưng được áp dụng cho thành phần theo mùa của chuỗi thời gian, S là chu kỳ của chuỗi thời gian (4 - quý, 12 - năm,...). Tìm các tham số của mô hình $ARIMA(p,d,q)(P,D,Q,S)$:

```
# Xác định các tham số p, d và q nhận bất kỳ giá trị nào trong khoảng từ 0 đến 2
p = d = q = range(0, 2)
# Tạo tất cả các kết hợp khác nhau của bộ ba p, q và q
pdq = list(itertools.product(p, d, q))
# Tạo tất cả các kết hợp khác nhau của bộ ba p, q và q theo mùa
seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]
print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))
```

```
Examples of parameter combinations for Seasonal ARIMA...
SARIMAX: (0, 0, 1) x (0, 0, 1, 12)
SARIMAX: (0, 0, 1) x (0, 1, 0, 12)
SARIMAX: (0, 1, 0) x (0, 1, 1, 12)
SARIMAX: (0, 1, 0) x (1, 0, 0, 12)
```


Để chọn được bộ ba tham số phù hợp ta sử dụng giá trị AIC để đo lường mức độ phù hợp của một mô hình với dữ liệu trong khi tính đến độ phức tạp tổng thể của mô hình. Do đó, ta cần phải tìm kiếm mô hình mang lại giá trị AIC thấp nhất. Tính AIC tương ứng với các tham số của mô hình:

```
warnings.filterwarnings("ignore")

for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(df, order=param, seasonal_order=param_seasonal,
                                             enforce_stationarity=False, enforce_invertibility=False)
            results = mod.fit()
            print('ARIMA{0}{1}{2} - AIC:{0}'.format(param, param_seasonal, results.aic))
        except:
            continue
```

```
ARIMA(0, 1, 1)x(0, 1, 0, 12) - AIC:71684.99279086998
ARIMA(0, 1, 1)x(0, 1, 1, 12) - AIC:68649.96220047477
ARIMA(0, 1, 1)x(1, 0, 0, 12) - AIC:68808.04680531674
ARIMA(0, 1, 1)x(1, 0, 1, 12) - AIC:68780.12458250842
ARIMA(0, 1, 1)x(1, 1, 0, 12) - AIC:70304.35942613718
ARIMA(0, 1, 1)x(1, 1, 1, 12) - AIC:68651.56151030216
ARIMA(1, 0, 0)x(0, 0, 0, 12) - AIC:68981.18575726163
ARIMA(1, 0, 0)x(0, 0, 1, 12) - AIC:68804.48712197739
ARIMA(1, 0, 0)x(0, 1, 0, 12) - AIC:71538.52340215625
ARIMA(1, 0, 0)x(0, 1, 1, 12) - AIC:68689.51844573769
ARIMA(1, 0, 0)x(1, 0, 0, 12) - AIC:68804.46726006188
ARIMA(1, 0, 0)x(1, 0, 1, 12) - AIC:68806.41901438916
ARIMA(1, 0, 0)x(1, 1, 0, 12) - AIC:70200.51795655343
ARIMA(1, 0, 0)x(1, 1, 1, 12) - AIC:69384.35473395934
ARIMA(1, 0, 1)x(0, 0, 0, 12) - AIC:68968.48174697117
ARIMA(1, 0, 1)x(0, 0, 1, 12) - AIC:68791.82807564377
ARIMA(1, 0, 1)x(0, 1, 0, 12) - AIC:71523.10130482464
ARIMA(1, 0, 1)x(0, 1, 1, 12) - AIC:69367.74075030185
```

Ta thấy ARIMA(0, 1, 1)x(1, 1, 1, 12) mang lại giá trị AIC thấp nhất là 68651,56 nên đây là lựa chọn tối ưu nhất.

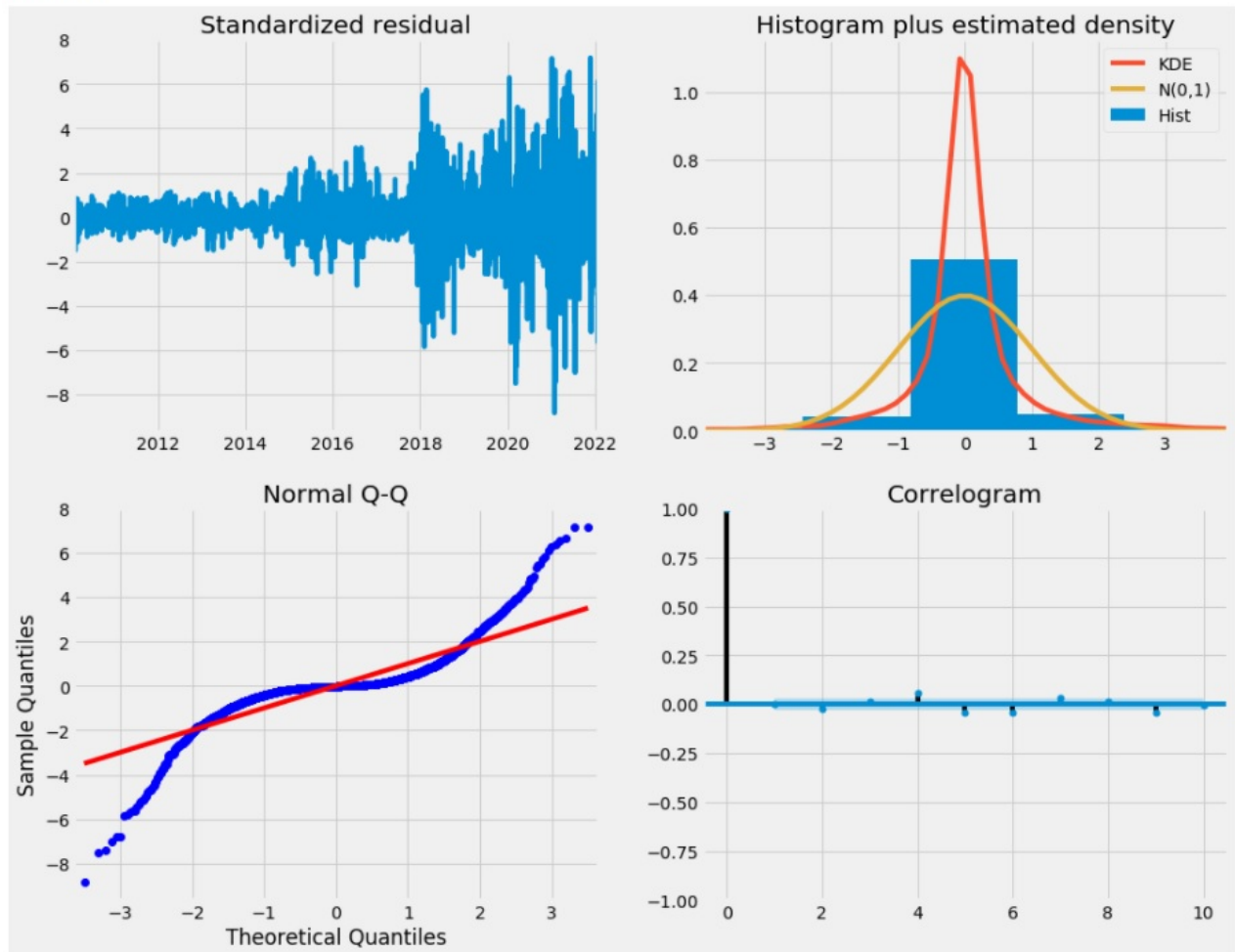
3.2.3 Kiểm tra chẩn đoán

```
mod = sm.tsa.statespace.SARIMAX(df, order=(0, 1, 1), seasonal_order=(1, 1, 1, 12),
                                enforce_stationarity=False, enforce_invertibility=False)
results = mod.fit()
print(results.summary().tables[1])
```

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	0.0012	0.008	0.156	0.876	-0.014	0.016
ar.S.L12	0.0094	0.009	1.023	0.306	-0.009	0.027
ma.S.L12	-0.9901	0.002	-434.550	0.000	-0.995	-0.986
sigma2	3.697e+05	2873.217	128.679	0.000	3.64e+05	3.75e+05

Ở đây ta chỉ xét 2 cột: cột coef hiển thị trọng số (tức là tầm quan trọng) của từng tính năng và cột P>|z| cho ta biết tầm quan trọng của từng trọng số đối tượng địa lý.

```
results.plot_diagnostics(figsize=(15, 12))
plt.show()
```



Cần phải đảm bảo phần dư của mô hình là các giá trị không tương quan và được phân phối bình thường với giá trị trung bình bằng 0. Nếu mô hình ARIMA theo mùa không đáp ứng các đặc tính này, thì đó là một dấu hiệu tốt cho thấy nó có thể được cải thiện hơn nữa.

Chẩn đoán mô hình ở đây có phần dư không được phân phối bình thường và có thể được cải thiện hơn nữa là do: Trong biểu đồ trên cùng bên phải, ta thấy rằng đường KDE màu đỏ không theo sát với đường $N(0,1)$ (trong đó $N(0,1)$ là ký hiệu chuẩn cho phân phối chuẩn với giá trị trung bình là 0 và độ lệch chuẩn là 1). Đây là một dấu hiệu cho thấy phần dư không được phân phối bình thường.

3.2.4 Ứng dụng mô hình dự báo chỉ số giá chứng khoán

```
# Nhận dự báo trước 500 bước trong tương lai
pred_uc = results.get_forecast(steps=500)
# Nhận khoảng tin cậy của các dự báo
pred_ci = pred_uc.conf_int()
```

```
ax = df.plot(label='observed', figsize=(20, 15))
pred_uc.predicted_mean.plot(ax=ax, label='Forecast')
ax.fill_between(pred_ci.index, pred_ci.iloc[:, 0], pred_ci.iloc[:, 1], color='k', alpha=.25)
ax.set_xlabel('Date')
ax.set_ylabel('Index')
plt.legend()
plt.show()
```



Dự báo cho thấy chuỗi thời gian dự kiến sẽ tiếp tục tăng với tốc độ ổn định và khi dự báo xa hơn về tương lai thì các khoảng tin cậy do mô hình tạo ra ngày càng lớn hơn.

Kết luận

Qua thời gian nghiên cứu về đề tài này, nhóm em đã nắm được quy trình xây dựng mô hình ARIMA cho dữ liệu chứng khoán và áp dụng mô hình này vào bài toán thực tế - bài toán dự báo chứng khoán của Vietcombank. Những kết quả mà nhóm chúng em đã đạt được có thể tổng kết như sau:

- Hiểu được một số khái niệm liên quan đến chứng khoán như: khái niệm chứng khoán, các loại chứng khoán hiện nay, chỉ số chứng khoán, một số phương pháp ước lượng chỉ số giá chứng khoán,...
- Cơ sở lý thuyết về chuỗi thời gian, mô hình ARIMA
- Đưa ra được đồ thị dự báo xu hướng của chỉ số chứng khoán của Vietcombank.

Bên cạnh những kết quả đã đạt được, nhóm vẫn còn một số vấn đề mà thời điểm này vẫn chưa giải quyết được:

- Đây chỉ là mô hình phân tích kỹ thuật, chưa thể dự báo một cách chính xác, bởi chỉ phụ thuộc vào một biến – thời gian, trong khi quá trình dự báo phụ thuộc vào nhiều yếu tố.
- Thuật toán để ước lượng cũng như đánh giá còn nhiều hạn chế.
- Cần phải xây dựng mô hình ARIMA đa biến: chỉ số của giá chứng khoán phụ thuộc vào nhiều biến khác nhau.

Tài liệu tham khảo

- [1] Bạch Đức Hiền, *Giáo trình Thị trường chứng khoán*, NXB Tài Chính, 2009.
- [2] Nguyễn Thị Vinh, *Phân tích chuỗi thời gian và các kỹ thuật dự báo*, Đại học Thủy Lợi, 2010.
- [3] P. Boye, Y. Y. Ziggah, “A Short-Term Stock Exchange Prediction Model Using Box-Jenkins Approach”, *Journal of Applied Mathematics and Physics*, 8, 766-779, 2020.
- [4] Peter J. B, Richard A. D, *Introduction to Time Series and Forecasting, Second Edition*, Springer-Verlag New York, Inc., 2002.
- [5] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Jung, *Time Series Analysis: Forecasting and Control*, Fifth Edition, John Wiley Sons, Inc., 2015.
- [6] Robert H. S, David S.S, *Time Series Analysis and Its Applications: With E Examples*, Fourth Edition, Springer-Verlag, 2016.
- [7] Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci, *Introduction to Time Series Analysis and Forecasting*, John Wiley Sons, Inc., 2008.
- [8] Bùi Kim Yến, *Phân tích và đầu tư chứng khoán*, NXB Thống Kê, 2008.
- [9] Nguyễn Thị Minh Huệ, Lê Thị Hương Lan, *Giáo trình thị trường chứng khoán*, NXB Đại Học Kinh Tế Quốc Dân, 2019.
- [10] [Dự báo chuỗi thời gian](#)
- [11] [Components of Time Series](#)
- [12] [Quy Trình Dự Báo, Phân Tích Dữ Liệu Và Lựa Chọn Phương Pháp Dự Báo](#)