

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - CƠ - TIN HỌC



MÔN HỌC: PHÂN TÍCH CHUỖI THỜI GIAN

ĐỀ TÀI: Phân tích lợi suất trái phiếu chính phủ Hoa Kỳ và dự báo chu kỳ kinh tế

Sinh viên:

Phạm Ngọc Hải - 21002139

Nguyễn Đức Nam - 21002159

Nguyễn Văn Thắng - 21002175

Hà Nội - 2024

Mục lục

1	Giới thiệu	5
1.1	Bối cảnh nghiên cứu	5
1.2	Mục tiêu và ý nghĩa nghiên cứu	5
1.3	Phạm vi và phương pháp nghiên cứu	6
1.4	Tổng quan về chu kỳ kinh tế và lợi suất trái phiếu chính phủ	6
2	Phương pháp - Tổng quan về quy trình nghiên cứu	7
2.1	Bước 1: Thu thập và chuẩn bị dữ liệu	7
2.2	Bước 2: Phân tích sơ bộ dữ liệu	8
2.3	Bước 3: Kiểm tra đặc tính của chuỗi thời gian	8
2.4	Bước 4: Xây dựng và đánh giá mô hình dự báo	9
2.5	Bước 5: Dự báo và kiểm tra mô hình	10
3	Cơ sở lý thuyết	10
3.1	Thu thập và chuẩn bị dữ liệu	10
3.1.1	Thu thập dữ liệu	10
3.1.2	Kiểm tra độ chính xác và tính đầy đủ của dữ liệu	11
3.1.3	Xử lý dữ liệu thiếu	11
3.1.4	Làm sạch dữ liệu và loại bỏ ngoại lai	12
3.1.5	Chuẩn hóa và chuyển đổi dữ liệu	12
3.1.6	Chuẩn bị dữ liệu cho phân tích	13
3.2	Phân tích sơ bộ dữ liệu	13
3.2.1	Khám phá dữ liệu	13
3.2.2	Kiểm tra tính dừng của chuỗi thời gian	14
3.2.3	Phân tích các yếu tố mùa vụ và chu kỳ	19
3.3	Kiểm tra đặc tính của chuỗi thời gian	23
3.4	Xây dựng và đánh giá mô hình dự báo	24
3.4.1	Mô hình ARIMA	25

3.4.2	Mô hình GARCH	26
3.4.3	Mô hình học máy và học sâu: LSTM	26
3.5	Ước lượng tham số và quá trình huấn luyện mô hình	28
3.5.1	Ước lượng tham số cho mô hình ARIMA	28
3.5.2	Ước lượng tham số cho mô hình GARCH	29
3.5.3	Ước lượng tham số cho mô hình LSTM	31
3.5.4	Đánh giá mô hình	32
3.5.5	Đánh giá phần dư	33
4	Thực nghiệm	34
4.1	Khám phá dữ liệu	34
4.1.1	Xử lý dữ liệu thiếu	35
4.1.2	Chuẩn hóa dữ liệu	36
4.2	Phân chia dữ liệu	36
4.3	Phân tích xu hướng, tính mùa, và phần dư	37
4.4	Đánh giá tính dừng	39
4.5	Sai phân bậc 1	42
4.6	Cài đặt mô hình ARIMA(2,1,2)	43
4.7	Cài đặt mô hình SARIMA	45
4.8	Các nỗ lực làm mượt chuỗi thời gian	48
4.9	Setup mô hình ARCH - GARCH	51
4.10	Thiết lập mô hình LSTM	54
5	Kết luận	57

Danh sách hình vẽ

1	Cấu trúc của dữ liệu	34
2	Trực quan hóa dữ liệu theo thời gian	35
3	Dữ liệu sau chuẩn hóa	36

4	Phân tích xu hướng, tính mùa, phần dư với chu kỳ 30 ngày	37
5	Phân tích xu hướng, tính mùa, phần dư với chu kỳ 365 ngày	38
6	Phân tích xu hướng, tính mùa, phần dư với chu kỳ 10 năm	38
7	Đồ thị ACF của chuỗi thời gian ‘Adj Close’	40
8	Đồ thị PACF của chuỗi thời gian ‘Adj Close’	41
9	Kết quả sai phân bậc 1 của chuỗi thời gian	42
10	Đồ thị ACF và PACF của chuỗi thời gian sau khi sai phân bậc 1	42
11	Kết quả chẩn đoán phần dư của mô hình ARIMA(2,1,2)	45
12	Chẩn đoán phần dư của mô hình SARIMA	48
14	Dữ liệu sau khi bổ sung các cột được làm mượt	49
13	Dữ liệu sau khi bổ sung các cột được làm mượt	49
15	Ví dụ về CMA30 (Centered Moving Average 30)	50
16	Ví dụ về sai phân bậc 3 của CMA30	51
17	Quá trình lựa chọn tham số cho các mô hình ARCH và GARCH	52
18	Dự đoán của mô hình LSTM trên tập kiểm tra	55
19	Đánh giá phân phối phần dư qua biểu đồ tần suất và QQ-plot	56
20	Đánh giá tự tương quan phần dư qua ACF	56

Danh sách bảng

1	Chỉ số AIC và BIC của các mô hình	53
---	---	----

1 Giới thiệu

1.1 Bối cảnh nghiên cứu

Trong bối cảnh nền kinh tế toàn cầu đang trải qua những biến động không ngừng, việc hiểu rõ các yếu tố ảnh hưởng đến sự thay đổi của nền kinh tế là vô cùng quan trọng. Một trong những yếu tố quan trọng nhất trong việc đánh giá tình trạng kinh tế và dự báo các xu hướng trong tương lai là lợi suất trái phiếu chính phủ. Lợi suất trái phiếu chính phủ, đặc biệt là trái phiếu chính phủ Hoa Kỳ, không chỉ phản ánh tình hình tài chính của quốc gia mà còn đóng vai trò là một chỉ số kinh tế quan trọng, liên quan mật thiết đến các quyết định đầu tư, chính sách tiền tệ và các chiến lược tài chính vĩ mô.

Lợi suất trái phiếu chính phủ thường được xem như là một chỉ báo của chi phí vay mượn trong nền kinh tế. Khi lợi suất trái phiếu tăng, điều này có thể phản ánh kỳ vọng của thị trường về việc lãi suất tăng hoặc sự bất ổn trong nền kinh tế, trong khi khi lợi suất giảm có thể báo hiệu kỳ vọng của thị trường về suy thoái hoặc tình trạng lạm phát thấp. Mối quan hệ giữa lợi suất trái phiếu và chu kỳ kinh tế đã được nghiên cứu nhiều trong lý thuyết tài chính, và nghiên cứu này nhằm phân tích sự tương quan đó trong bối cảnh hiện tại, với dữ liệu từ năm 1960 đến nay.

1.2 Mục tiêu và ý nghĩa nghiên cứu

Mục tiêu chính của nghiên cứu này là phân tích chuỗi thời gian của lợi suất trái phiếu chính phủ Hoa Kỳ, với dữ liệu hàng ngày từ năm 1960 đến 2024, nhằm đưa ra cái nhìn toàn diện về sự biến động của lợi suất và mối liên hệ của nó với các chu kỳ kinh tế. Nghiên cứu sẽ sử dụng các phương pháp phân tích chuỗi thời gian hiện đại để ước lượng các mô hình dự báo chu kỳ kinh tế, dựa trên các đặc điểm của lợi suất trái phiếu chính phủ. Ngoài ra, nghiên cứu này trong tương lai cũng sẽ tìm hiểu cách các yếu tố vĩ mô, như tỷ lệ lạm phát, tăng trưởng GDP, và tỷ lệ thất nghiệp, ảnh hưởng đến lợi suất trái phiếu chính phủ và tác động của chúng đến chu kỳ kinh tế.

Mối liên hệ giữa lợi suất trái phiếu và chu kỳ kinh tế có thể giúp các nhà đầu tư, chính phủ, và các tổ chức tài chính đưa ra các quyết định chính sách đúng đắn. Việc dự báo chính xác chu kỳ

kinh tế có thể giúp giảm thiểu các rủi ro tài chính và cải thiện khả năng dự đoán các sự kiện quan trọng trong tương lai. Các phương pháp phân tích chuỗi thời gian, đặc biệt là các mô hình ARIMA, GARCH, hay các phương pháp học máy hiện đại, sẽ được áp dụng để phân tích dữ liệu và đưa ra các dự báo chính xác.

1.3 Phạm vi và phương pháp nghiên cứu

Nghiên cứu này tập trung vào dữ liệu lợi suất trái phiếu chính phủ Hoa Kỳ từ năm 1960 đến 2024, với các cột dữ liệu gồm "Ngày", "Giá mở cửa", "Giá cao", "Giá thấp", "Giá đóng cửa", "Giá đóng cửa điều chỉnh", và "Khối lượng giao dịch". Dữ liệu này sẽ được xử lý và phân tích để rút ra các đặc trưng của chuỗi thời gian, bao gồm xu hướng dài hạn, sự biến động theo chu kỳ, và mối liên hệ với các yếu tố vĩ mô. Các phương pháp phân tích chuỗi thời gian sẽ được áp dụng, bao gồm phân tích đặc trưng chuỗi, xác định mùa vụ, kiểm tra tính dừng của chuỗi thời gian, và xây dựng các mô hình dự báo.

Một trong những điểm quan trọng của nghiên cứu này là việc sử dụng các mô hình dự báo để ước tính chu kỳ kinh tế dựa trên các biến số từ lợi suất trái phiếu. Các mô hình ARIMA (AutoRegressive Integrated Moving Average) và các biến thể của nó, như SARIMA (Seasonal ARIMA) và ARCH/GARCH, sẽ được nghiên cứu và áp dụng để dự báo lợi suất và chu kỳ kinh tế. Ngoài ra, các phương pháp học máy như Random Forest, XGBoost, và các mạng nơ-ron nhân tạo cũng sẽ được xem xét như những công cụ hỗ trợ để cải thiện độ chính xác của các dự báo.

1.4 Tổng quan về chu kỳ kinh tế và lợi suất trái phiếu chính phủ

Chu kỳ kinh tế là quá trình biến động của nền kinh tế qua các giai đoạn mở rộng và suy thoái. Mỗi chu kỳ kinh tế thường bao gồm các giai đoạn phục hồi, tăng trưởng, đạt đỉnh và suy thoái. Các chỉ số kinh tế vĩ mô, bao gồm GDP, tỷ lệ thất nghiệp, tỷ lệ lạm phát và lợi suất trái phiếu, thường được sử dụng để đánh giá các giai đoạn khác nhau trong chu kỳ kinh tế. Lợi suất trái phiếu chính phủ là một trong những chỉ báo quan trọng về sức khỏe tài chính của quốc gia, vì nó phản ánh kỳ vọng của thị trường về chính sách tiền tệ, lạm phát và tăng trưởng kinh tế.

Lợi suất trái phiếu chính phủ Mỹ, với đặc tính an toàn và thanh khoản cao, thường được sử dụng

làm thước đo cho chi phí vay mượn trong nền kinh tế. Mỗi quan hệ giữa lợi suất trái phiếu và chu kỳ kinh tế có thể giúp các nhà phân tích hiểu rõ hơn về những yếu tố ảnh hưởng đến nền kinh tế và dự báo được các giai đoạn chuyển tiếp của chu kỳ kinh tế. Trong các chu kỳ kinh tế, lợi suất trái phiếu có xu hướng thay đổi, với sự gia tăng trong các giai đoạn tăng trưởng mạnh mẽ và giảm xuống trong các giai đoạn suy thoái.

2 Phương pháp - Tổng quan về quy trình nghiên cứu

Quy trình nghiên cứu

Quy trình nghiên cứu này được thiết kế để phân tích chuỗi thời gian của lợi suất trái phiếu chính phủ Hoa Kỳ và dự báo chu kỳ kinh tế. Quy trình phân tích gồm các bước chính sau đây: (i) Thu thập và chuẩn bị dữ liệu, (ii) Phân tích sơ bộ dữ liệu, (iii) Kiểm tra đặc tính của chuỗi thời gian, (iv) Xây dựng và đánh giá mô hình dự báo, và (v) Dự báo và kiểm tra mô hình. Mỗi bước sẽ được mô tả chi tiết dưới đây.

2.1 Bước 1: Thu thập và chuẩn bị dữ liệu

Bước đầu tiên trong quy trình là thu thập dữ liệu về lợi suất trái phiếu chính phủ Hoa Kỳ từ năm 1960 đến 2024. Dữ liệu này được thu thập từ các nguồn dữ liệu tài chính đáng tin cậy, bao gồm các cột thông tin như "Ngày", "Giá mở cửa", "Giá cao", "Giá thấp", "Giá đóng cửa", "Giá đóng cửa điều chỉnh", và "Khối lượng giao dịch". Trong quá trình thu thập, việc kiểm tra độ chính xác và tính đầy đủ của dữ liệu là rất quan trọng. Dữ liệu có thể thiếu hoặc có sự sai lệch do lỗi thu thập, vì vậy việc làm sạch dữ liệu là bước không thể thiếu. Các giá trị thiếu sẽ được xử lý bằng cách sử dụng các phương pháp như nội suy, điền giá trị trung bình, hoặc bỏ qua các điểm dữ liệu không cần thiết.

Sau khi thu thập và làm sạch dữ liệu, bước tiếp theo là chuẩn bị dữ liệu cho việc phân tích. Việc chuyển đổi các cột dữ liệu thành các dạng dễ phân tích và xử lý sẽ được thực hiện. Điều này bao gồm việc chuyển đổi các dữ liệu dạng ngày tháng thành kiểu dữ liệu chuẩn, loại bỏ các điểm ngoại lai (outliers) có thể làm sai lệch kết quả phân tích, và chuẩn hóa các dữ liệu theo yêu cầu của mô

hình.

2.2 Bước 2: Phân tích sơ bộ dữ liệu

Sau khi dữ liệu được làm sạch và chuẩn bị, bước tiếp theo là phân tích sơ bộ dữ liệu. Mục đích của bước này là để hiểu rõ hơn về đặc điểm của chuỗi thời gian, bao gồm xu hướng, chu kỳ và các yếu tố mùa vụ. Phân tích sơ bộ bao gồm các bước:

- **Khám phá dữ liệu:** Bước này liên quan đến việc vẽ đồ thị chuỗi thời gian để quan sát sự biến động của lợi suất trái phiếu theo thời gian. Các biểu đồ như đồ thị đường thẳng, biểu đồ phân phối và biểu đồ boxplot sẽ giúp phát hiện các xu hướng tổng quát, các đỉnh và đáy bất thường, và phân bố của lợi suất.
- **Kiểm tra tính dừng của chuỗi thời gian:** Một đặc điểm quan trọng của chuỗi thời gian là tính dừng (stationarity). Một chuỗi thời gian dừng có các đặc tính thống kê không thay đổi theo thời gian, như trung bình và phương sai. Việc kiểm tra tính dừng được thực hiện thông qua các kiểm định như Augmented Dickey-Fuller (ADF) test. Nếu chuỗi không dừng, cần phải biến đổi chuỗi để đạt được tính dừng, ví dụ bằng cách lấy sai phân (differencing).
- **Phân tích các yếu tố mùa vụ và chu kỳ:** Để hiểu rõ hơn về các yếu tố ảnh hưởng đến lợi suất, các phân tích chi tiết về mùa vụ và chu kỳ cũng sẽ được thực hiện. Việc tách các yếu tố mùa vụ sẽ giúp làm rõ các ảnh hưởng có tính chu kỳ của các yếu tố như chính sách tài chính, thay đổi trong tăng trưởng GDP, tỷ lệ thất nghiệp, v.v.

2.3 Bước 3: Kiểm tra đặc tính của chuỗi thời gian

Trong bước này, các đặc tính thống kê của chuỗi thời gian sẽ được kiểm tra kỹ lưỡng hơn, bao gồm độ tự tương quan (autocorrelation), sự phân bố của chuỗi, và các cấu trúc trong dữ liệu. Một trong những phương pháp quan trọng trong phân tích chuỗi thời gian là việc kiểm tra tính tự tương quan của chuỗi. Các kiểm định như ACF (AutoCorrelation Function) và PACF (Partial AutoCorrelation Function) giúp xác định mức độ tự tương quan và có thể chỉ ra các mô hình AR (AutoRegressive) hoặc MA (Moving Average) phù hợp.

Ngoài ra, trong bước này cũng sẽ thực hiện kiểm tra các mô hình ARIMA (AutoRegressive Integrated Moving Average) để xác định các tham số tối ưu cho mô hình. Các tham số này bao gồm p (số bậc tự hồi quy), d (số lần sai phân cần thiết để chuỗi trở thành dừng), và q (số bậc của phần trung bình động). Việc lựa chọn các tham số này dựa trên các kiểm định ACF và PACF, cũng như các tiêu chí thông qua các phép đo như AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion).

2.4 Bước 4: Xây dựng và đánh giá mô hình dự báo

Sau khi các bước chuẩn bị dữ liệu và kiểm tra đặc tính chuỗi thời gian đã hoàn tất, bước tiếp theo là xây dựng các mô hình dự báo chu kỳ kinh tế. Một số mô hình phổ biến trong phân tích chuỗi thời gian bao gồm:

- **Mô hình ARIMA:** Mô hình này được sử dụng rộng rãi trong phân tích chuỗi thời gian, đặc biệt khi dữ liệu có tính dừng. ARIMA là sự kết hợp giữa ba thành phần: tự hồi quy (AR), sai phân (I), và trung bình động (MA). Mô hình này được áp dụng để dự báo lợi suất trái phiếu trong các chu kỳ kinh tế.
- **Mô hình GARCH (Generalized Autoregressive Conditional Heteroskedasticity):** Đây là mô hình được sử dụng để dự báo sự biến động (volatility) trong chuỗi thời gian, đặc biệt là trong các chu kỳ kinh tế khi có sự thay đổi lớn về tỷ lệ lạm phát hay các yếu tố vĩ mô khác.
- **Mô hình học máy và học sâu:** Mô hình LSTM được áp dụng để dự báo chu kỳ kinh tế. Những mô hình này có thể xử lý các mối quan hệ phi tuyến tính trong dữ liệu và cung cấp các dự báo chính xác hơn trong những trường hợp phức tạp.

Sau khi xây dựng các mô hình, việc đánh giá mô hình sẽ được thực hiện thông qua các chỉ tiêu như MSE (Mean Squared Error), RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), và R^2 . Các mô hình tốt nhất sẽ được chọn để thực hiện dự báo chu kỳ kinh tế.

2.5 Bước 5: Dự báo và kiểm tra mô hình

Cuối cùng, sau khi các mô hình đã được xây dựng và lựa chọn, bước cuối cùng là thực hiện dự báo và kiểm tra hiệu suất của mô hình. Dự báo sẽ được thực hiện cho các khoảng thời gian trong tương lai và so sánh với giá trị thực tế khi có sẵn dữ liệu mới. Các mô hình sẽ được kiểm tra lại định kỳ để đảm bảo độ chính xác trong các dự báo dài hạn.

Trong khuôn khổ nghiên cứu này sẽ đơn giản đưa ra dự báo cho 365 ngày sắp tới.

3 Cơ sở lý thuyết

3.1 Thu thập và chuẩn bị dữ liệu

Bước đầu tiên trong quy trình phân tích chuỗi thời gian là thu thập và chuẩn bị dữ liệu. Dữ liệu trong nghiên cứu này bao gồm các thông tin về lợi suất trái phiếu chính phủ Hoa Kỳ, được thu thập từ năm 1960 đến 2024. Dữ liệu này được cung cấp từ các nguồn tài chính đáng tin cậy như Bloomberg, Yahoo Finance, FRED (Federal Reserve Economic Data) hoặc các cơ sở dữ liệu tài chính công cộng khác. Các nguồn này cung cấp dữ liệu với độ chính xác cao và có tính sẵn có, đặc biệt là đối với các thông tin về các công cụ tài chính vĩ mô như lợi suất trái phiếu.

3.1.1 Thu thập dữ liệu

Dữ liệu thu thập bao gồm các cột chính sau:

- **Ngày:** Thời gian ghi nhận dữ liệu, thường được cung cấp dưới dạng ngày tháng năm.
- **Giá mở cửa (Open):** Giá của trái phiếu tại thời điểm mở cửa giao dịch trong một ngày.
- **Giá cao (High):** Giá cao nhất đạt được trong ngày.
- **Giá thấp (Low):** Giá thấp nhất đạt được trong ngày.
- **Giá đóng cửa (Close):** Giá cuối cùng giao dịch trong ngày.

- **Giá đóng cửa điều chỉnh (Adj Close):** Giá đóng cửa đã được điều chỉnh để phản ánh các thay đổi như cổ tức, chia tách cổ phiếu, v.v.
- **Khối lượng giao dịch (Volume):** Số lượng giao dịch được thực hiện trong ngày.

Các dữ liệu này là các yếu tố quan trọng để phân tích lợi suất trái phiếu chính phủ và xác định các xu hướng kinh tế. Tuy nhiên, trong quá trình thu thập, dữ liệu có thể bị thiếu hoặc sai lệch do lỗi trong quá trình nhập liệu, giao dịch thị trường không bình thường, hoặc các sự kiện ngoại lệ.

3.1.2 Kiểm tra độ chính xác và tính đầy đủ của dữ liệu

Một trong những vấn đề quan trọng khi làm việc với dữ liệu tài chính là đảm bảo tính đầy đủ và chính xác của nó. Dữ liệu có thể bị thiếu hoặc sai lệch trong các trường hợp như các ngày nghỉ lễ, sự kiện giao dịch bất thường, hoặc các vấn đề trong quá trình thu thập từ các nguồn dữ liệu. Do đó, việc kiểm tra dữ liệu là một bước quan trọng để loại bỏ các sự bất thường.

Kiểm tra độ chính xác có thể được thực hiện bằng cách so sánh dữ liệu thu thập được với các dữ liệu từ các nguồn khác (nếu có sẵn), kiểm tra các mốc thời gian hoặc sự kiện quan trọng trong chuỗi thời gian như các cuộc khủng hoảng tài chính. Điều này giúp phát hiện các lỗi nhập liệu, ví dụ như giá trị sai hoặc thiếu.

Ngoài ra, tính đầy đủ của dữ liệu cần được đảm bảo, đặc biệt là trong các chuỗi thời gian dài hạn. Việc thiếu dữ liệu ở một số thời điểm có thể ảnh hưởng đến kết quả phân tích, đặc biệt trong các mô hình dự báo. Một phương pháp phổ biến để kiểm tra tính đầy đủ là vẽ biểu đồ chuỗi thời gian và quan sát các khoảng trống, sau đó xác định liệu có thể điền các giá trị thiếu một cách hợp lý hay không.

3.1.3 Xử lý dữ liệu thiếu

Khi phát hiện dữ liệu thiếu, có một số phương pháp để xử lý:

- **Nội suy (Interpolation):** Đây là phương pháp phổ biến để điền các giá trị thiếu trong chuỗi thời gian. Nội suy tuyến tính là phương pháp thường được sử dụng, trong đó giá trị thiếu được tính bằng cách sử dụng các giá trị xung quanh (trước và sau điểm thiếu) để ước tính giá

trị thiếu. Nội suy có thể áp dụng cho các giá trị thiếu trong một khoảng thời gian ngắn hoặc trong các mô hình phân tích có độ chính xác cao.

- **Điền giá trị trung bình hoặc trung vị:** Trong trường hợp chuỗi thời gian có một số điểm dữ liệu thiếu không thể nội suy, một phương pháp đơn giản là điền giá trị trung bình hoặc trung vị của chuỗi vào những vị trí thiếu. Phương pháp này đơn giản nhưng có thể làm giảm độ chính xác của mô hình nếu giá trị thiếu không phân bố ngẫu nhiên.
- **Loại bỏ điểm dữ liệu không cần thiết:** Đôi khi, nếu các điểm dữ liệu thiếu quá nhiều hoặc không thể điền một cách hợp lý, một lựa chọn khác là loại bỏ các dòng dữ liệu đó khỏi tập dữ liệu. Tuy nhiên, việc này cần được thực hiện cẩn thận, đặc biệt khi các điểm dữ liệu bị thiếu xuất hiện ở những thời điểm quan trọng.

Phương pháp xử lý dữ liệu thiếu phải được chọn dựa trên bản chất của dữ liệu và yêu cầu của mô hình phân tích.

3.1.4 Làm sạch dữ liệu và loại bỏ ngoại lai

Trong khi chuẩn bị dữ liệu, bước tiếp theo là loại bỏ các điểm ngoại lai (outliers), những giá trị có thể làm sai lệch kết quả phân tích. Các ngoại lai có thể do nhiều nguyên nhân khác nhau, chẳng hạn như lỗi ghi nhận giá trị không hợp lý, hoặc các sự kiện đặc biệt ngoài tầm kiểm soát. Để phát hiện ngoại lai, có thể sử dụng các phương pháp thống kê như phương pháp boxplot, phân tích tần suất, hoặc kiểm tra theo các quy tắc như Z-score.

Nếu các điểm ngoại lai chỉ xuất hiện trong một khoảng thời gian ngắn và không ảnh hưởng đáng kể đến phân tích tổng thể, chúng có thể được loại bỏ hoặc thay thế bằng các giá trị hợp lý hơn. Tuy nhiên, nếu các ngoại lai mang tính chất quan trọng (ví dụ: các cú sốc kinh tế, khủng hoảng tài chính), cần phải lưu giữ và phân tích chúng một cách riêng biệt.

3.1.5 Chuẩn hóa và chuyển đổi dữ liệu

Để đảm bảo tính khả thi trong việc phân tích và xây dựng mô hình, dữ liệu cần phải được chuẩn hóa và chuyển đổi sao cho phù hợp với yêu cầu của các mô hình phân tích chuỗi thời gian. Một

trong những bước quan trọng là chuyển đổi các dữ liệu ngày tháng thành dạng chuẩn (ví dụ: năm-tháng-ngày), giúp mô hình dễ dàng xử lý các yếu tố theo thời gian.

Các cột dữ liệu như "Giá mở cửa", "Giá đóng cửa", và "Khối lượng giao dịch" có thể cần phải được chuẩn hóa, tức là đưa về cùng một phạm vi hoặc phân phối. Phương pháp chuẩn hóa thường sử dụng là Min-Max Scaling hoặc Z-score Standardization, giúp các giá trị trong các cột này có thể so sánh được và không bị ảnh hưởng bởi sự chênh lệch đơn vị.

3.1.6 Chuẩn bị dữ liệu cho phân tích

Sau khi đã thu thập và làm sạch dữ liệu, bước tiếp theo là chuẩn bị dữ liệu cho quá trình phân tích. Việc này bao gồm việc chia nhỏ dữ liệu thành các tập huấn luyện, kiểm tra và validation. Việc chia dữ liệu này giúp xây dựng các mô hình chính xác hơn và đánh giá hiệu quả mô hình một cách khách quan.

Phân tích chuỗi thời gian yêu cầu dữ liệu phải có tính liên tục và không gián đoạn, vì vậy việc sắp xếp lại dữ liệu theo thứ tự thời gian là rất quan trọng. Cùng với đó, việc phân chia dữ liệu thành các phần có thể kiểm tra lại giúp phát hiện được mô hình tốt nhất và giảm thiểu hiện tượng overfitting.

3.2 Phân tích sơ bộ dữ liệu

Phân tích sơ bộ dữ liệu là bước quan trọng để hiểu rõ hơn về đặc điểm của chuỗi thời gian, bao gồm các xu hướng dài hạn, chu kỳ, yếu tố mùa vụ và các biến động ngắn hạn. Việc phân tích này giúp xác định các đặc điểm chính của chuỗi dữ liệu, từ đó đưa ra các quyết định quan trọng trong việc xây dựng mô hình dự báo. Bước phân tích sơ bộ bao gồm ba yếu tố chính: khám phá dữ liệu, kiểm tra tính dừng của chuỗi thời gian và phân tích các yếu tố mùa vụ và chu kỳ. Dưới đây là phần chi tiết về từng bước trong quy trình phân tích sơ bộ dữ liệu.

3.2.1 Khám phá dữ liệu

Khám phá dữ liệu (exploratory data analysis - EDA) là bước đầu tiên trong quá trình phân tích sơ bộ, nhằm giúp nhà nghiên cứu hiểu rõ hơn về cấu trúc của dữ liệu và phát hiện các đặc điểm quan trọng như xu hướng, tính chất phân phối và các điểm ngoại lai.

Vẽ đồ thị chuỗi thời gian Để quan sát sự biến động của lợi suất trái phiếu theo thời gian, việc vẽ đồ thị chuỗi thời gian là bước quan trọng nhất. Một đồ thị chuỗi thời gian thường bao gồm các giá trị của lợi suất trái phiếu theo trục tung (y-axis) và thời gian (theo ngày, tháng, hoặc năm) trên trục hoành (x-axis). Qua đó, ta có thể quan sát sự biến động và các xu hướng trong chuỗi dữ liệu. Ví dụ, việc vẽ đồ thị giúp phát hiện các xu hướng tăng hoặc giảm dài hạn trong lợi suất, đồng thời nhận diện các đỉnh và đáy bất thường, ví dụ như trong các giai đoạn khủng hoảng tài chính. Các đỉnh và đáy này có thể chỉ ra những thay đổi đáng kể trong môi trường kinh tế hoặc các chính sách tài chính.

Biểu đồ phân phối Biểu đồ phân phối, như histograms hoặc kernel density plots, giúp phân tích phân phối của lợi suất trái phiếu. Biểu đồ này cho phép chúng ta quan sát xem lợi suất có phân phối đều, chuẩn, hay có sự lệch về phía một giá trị nhất định. Phân phối lợi suất có thể cung cấp thông tin về sự biến động của thị trường tài chính và các yếu tố ảnh hưởng đến giá trị của trái phiếu.

Biểu đồ boxplot Boxplot là một công cụ hữu ích để nhận diện các điểm ngoại lai (outliers) trong dữ liệu. Boxplot giúp chia dữ liệu thành các phần tử theo quartile (phần vị) và hiển thị giá trị tối đa, tối thiểu, giá trị trung vị (median) và các phần tử ngoại lai. Bằng cách sử dụng boxplot, ta có thể phát hiện được các giá trị bất thường hoặc sai lệch, những giá trị này có thể gây ảnh hưởng lớn đến mô hình phân tích và dự báo.

3.2.2 Kiểm tra tính dừng của chuỗi thời gian

Một đặc điểm quan trọng của chuỗi thời gian là tính dừng (stationarity). Chuỗi thời gian dừng có các đặc tính thống kê không thay đổi theo thời gian, như trung bình, phương sai, và hiệp phương sai. Điều này có nghĩa là các đặc tính của chuỗi dữ liệu không phụ thuộc vào thời gian, điều này rất quan trọng khi xây dựng các mô hình dự báo như ARIMA, vì các mô hình này giả định chuỗi dữ liệu là dừng.

Kiểm tra tính dừng với Augmented Dickey-Fuller (ADF) Test Kiểm tra tính dừng của chuỗi thời gian là bước quan trọng trong phân tích chuỗi thời gian, vì hầu hết các mô hình dự báo, chẳng hạn như ARIMA, yêu cầu chuỗi dữ liệu phải là chuỗi dừng (stationary). Chuỗi thời gian được coi là dừng nếu các đặc trưng thống kê của nó (như trung bình, phương sai, và hiệp phương sai) không thay đổi theo thời gian. Ngược lại, chuỗi không dừng có thể có xu hướng (trend) hoặc sự thay đổi mùa vụ (seasonality), điều này cần được xử lý trước khi tiếp tục phân tích. Một trong những phương pháp phổ biến nhất để kiểm tra tính dừng là kiểm tra Augmented Dickey-Fuller (ADF) test.

Giả thuyết của ADF Test Kiểm định ADF là một bài kiểm tra giả thuyết thống kê, được sử dụng để kiểm tra sự tồn tại của đơn gốc (unit root) trong chuỗi thời gian, nghĩa là kiểm tra xem chuỗi dữ liệu có phải là chuỗi dừng hay không. Kiểm định này được mở rộng từ kiểm định Dickey-Fuller (DF) ban đầu, nhưng với một số cải tiến, bao gồm thêm các biến phụ (lag) để giải quyết vấn đề tự tương quan trong chuỗi dữ liệu.

- **Giả thuyết không dừng (Null Hypothesis, H_0):** Chuỗi có đơn gốc, tức là chuỗi không dừng.

- **Giả thuyết đối (Alternative Hypothesis, H_1):** Chuỗi không có đơn gốc, tức là chuỗi dừng.

Do đó, nếu p-value của kiểm định nhỏ hơn mức ý nghĩa α (thường là 0.05), ta bác bỏ giả thuyết không dừng H_0 và kết luận rằng chuỗi thời gian là dừng. Ngược lại, nếu p-value lớn hơn α , ta không thể bác bỏ giả thuyết không dừng và chuỗi thời gian được coi là không dừng.

Mô hình ADF Kiểm định ADF sử dụng một mô hình hồi quy để kiểm tra đơn gốc trong chuỗi thời gian. Mô hình cơ bản của kiểm định này có thể được viết dưới dạng sau:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \dots + \delta_p \Delta y_{t-p} + \varepsilon_t$$

Trong đó: - y_t là giá trị chuỗi thời gian tại thời điểm t ,

- $\Delta y_t = y_t - y_{t-1}$ là sự thay đổi của chuỗi thời gian,

- α là hệ số chặn (intercept),

- βt là thành phần xu hướng (trend component),

- γ là hệ số kiểm tra đơn gốc,

- ε_t là phần dư ngẫu nhiên (error term),

- $\delta_1, \dots, \delta_p$ là các hệ số của các sai phân bậc cao (lag terms).

Nếu $\gamma = 0$, tức là nếu hệ số của y_{t-1} trong mô hình hồi quy không đáng kể, thì chuỗi có đơn gốc và không dừng. Nếu $\gamma < 0$, chuỗi không có đơn gốc và là chuỗi dừng.

Lựa chọn độ trễ (Lag length) trong ADF Việc lựa chọn số lượng độ trễ p trong mô hình ADF là rất quan trọng, vì độ trễ quá ngắn có thể không bắt được hết tự tương quan trong chuỗi, còn độ trễ quá dài có thể làm cho mô hình trở nên quá phức tạp và dễ bị overfitting. Phương pháp phổ biến để chọn độ trễ là sử dụng các tiêu chí thông tin như Akaike Information Criterion (AIC) hoặc Bayesian Information Criterion (BIC).

$AIC = -2\ln L + 2k$ với k là số lượng tham số trong mô hình và L là giá trị hàm likelihood.

$BIC = -2\ln L + k\ln n$ với n là số lượng quan sát.

Kết quả kiểm định ADF phụ thuộc vào mức độ trễ được chọn, vì vậy việc lựa chọn độ trễ tối ưu giúp cải thiện độ chính xác của kiểm định.

Kết quả của ADF Test Kết quả của kiểm định ADF sẽ bao gồm giá trị thống kê kiểm định và p-value. Nếu p-value nhỏ hơn mức ý nghĩa đã chọn (thường là 0.05), ta có thể bác bỏ giả thuyết không dừng H_0 và kết luận rằng chuỗi thời gian là dừng. Ngược lại, nếu p-value lớn hơn mức ý nghĩa, ta không thể bác bỏ giả thuyết không dừng và chuỗi không dừng.

$p\text{-value} \leq 0.05 \Rightarrow \text{Bác bỏ } H_0 \Rightarrow \text{Chuỗi là dừng.}$

$p\text{-value} > 0.05 \Rightarrow \text{Không bác bỏ } H_0 \Rightarrow \text{Chuỗi không dừng.}$

Ứng dụng của ADF Test trong phân tích chuỗi thời gian Kiểm định ADF là công cụ quan trọng trong việc xác định tính chất dừng của chuỗi thời gian. Việc kiểm tra tính dừng rất quan trọng vì hầu hết các mô hình dự báo, như ARIMA, yêu cầu chuỗi thời gian phải là chuỗi dừng. Nếu chuỗi không dừng, các bước như sai phân cần được thực hiện để làm cho chuỗi trở thành dừng. Một chuỗi dừng sẽ có các tính chất thống kê ổn định và có thể dự báo tốt hơn trong các mô hình dự báo.

Ngoài ADF, còn có các phương pháp kiểm tra tính dừng khác như kiểm định Phillips-Perron (PP) hoặc kiểm định Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Tuy nhiên, ADF vẫn là phương pháp phổ biến và mạnh mẽ trong phân tích chuỗi thời gian.

Lấy sai phân (Differencing) Khi chuỗi thời gian không dừng, một trong những phương pháp phổ biến để làm cho chuỗi trở nên dừng là thực hiện quá trình *lấy sai phân* (differencing). Sai phân là một kỹ thuật đơn giản giúp loại bỏ xu hướng dài hạn hoặc mùa vụ trong chuỗi dữ liệu, từ đó giảm bớt các yếu tố không ổn định như xu hướng tuyến tính hoặc tăng trưởng theo thời gian. Việc thực hiện sai phân làm giảm sự phụ thuộc của chuỗi vào thời gian, giúp chuỗi trở nên ổn định hơn và có thể được phân tích hoặc dự báo bằng các mô hình như ARIMA.

Sai phân bậc 1 (First Differencing) Sai phân bậc 1, hay còn gọi là *first differencing*, được thực hiện bằng cách tính hiệu số giữa các giá trị liên tiếp trong chuỗi thời gian. Cụ thể, với chuỗi dữ liệu y_t , sai phân bậc 1 được tính như sau:

$$\Delta y_t = y_t - y_{t-1}$$

Trong đó:

- y_t là giá trị tại thời điểm t ,
- y_{t-1} là giá trị tại thời điểm $t - 1$,
- Δy_t là sự thay đổi (sai phân) giữa hai giá trị liên tiếp.

Mục tiêu của việc lấy sai phân là làm giảm sự hiện diện của xu hướng dài hạn. Khi chuỗi có xu hướng tăng hoặc giảm đều đặn theo thời gian, sai phân bậc 1 giúp loại bỏ phần này, chỉ còn lại các yếu tố ngẫu nhiên hoặc các yếu tố ngắn hạn khác. Điều này làm cho chuỗi trở nên dừng hơn, vì giá

trị của chuỗi không còn phụ thuộc vào thời gian mà chỉ vào các biến động ngắn hạn.

Ví dụ, giả sử ta có một chuỗi thời gian y_t với các giá trị tăng dần theo thời gian. Sau khi thực hiện sai phân, chuỗi Δy_t sẽ chỉ phản ánh sự thay đổi trong giá trị, mà không còn sự tăng trưởng đều đặn của chuỗi ban đầu.

Sai phân bậc 2 (Second Differencing) Nếu chuỗi thời gian vẫn không dừng sau khi lấy sai phân bậc 1, có thể cần phải thực hiện sai phân bậc 2, tức là lấy sai phân của chuỗi đã được sai phân trước đó. Quá trình này loại bỏ các xu hướng không chỉ ở mức tuyến tính mà còn ở các xu hướng bậc cao hơn, chẳng hạn như chu kỳ tăng giảm không đều. Sai phân bậc 2 được tính như sau:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

Trong đó:

- $\Delta^2 y_t$ là sai phân bậc 2 tại thời điểm t ,

- y_t, y_{t-1} , và y_{t-2} là các giá trị trong chuỗi thời gian tại các thời điểm tương ứng.

Sai phân bậc 2 có tác dụng mạnh mẽ hơn trong việc loại bỏ các xu hướng không chỉ tuyến tính mà còn các xu hướng phức tạp hơn, như chu kỳ hoặc các biến động không đều theo thời gian. Sau khi thực hiện sai phân bậc 2, chuỗi thời gian có thể trở nên ổn định hơn và có thể áp dụng các mô hình phân tích chuỗi thời gian.

Sai phân bậc d (Differencing of Order d) Quá trình sai phân có thể được thực hiện nhiều lần để đạt được sự dừng ổn định của chuỗi thời gian. Nếu chuỗi không dừng sau một lần sai phân, ta có thể tiếp tục thực hiện sai phân bậc cao hơn cho đến khi chuỗi trở thành dừng. Sai phân bậc d được tính bằng cách lấy sai phân d lần liên tiếp. Mỗi lần sai phân làm giảm độ phức tạp của xu hướng trong chuỗi, và giá trị sai phân có thể được ký hiệu là $\Delta^d y_t$, với d là bậc sai phân:

$$\Delta^d y_t = \Delta(\Delta^{d-1} y_t)$$

Ví dụ, đối với chuỗi không dừng do xu hướng bậc 2 (như chu kỳ tăng giảm không đều), có thể thực hiện sai phân bậc 2 hoặc bậc 3 cho đến khi chuỗi trở thành dừng.

Kiểm tra tính dừng sau khi sai phân Sau khi thực hiện quá trình sai phân, ta cần kiểm tra lại tính dừng của chuỗi bằng các kiểm định như Augmented Dickey-Fuller (ADF) để xác nhận rằng chuỗi đã trở thành dừng. Nếu chuỗi đã dừng, ta có thể tiến hành các phân tích tiếp theo. Ngược lại, nếu chuỗi vẫn chưa dừng, có thể cần thực hiện sai phân thêm hoặc áp dụng các phương pháp khác như loại bỏ các yếu tố mùa vụ (seasonal adjustment).

Lý thuyết đằng sau sai phân Về lý thuyết, sai phân giúp loại bỏ xu hướng trong chuỗi thời gian bằng cách tính toán sự thay đổi giữa các giá trị liên tiếp. Khi có xu hướng, giá trị tại thời điểm t có thể có sự phụ thuộc mạnh vào giá trị tại thời điểm $t - 1$, tạo ra sự thay đổi chậm theo thời gian. Sai phân giúp làm cho chuỗi dữ liệu không còn có sự phụ thuộc mạnh mẽ vào thời gian, từ đó đạt được tính ổn định về mặt thống kê.

Về mặt lý thuyết toán học, chuỗi thời gian không dừng có thể được mô hình hóa dưới dạng một quá trình có đơn gốc (unit root process). Quá trình này có tính chất là các giá trị của nó có xu hướng thay đổi liên tục mà không có sự ổn định về mặt thống kê. Việc thực hiện sai phân giúp loại bỏ đơn gốc, làm cho chuỗi trở thành một quá trình dừng, trong đó các đặc tính thống kê của chuỗi (như trung bình, phương sai) không thay đổi theo thời gian.

Ứng dụng sai phân trong mô hình ARIMA Trong các mô hình dự báo chuỗi thời gian như ARIMA (AutoRegressive Integrated Moving Average), quá trình sai phân được sử dụng để làm cho chuỗi trở thành dừng. Mô hình ARIMA bao gồm ba thành phần chính: tự hồi quy (AR), sai phân (I), và trung bình động (MA). Trong đó, thành phần sai phân (I) giúp chuyển chuỗi không dừng thành chuỗi dừng. Do đó, việc chọn bậc sai phân d là rất quan trọng trong việc áp dụng mô hình ARIMA. Thông qua việc thử nghiệm với các giá trị khác nhau của d , ta có thể tìm được giá trị tối ưu giúp cải thiện độ chính xác của mô hình.

3.2.3 Phân tích các yếu tố mùa vụ và chu kỳ

Trong phân tích chuỗi thời gian, ngoài việc nhận diện các xu hướng dài hạn và kiểm tra tính dừng của chuỗi, việc phân tích các yếu tố mùa vụ và chu kỳ cũng đóng vai trò quan trọng. Các yếu tố này có thể bao gồm các ảnh hưởng từ các sự kiện định kỳ, sự thay đổi trong chính sách tài chính,

biến động trong các yếu tố kinh tế vĩ mô như tăng trưởng GDP, tỷ lệ thất nghiệp, hay sự thay đổi trong các yếu tố xã hội. Việc phân tách các yếu tố mùa vụ và chu kỳ không chỉ giúp làm rõ các yếu tố ảnh hưởng đến lợi suất trái phiếu, mà còn cung cấp các thông tin quan trọng cho việc xây dựng mô hình dự báo.

Phân tích các yếu tố mùa vụ (Seasonality Analysis) Yếu tố mùa vụ trong chuỗi thời gian đề cập đến các biến động có tính chu kỳ, diễn ra với tần suất xác định trong một năm hoặc trong một khoảng thời gian ngắn hơn. Chúng thường được gây ra bởi các yếu tố như thời tiết, chính sách, và các sự kiện định kỳ khác. Đặc điểm của yếu tố mùa vụ là nó có xu hướng lặp lại một cách đều đặn sau các khoảng thời gian nhất định, ví dụ như hàng tháng, hàng quý hoặc hàng năm.

Phương pháp Decomposition (Phân tích tách rời) Một phương pháp phổ biến để phân tích các yếu tố mùa vụ trong chuỗi thời gian là sử dụng phương pháp *Decomposition* (phân tích tách rời). Phương pháp này tách chuỗi thời gian thành các thành phần cơ bản: xu hướng (trend), mùa vụ (seasonality), và phần dư (residual), giúp phân tích và nhận diện các yếu tố ảnh hưởng đến dữ liệu.

Công thức tổng quát của một chuỗi thời gian có thể được mô tả như sau:

$$y_t = T_t + S_t + R_t$$

Trong đó:

- y_t là giá trị chuỗi thời gian tại thời điểm t ,
- T_t là thành phần xu hướng (trend) tại thời điểm t ,
- S_t là thành phần mùa vụ (seasonality) tại thời điểm t ,
- R_t là phần dư (residual) hoặc phần ngẫu nhiên còn lại.

Có hai phương pháp tách rời chuỗi thời gian phổ biến:

- **Decomposition additif (Tách cộng):** Đây là phương pháp tách rời chuỗi bằng cách cộng ba thành phần lại với nhau. Phương pháp này giả định rằng các thành phần (xu hướng, mùa vụ,

dư) có thể cộng lại để tạo thành giá trị của chuỗi thời gian.

$$y_t = T_t + S_t + R_t$$

- **Decomposition multiplicatif (Tách nhân):** Phương pháp này giả định rằng các thành phần mùa vụ, xu hướng, và phần dư có mối quan hệ nhân với nhau, và chuỗi thời gian có thể được mô hình hóa dưới dạng:

$$y_t = T_t \times S_t \times R_t$$

Phương pháp decomposition giúp phân tách rõ ràng các yếu tố mùa vụ khỏi các yếu tố xu hướng dài hạn và phần dư ngẫu nhiên. Sau khi tách thành các thành phần, ta có thể phân tích các yếu tố mùa vụ để hiểu rõ hơn về sự thay đổi theo chu kỳ của chuỗi thời gian, cũng như tác động của các yếu tố này lên lợi suất trái phiếu.

Phương pháp Smoothing (Làm mượt) Một phương pháp khác để phân tích các yếu tố mùa vụ là sử dụng các kỹ thuật làm mượt chuỗi thời gian, như phương pháp trung bình trượt (Moving Average) hay Exponential Smoothing. Các phương pháp này giúp loại bỏ các biến động ngẫu nhiên và làm nổi bật các xu hướng hoặc mùa vụ trong chuỗi.

Moving Average (Trung bình trượt): Phương pháp này sử dụng trung bình cộng của một số giá trị liên tiếp trong chuỗi để làm mượt các biến động ngẫu nhiên và phát hiện các xu hướng dài hạn hoặc mùa vụ. Các phương pháp trung bình trượt có thể là trung bình trượt đơn (Simple Moving Average) hoặc trung bình trượt có trọng số (Weighted Moving Average).

$$S_t = \frac{1}{n} \sum_{i=0}^{n-1} y_{t-i}$$

Trong đó, S_t là giá trị trung bình trượt tại thời điểm t , n là số kỳ tính trung bình trượt.

Exponential Smoothing (Làm mượt theo hàm số mũ): Phương pháp này làm mượt chuỗi thời gian bằng cách gán trọng số giảm dần cho các giá trị trong quá khứ. Phương pháp này đặc biệt hữu ích khi chuỗi có sự thay đổi theo mùa vụ hoặc xu hướng dài hạn.

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}$$

Trong đó, S_t là giá trị làm mượt tại thời điểm t , α là hệ số smoothing (thường dao động từ 0 đến 1), và S_{t-1} là giá trị làm mượt tại thời điểm $t - 1$.

Phân tích chu kỳ (Cycle Analysis) Khác với các yếu tố mùa vụ, các yếu tố chu kỳ không có tính chất cố định, mà thay đổi theo thời gian, thường liên quan đến các biến động vĩ mô như thay đổi trong chính sách tài chính, lãi suất, hoặc tình trạng của nền kinh tế. Các yếu tố chu kỳ có thể kéo dài từ vài tháng đến vài năm và thường liên quan đến các yếu tố kinh tế vĩ mô như GDP, tỷ lệ thất nghiệp, hoặc lạm phát.

Một phương pháp phổ biến để phân tích các yếu tố chu kỳ là sử dụng mô hình *Fourier Transformation* hoặc *Wavelet Transformation*. Những phương pháp này cho phép phân tích và nhận diện các chu kỳ không cố định trong chuỗi thời gian.

Fourier Transformation: Phương pháp Fourier giúp tách chuỗi thời gian thành các sóng hài (sinusoidal waves) với các tần số khác nhau, từ đó nhận diện các chu kỳ trong chuỗi dữ liệu. Một chuỗi thời gian có thể được mô hình hóa như một tổng của các sóng hài có tần số và biên độ khác nhau.

$$y_t = \sum_{k=1}^n A_k \sin(2\pi f_k t + \phi_k)$$

Trong đó:

- A_k là biên độ của sóng hài thứ k ,
- f_k là tần số của sóng hài thứ k ,
- ϕ_k là pha của sóng hài thứ k ,
- t là thời gian.

Thông qua Fourier Transformation, ta có thể xác định các chu kỳ trong chuỗi thời gian và phân tách các yếu tố chu kỳ ra khỏi các yếu tố mùa vụ hoặc xu hướng dài hạn.

Ứng dụng phân tích mùa vụ và chu kỳ trong dự báo Khi phân tích và tách biệt các yếu tố mùa vụ và chu kỳ, ta có thể sử dụng các mô hình dự báo như ARIMA (AutoRegressive Integrated Moving Average) hoặc SARIMA (Seasonal ARIMA), trong đó thành phần mùa vụ (Seasonal) được đưa vào để cải thiện độ chính xác của mô hình. Mô hình SARIMA bao gồm các tham số bổ sung để mô hình hóa các yếu tố mùa vụ, từ đó giúp dự báo chính xác hơn đối với chuỗi thời gian có tính mùa vụ rõ ràng.

$$\text{SARIMA}(p, d, q)(P, D, Q)_s$$

Trong đó:

- p, d, q là các tham số cho mô hình ARIMA không mùa vụ,
- P, D, Q là các tham số cho mô hình mùa vụ,
- s là chu kỳ mùa vụ (ví dụ $s = 12$ cho dữ liệu hàng tháng có chu kỳ hàng năm).

3.3 Kiểm tra đặc tính của chuỗi thời gian

Trong bước này, các đặc tính thống kê của chuỗi thời gian sẽ được kiểm tra kỹ lưỡng hơn để hiểu rõ hơn về tính chất của chuỗi. Các đặc tính cần kiểm tra bao gồm độ tự tương quan (autocorrelation), sự phân bố của chuỗi, và các cấu trúc trong dữ liệu. Một trong những phương pháp quan trọng trong phân tích chuỗi thời gian là việc kiểm tra tính tự tương quan của chuỗi. Kiểm tra này sẽ giúp nhận diện các mô hình AR (AutoRegressive) hoặc MA (Moving Average) phù hợp.

Độ tự tương quan (Autocorrelation) Độ tự tương quan (ACF - AutoCorrelation Function) đo lường mức độ tương quan giữa các giá trị trong chuỗi thời gian tại các thời điểm khác nhau. Một chuỗi thời gian có tính tự tương quan mạnh mẽ có thể có các phụ thuộc rõ ràng giữa các giá trị của chuỗi tại các thời điểm khác nhau.

Công thức tính độ tự tương quan tại độ trễ k được xác định bởi:

$$\rho_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Trong đó:

- ρ_k là hệ số tự tương quan tại độ trễ k ,
- y_t là giá trị chuỗi thời gian tại thời điểm t ,
- \bar{y} là giá trị trung bình của chuỗi thời gian,
- T là tổng số thời điểm trong chuỗi.

Hệ số tự tương quan ρ_k sẽ có giá trị từ -1 đến 1. Một giá trị ρ_k gần 1 cho thấy sự tự tương quan mạnh mẽ tại độ trễ k , trong khi đó giá trị gần 0 cho thấy không có sự tự tương quan.

Partial AutoCorrelation Function (PACF) Để xác định chính xác hơn các tham số trong mô hình AR, PACF (Partial AutoCorrelation Function) là một công cụ quan trọng. PACF đo lường sự tự tương quan của một chuỗi thời gian với độ trễ k , nhưng loại bỏ sự ảnh hưởng của các độ trễ trước đó. Điều này giúp xác định rõ ràng mức độ ảnh hưởng của mỗi độ trễ riêng biệt mà không bị nhiễu từ các độ trễ ngắn hơn.

Công thức tính PACF không đơn giản như ACF vì nó phải loại bỏ sự ảnh hưởng của các độ trễ trước đó. Tuy nhiên, PACF có thể được tính bằng cách sử dụng mô hình hồi quy AR để xác định các hệ số tự hồi quy tại mỗi độ trễ. PACF có thể giúp xác định số bậc tự hồi quy p trong mô hình AR.

Kiểm tra tính dừng của chuỗi thời gian Một bước quan trọng trong việc phân tích chuỗi thời gian là kiểm tra tính dừng của chuỗi (stationarity). Chuỗi thời gian là dừng nếu các đặc tính thống kê của chuỗi (như trung bình và phương sai) không thay đổi theo thời gian. Kiểm tra tính dừng có thể được thực hiện bằng cách sử dụng các kiểm định thống kê như Augmented Dickey-Fuller (ADF) test, như đã đề cập trong phần trước.

3.4 Xây dựng và đánh giá mô hình dự báo

Sau khi hoàn tất bước chuẩn bị dữ liệu và kiểm tra đặc tính chuỗi thời gian, bước tiếp theo là xây dựng các mô hình dự báo chu kỳ kinh tế. Các mô hình phổ biến trong phân tích chuỗi thời gian bao gồm ARIMA, GARCH, và mô hình học máy/học sâu, đặc biệt là LSTM (Long Short-Term

Memory). Mỗi mô hình có các đặc điểm riêng và được áp dụng trong những tình huống cụ thể để dự báo chu kỳ kinh tế.

3.4.1 Mô hình ARIMA

Giới thiệu ARIMA (AutoRegressive Integrated Moving Average) là một mô hình rất phổ biến trong phân tích chuỗi thời gian. Mô hình ARIMA kết hợp ba thành phần: tự hồi quy (AR), sai phân (I) và trung bình động (MA). Mô hình này đặc biệt hữu ích khi chuỗi thời gian có xu hướng dừng và có cấu trúc tự tương quan.

Công thức mô hình ARIMA Mô hình ARIMA(p,d,q) có thể được biểu diễn bằng công thức:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + e_t$$

Trong đó:

- y_t là giá trị chuỗi thời gian tại thời điểm t ,
- c là một hằng số,
- ϕ_i là các hệ số của phần tự hồi quy (AR),
- e_t là sai số ngẫu nhiên tại thời điểm t ,
- θ_j là các hệ số của phần trung bình động (MA),
- p là số bậc của phần tự hồi quy (AR),
- q là số bậc của phần trung bình động (MA),
- d là số lần sai phân cần thiết để làm cho chuỗi thời gian trở thành dừng.

Cách lựa chọn tham số p, d, q - **p (AR)**: Được xác định từ PACF (Partial Autocorrelation Function). Nếu PACF cắt đứt sau độ trễ p , thì ta chọn p .

- **q (MA)**: Được xác định từ ACF (AutoCorrelation Function). Nếu ACF cắt đứt sau độ trễ q , ta chọn q .

- **d (I)**: Số lần sai phân cần thiết để chuỗi trở thành dừng, được xác định qua việc kiểm tra tính dừng của chuỗi.

3.4.2 Mô hình GARCH

Giới thiệu GARCH (Generalized Autoregressive Conditional Heteroskedasticity) là mô hình dùng để dự báo sự biến động (volatility) trong chuỗi thời gian. Mô hình này rất hữu ích trong các tình huống khi dữ liệu có sự thay đổi lớn về độ biến động theo thời gian, chẳng hạn như trong các chu kỳ kinh tế, nơi tỷ lệ lạm phát, thất nghiệp, hoặc các yếu tố vĩ mô khác thay đổi mạnh mẽ.

Công thức mô hình GARCH Mô hình GARCH(p,q) có thể được biểu diễn như sau:

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

Trong đó:

- y_t là giá trị chuỗi thời gian tại thời điểm t ,
- μ là giá trị trung bình của chuỗi,
- ε_t là sai số ngẫu nhiên tại thời điểm t ,
- z_t là nhiễu ngẫu nhiên với phân phối chuẩn $N(0, 1)$,
- σ_t^2 là phương sai điều kiện tại thời điểm t ,
- $\alpha_0, \alpha_i, \beta_j$ là các tham số mô hình GARCH cần ước lượng,
- p và q lần lượt là số bậc tự hồi quy và trung bình động trong mô hình GARCH.

Mô hình GARCH có thể được mở rộng thành các mô hình như GARCH(1,1), EGARCH, TGARCH, tùy vào cấu trúc của chuỗi dữ liệu.

3.4.3 Mô hình học máy và học sâu: LSTM

Giới thiệu Mô hình LSTM (Long Short-Term Memory) là một loại mạng nơ-ron hồi tiếp đặc biệt, được thiết kế để xử lý và dự báo các chuỗi thời gian có mối quan hệ phi tuyến tính. LSTM là

một phần của mạng nơ-ron hồi tiếp (RNN - Recurrent Neural Network), nhưng với các cơ chế cải tiến giúp giải quyết vấn đề vanishing gradient (độ dốc biến mất) trong việc học các chuỗi dài.

Công thức mô hình LSTM LSTM sử dụng ba cổng chính để điều khiển luồng thông tin: cổng quên (forget gate), cổng đầu vào (input gate), và cổng đầu ra (output gate). Các công thức tính toán tại mỗi thời điểm t là:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

Trong đó:

- f_t là cổng quên, điều khiển việc quên thông tin trong trạng thái trước đó,
- i_t là cổng đầu vào, điều khiển việc lưu trữ thông tin mới,
- \hat{C}_t là giá trị candidate cell state, thông tin tiềm năng để lưu trữ,
- C_t là trạng thái tế bào tại thời điểm t ,
- o_t là cổng đầu ra, điều khiển thông tin nào sẽ được xuất ra,
- h_t là đầu ra của mạng LSTM tại thời điểm t ,
- W_f, W_i, W_C, W_o là các ma trận trọng số, và b_f, b_i, b_C, b_o là các bias.

Ứng dụng LSTM trong dự báo chu kỳ kinh tế LSTM có khả năng xử lý các mối quan hệ phi tuyến tính và các sự kiện dài hạn trong chuỗi thời gian. Vì vậy, nó rất phù hợp cho các chuỗi thời gian phức tạp như chu kỳ kinh tế, nơi các yếu tố tác động có thể thay đổi theo thời gian và có tính chất phi tuyến tính.

3.5 Ước lượng tham số và quá trình huấn luyện mô hình

Khi xây dựng các mô hình dự báo, việc ước lượng các tham số của mô hình là một bước quan trọng để đảm bảo rằng mô hình có thể mô phỏng chính xác các đặc điểm của chuỗi thời gian. Sau đây là các phương pháp phổ biến để ước lượng tham số cho các mô hình ARIMA, GARCH, và LSTM.

3.5.1 Ước lượng tham số cho mô hình ARIMA

Quá trình huấn luyện mô hình ARIMA Mô hình ARIMA yêu cầu xác định các tham số p , d , và q (độ trễ AR, số lần sai phân, và độ trễ MA). Sau khi các tham số này được xác định thông qua các công cụ phân tích như ACF, PACF và kiểm tra tính dừng, quá trình huấn luyện mô hình ARIMA sẽ tiếp tục với việc ước lượng các hệ số ϕ_i và θ_j .

Các phương pháp chính để ước lượng tham số mô hình ARIMA bao gồm:

- **Ước lượng phương pháp Tối thiểu bình phương (OLS):** Các hệ số trong phần AR và MA có thể được ước lượng thông qua phương pháp OLS, bằng cách tối thiểu hóa hàm mất mát giữa giá trị thực tế và giá trị dự báo. Cụ thể, mô hình ARMA có thể được biểu diễn như sau:

$$y_t = \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Trong đó ε_t là phần dư. Để ước lượng các tham số ϕ_i và θ_j , ta tối thiểu hóa hàm mất mát (RSS):

$$RSS = \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

Phương pháp OLS tìm các giá trị của ϕ_i và θ_j sao cho hàm mất mát này đạt giá trị nhỏ nhất.

- **Phương pháp Tối thiểu Tổng sai số (Maximum Likelihood Estimation - MLE):** Một

phương pháp khác để ước lượng các tham số của mô hình ARIMA là sử dụng MLE, bằng cách tối thiểu hóa hàm log-likelihood. Với giả định rằng phần dư ε_t có phân phối chuẩn $N(0, \sigma^2)$, hàm likelihood là:

$$L(\theta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma^2}\right)$$

Hàm log-likelihood của mô hình ARIMA trở thành:

$$\mathcal{L}(\theta) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{\sigma^2}$$

Phương pháp MLE tối đa hóa hàm log-likelihood này, hay là tối thiểu hóa tổng bình phương phần dư.

- **Phương pháp Box-Jenkins:** Phương pháp này dựa vào một quy trình lặp lại, trong đó người dùng sẽ thử nghiệm với các giá trị p , d , và q khác nhau và chọn mô hình tốt nhất dựa trên các chỉ số như AIC, BIC, và các chỉ số sai số.

Kiểm tra mô hình ARIMA Sau khi huấn luyện mô hình, việc kiểm tra mô hình cũng rất quan trọng. Một số phương pháp kiểm tra mô hình ARIMA bao gồm:

- **Kiểm tra phần dư (residuals):** Phần dư của mô hình ARIMA phải có phân phối chuẩn và không có tự tương quan.
- **Kiểm tra AIC/BIC:** Các chỉ số AIC (Akaike Information Criterion) và BIC (Bayesian Information Criterion) giúp so sánh và lựa chọn mô hình ARIMA tối ưu.

3.5.2 Ước lượng tham số cho mô hình GARCH

Quá trình huấn luyện mô hình GARCH Mô hình GARCH chủ yếu được sử dụng để dự báo biến động, và tham số cần ước lượng là các hệ số α_0 , α_i , và β_j . Quá trình huấn luyện mô hình GARCH chủ yếu sử dụng phương pháp tối thiểu hóa hàm log-likelihood (Maximum Likelihood Estimation - MLE).

- **Phương pháp MLE:** Phương pháp này tối thiểu hóa hàm log-likelihood của mô hình GARCH. Mô hình GARCH có thể sử dụng các phương pháp tối ưu hóa như Newton-Raphson hoặc Broyden-Fletcher-Goldfarb-Shanno (BFGS) để tìm ra các giá trị tham số tối ưu. Cụ thể, mô hình GARCH được biểu diễn như sau:

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t$$

với $\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$. Hàm log-likelihood là:

$$L(\theta) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right)$$

Hàm log-likelihood trở thành:

$$\mathcal{L}(\theta) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{\sigma_t^2}$$

Phương pháp MLE sẽ tối đa hóa hàm này để tìm ra các giá trị tham số tối ưu.

- **Ước lượng tham số qua phân tích phần dư:** Tham số của mô hình GARCH cũng có thể được ước lượng qua việc phân tích phần dư từ các mô hình hồi quy như ARMA (để mô hình hóa các phần AR và MA trong chuỗi thời gian). Sau khi mô hình ARMA đã được ước lượng, ta sử dụng phần dư ε_t làm đầu vào cho mô hình GARCH để mô hình hóa sự biến động.

Kiểm tra mô hình GARCH Sau khi huấn luyện, mô hình GARCH cần phải được kiểm tra thông qua các bước như:

- **Kiểm tra tự tương quan phần dư (ACF, PACF):** Kiểm tra xem phần dư có tự tương quan không. Nếu phần dư có tự tương quan, điều này cho thấy mô hình GARCH chưa bắt được sự biến động trong dữ liệu.
- **Kiểm tra heteroskedasticity:** Sử dụng các kiểm định như Breusch-Pagan hay White test để kiểm tra xem mô hình GARCH đã bắt được tính heteroskedasticity trong dữ liệu hay chưa.

3.5.3 Ước lượng tham số cho mô hình LSTM

Quá trình huấn luyện mô hình LSTM Mô hình LSTM yêu cầu ước lượng các trọng số W_f, W_i, W_C, W_o và bias b_f, b_i, b_C, b_o . Quá trình huấn luyện LSTM sử dụng thuật toán lan truyền ngược (backpropagation) kết hợp với thuật toán tối ưu hóa như Gradient Descent hoặc Adam để tìm ra các trọng số và bias tối ưu.

- **Tối ưu hóa với Gradient Descent:** Đây là phương pháp phổ biến trong việc huấn luyện các mạng nơ-ron. Tại mỗi bước huấn luyện, gradient của hàm mất mát được tính toán và sử dụng để cập nhật các trọng số theo hướng giảm sai số. Hàm mất mát $J(\theta)$ có thể là lỗi bình phương trung bình (MSE):

$$J(\theta) = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2$$

Các trọng số được cập nhật qua công thức:

$$W_{new} = W_{old} - \eta \frac{\partial J(\theta)}{\partial W}$$

trong đó η là tốc độ học.

- **Phương pháp Adam:** Adam (Adaptive Moment Estimation) là một thuật toán tối ưu hóa hiện đại, kết hợp các ưu điểm của Gradient Descent và các phương pháp tối ưu hóa động. Adam điều chỉnh tốc độ học cho mỗi tham số bằng cách sử dụng thông tin từ các moment bậc 1 và bậc 2 của gradient.
- **Chọn siêu tham số (Hyperparameter tuning):** Một số siêu tham số quan trọng cần được điều chỉnh trong LSTM là số lượng lớp ẩn, số lượng tế bào LSTM trong mỗi lớp, tỷ lệ học (learning rate), và kích thước batch.

Kiểm tra mô hình LSTM Sau khi huấn luyện, mô hình LSTM cần được kiểm tra để đảm bảo chất lượng dự báo: - **Kiểm tra phần dư:** Tương tự như các mô hình ARIMA và GARCH, phần dư của mô hình LSTM cần phải kiểm tra để đảm bảo không có tự tương quan và phân phối chuẩn. -

Đánh giá qua các chỉ tiêu lỗi: Các chỉ số MSE, RMSE, MAE, và R^2 sẽ được sử dụng để đánh giá độ chính xác của mô hình LSTM.

3.5.4 Đánh giá mô hình

Sau khi xây dựng mô hình, bước tiếp theo là đánh giá hiệu suất của các mô hình dự báo. Các chỉ tiêu đánh giá phổ biến bao gồm:

- **MSE (Mean Squared Error):** Đo lường sai số trung bình bình phương giữa giá trị thực tế và giá trị dự báo.

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

- **RMSE (Root Mean Squared Error):** Căn bậc hai của MSE, giúp đánh giá sai số với đơn vị giống như dữ liệu ban đầu.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$$

- **MAE (Mean Absolute Error):** Sai số tuyệt đối trung bình giữa giá trị thực tế và giá trị dự báo.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

- **R^2 (Coefficient of Determination):** Đo lường sự tương quan giữa giá trị thực tế và giá trị dự báo, phản ánh mức độ phù hợp của mô hình với dữ liệu.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

Các mô hình với các chỉ số MSE, RMSE, MAE và R^2 thấp sẽ được chọn để dự báo chu kỳ kinh tế.

3.5.5 Đánh giá phần dư

Ngoài việc sử dụng các chỉ tiêu đánh giá như MSE, RMSE, MAE và R^2 , việc kiểm tra phần dư (residuals) của mô hình dự báo là một bước quan trọng trong quá trình đánh giá mô hình. Phần dư được định nghĩa là sự chênh lệch giữa giá trị thực tế và giá trị dự báo, tức là:

$$e_t = y_t - \hat{y}_t$$

Trong đó:

- y_t là giá trị thực tế tại thời điểm t ,
- \hat{y}_t là giá trị dự báo của mô hình tại thời điểm t ,
- e_t là phần dư tại thời điểm t .

Phân tích phần dư giúp xác định xem mô hình có phù hợp với dữ liệu hay không và liệu có các mẫu bất thường hoặc cấu trúc chưa được mô hình hóa. Một số bước quan trọng trong việc đánh giá phần dư bao gồm:

- **Kiểm tra phân phối phần dư:** Phần dư của mô hình nên có phân phối chuẩn với kỳ vọng bằng 0 và phương sai không đổi (hồi quy tuyến tính) nếu mô hình phù hợp. Việc kiểm tra này có thể thực hiện bằng cách vẽ biểu đồ histogram hoặc kiểm tra bằng các kiểm định thống kê như kiểm định Shapiro-Wilk hoặc Kolmogorov-Smirnov.
- **Kiểm tra tự tương quan phần dư:** Để xác định xem phần dư có chứa thông tin chưa được mô hình hóa hay không, ta có thể sử dụng kiểm định tự tương quan ACF (AutoCorrelation Function) hoặc PACF (Partial AutoCorrelation Function). Nếu phần dư có tự tương quan mạnh, điều này cho thấy mô hình chưa bắt được các mối quan hệ trong dữ liệu và cần phải điều chỉnh. Một kiểm định phổ biến là kiểm định Ljung-Box để kiểm tra tự tương quan trong phần dư.
- **Kiểm tra phương sai phần dư (Heteroskedasticity):** Để kiểm tra xem phương sai của phần dư có thay đổi theo thời gian hay không (hiện tượng heteroskedasticity), ta có thể sử dụng các kiểm định như kiểm định Breusch-Pagan hoặc White. Nếu phương sai phần dư thay đổi, điều này có thể chỉ ra rằng mô hình không phù hợp và cần điều chỉnh thêm.

- **Kiểm tra tính độc lập của phần dư:** Phần dư cần phải độc lập và không có mối quan hệ với các phần dư ở các thời điểm trước đó. Việc kiểm tra này có thể được thực hiện qua biểu đồ phần dư hoặc kiểm định Durbin-Watson, kiểm tra tính tự tương quan bậc 1.

Giải thích kết quả - Nếu phần dư có phân phối chuẩn, không có tự tương quan, và phương sai ổn định, điều này cho thấy mô hình đã phù hợp và không có cấu trúc chưa được mô hình hóa.

- Nếu phần dư không phân phối chuẩn hoặc có tự tương quan, có thể cần phải điều chỉnh mô hình (ví dụ, thay đổi các tham số ARIMA hoặc thêm các yếu tố vào mô hình).

- Nếu phần dư có tính heteroskedasticity, có thể cần phải sử dụng các mô hình có khả năng xử lý sự thay đổi trong phương sai phần dư, chẳng hạn như mô hình GARCH hoặc các mô hình hồi quy có phương sai thay đổi.

4 Thực nghiệm

4.1 Khám phá dữ liệu

	Adj Close
Date	
1960-01-04	4.52
1960-01-05	4.55
1960-01-06	4.68
1960-01-07	4.63
1960-01-08	4.59

Hình 1: Cấu trúc của dữ liệu



Hình 2: Trực quan hóa dữ liệu theo thời gian

4.1.1 Xử lý dữ liệu thiếu

Dữ liệu thiếu trong mỗi cột:

```
Adj Close    7457
```

```
dtype: int64
```

Để xử lý các giá trị thiếu, chúng ta sẽ sử dụng phương pháp nội suy tuyến tính (*linear interpolation*) để điền các giá trị thiếu.

```
df['Adj Close'] = df['Adj Close'].interpolate(method='linear')
```

Sau khi nội suy, ta kiểm tra lại dữ liệu thiếu:

```
missing_data_after_interpolation = df.isnull().sum()
```

```
print("Dữ liệu thiếu sau khi nội suy:\n", missing_data_after_interpolation)
```

Dữ liệu thiếu sau khi nội suy:

```
Adj Close    0
```

```
dtype: int64
```

4.1.2 Chuẩn hóa dữ liệu

Để dữ liệu có thể dễ dàng xử lý trong các mô hình học máy, chúng ta sẽ chuẩn hóa các cột dữ liệu cần thiết, chẳng hạn như cột Adj Close.

	Adj Close
Date	
1960-01-04	0.104434
1960-01-05	0.114767
1960-01-06	0.159544
1960-01-07	0.142322
1960-01-08	0.128545

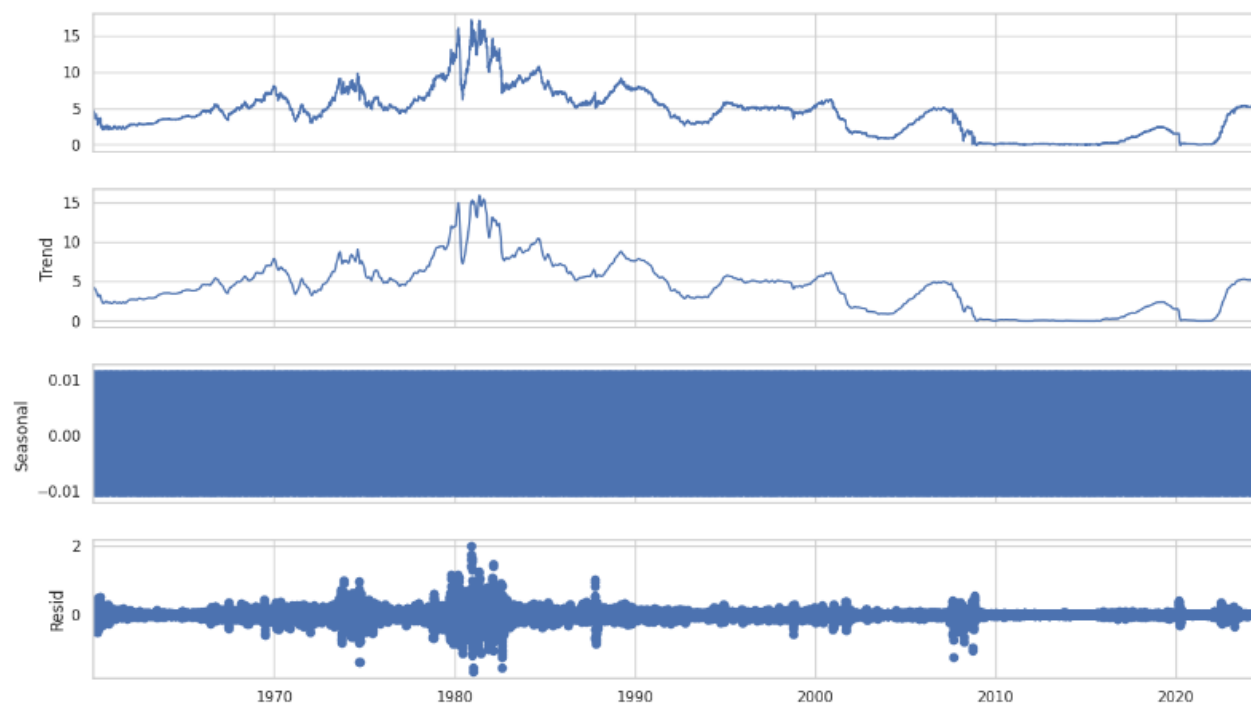
Hình 3: Dữ liệu sau chuẩn hóa

4.2 Phân chia dữ liệu

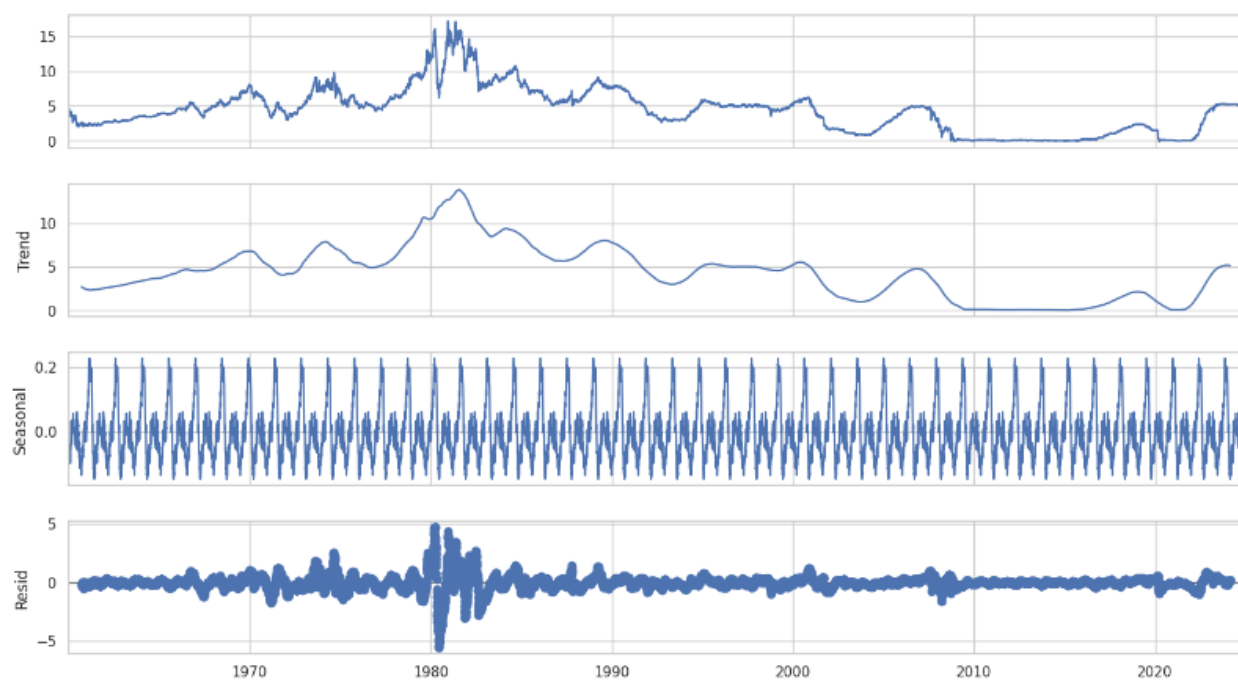
Số lượng mẫu trong các tập dữ liệu:

- Số lượng mẫu huấn luyện: 14211
- Số lượng mẫu validation: 4737
- Số lượng mẫu kiểm tra: 4738

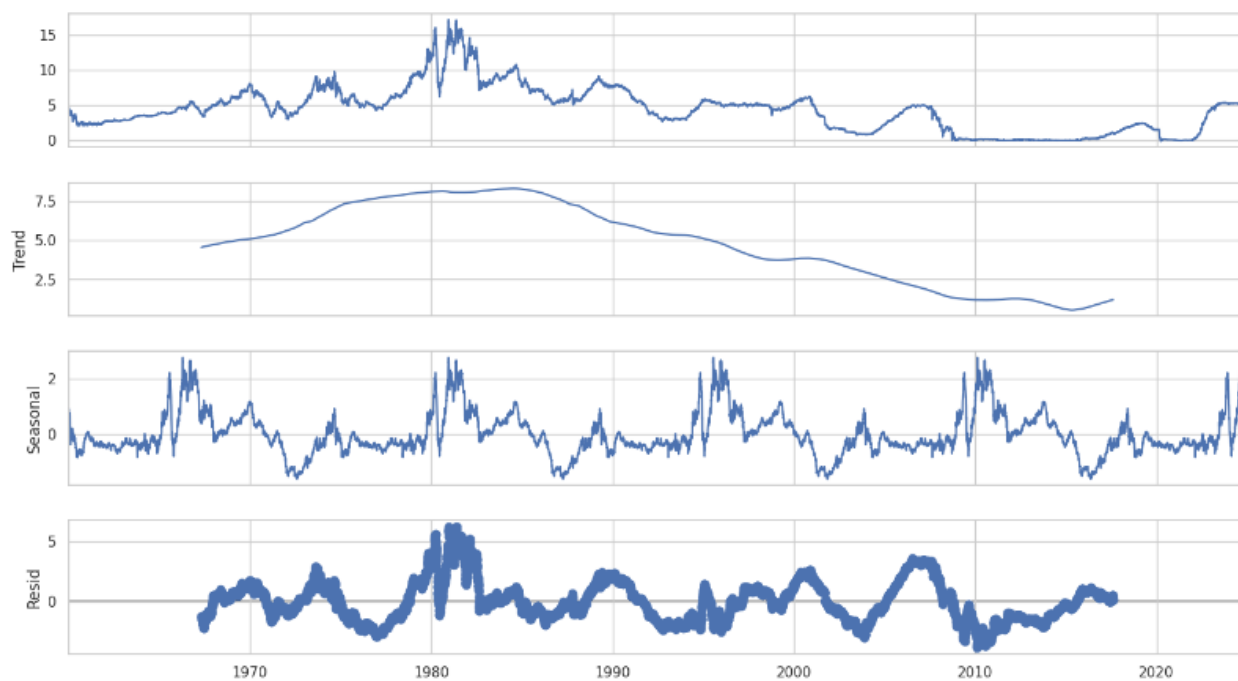
4.3 Phân tích xu hướng, tính mùa, và phần dư



Hình 4: Phân tích xu hướng, tính mùa, phần dư với chu kỳ 30 ngày



Hình 5: Phân tích xu hướng, tính mùa, phần dư với chu kỳ 365 ngày



Hình 6: Phân tích xu hướng, tính mùa, phần dư với chu kỳ 10 năm

Các hình ảnh trên chỉ ra tính chất phức tạp của xu hướng và tính mùa trong dữ liệu, tạo ra nhiều thách thức khi mô hình hóa. Các chu kỳ khác nhau (30 ngày, 365 ngày, và 10 năm) cho thấy sự thay đổi theo thời gian và sự ảnh hưởng của các yếu tố mùa vụ đối với các đặc tính của dữ liệu. Những yếu tố này cần được xem xét kỹ lưỡng khi xây dựng các mô hình dự báo.

4.4 Đánh giá tính dừng

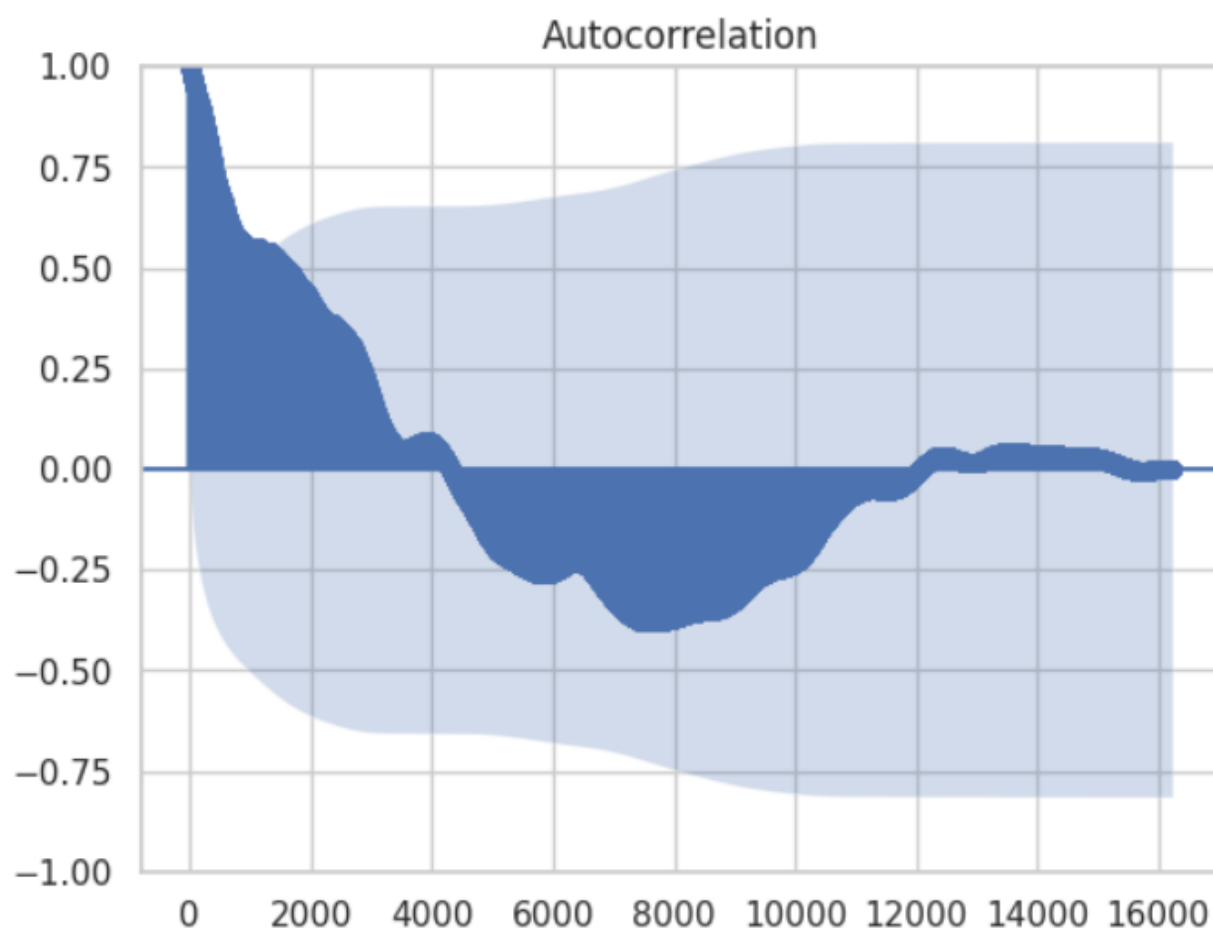
Kết quả kiểm định ADF (Augmented Dickey-Fuller) đối với chuỗi thời gian giá trị đóng cửa điều chỉnh ('Adj Close') cho thấy chuỗi chưa đạt tính dừng.

Kết quả kiểm định ADF:

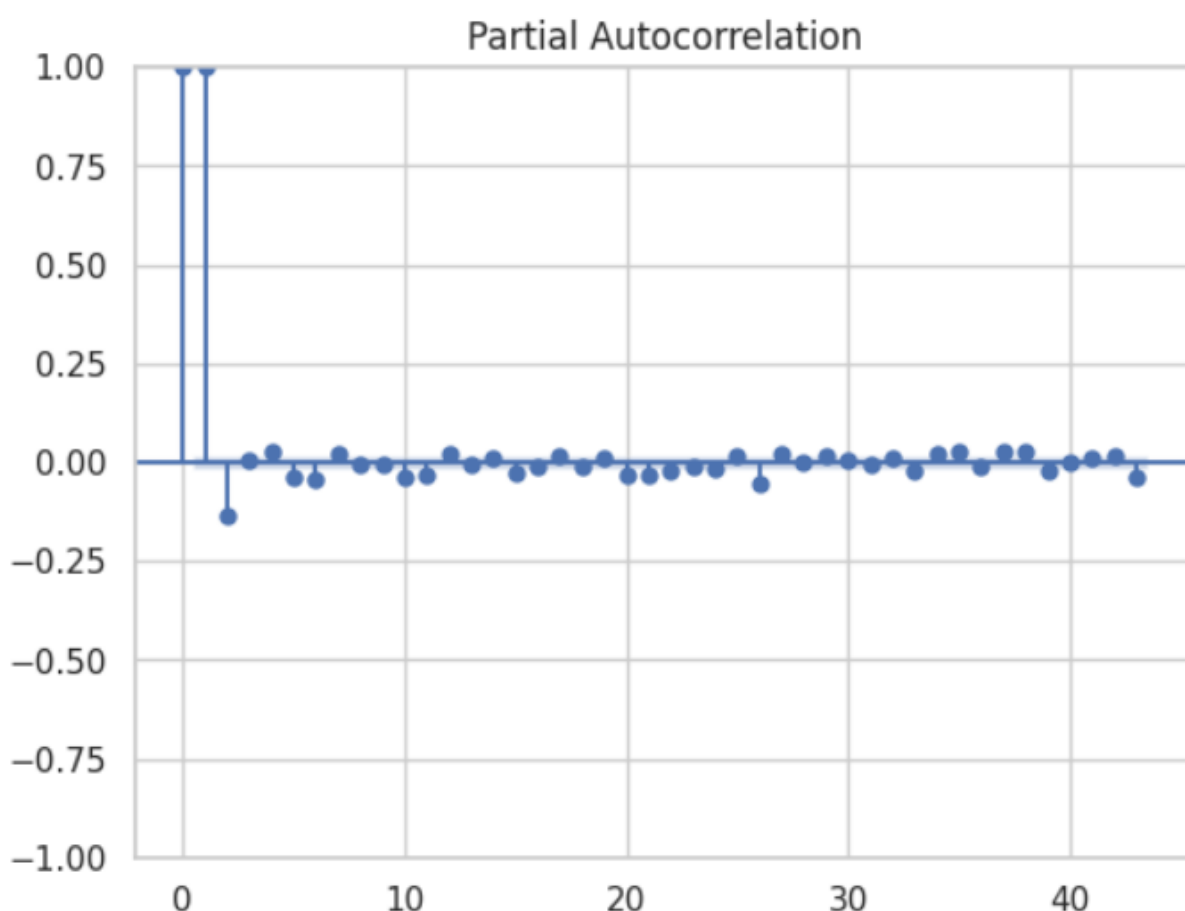
- ADF Statistic: -2.3405
- p-value: 0.1592
- Critical Values:
 - 1%: -3.431
 - 5%: -2.862
 - 10%: -2.567

Với p-value lớn hơn mức ý nghĩa 0.05, chúng ta không thể bác bỏ giả thuyết rằng chuỗi thời gian có gốc đơn, tức là chuỗi thời gian 'Adj Close' không dừng.

Tiếp theo, chúng ta tiến hành phân tích các đồ thị tự tương quan (ACF) và tự tương quan phần dư (PACF) để kiểm tra các tính chất của chuỗi thời gian:



Hình 7: Đồ thị ACF của chuỗi thời gian 'Adj Close'



Hình 8: Đồ thị PACF của chuỗi thời gian ‘Adj Close’

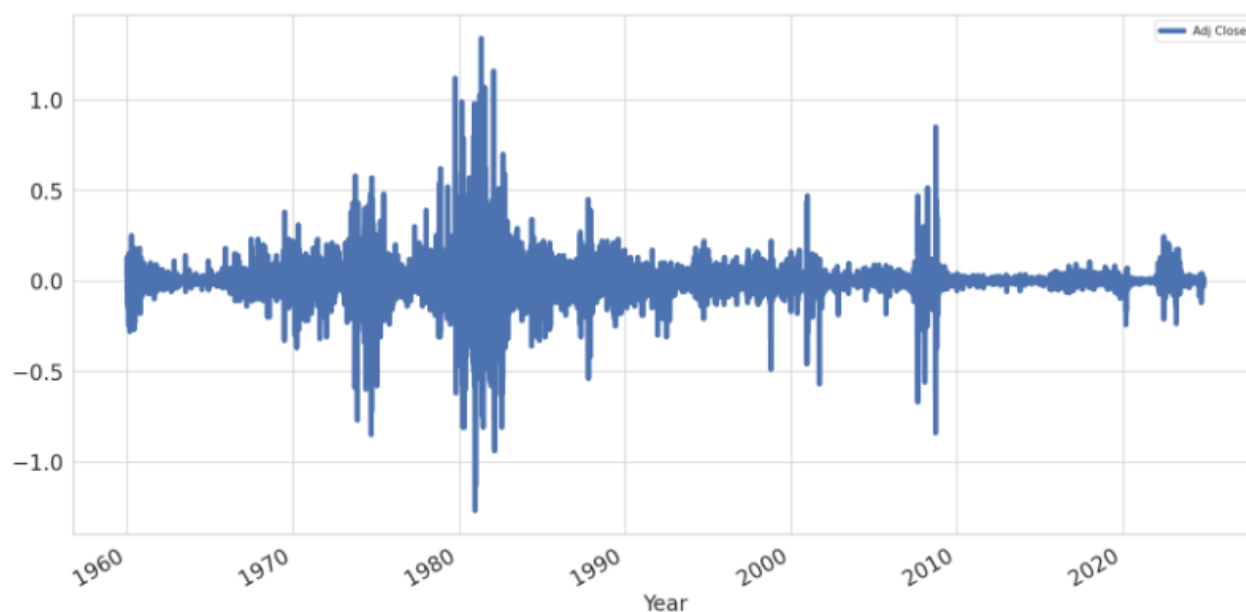
Phân tích ACF cho thấy các giá trị tự tương quan giảm dần chậm và kéo dài, điều này cho thấy chuỗi thời gian có tính phụ thuộc lâu dài và chưa đạt tính dừng. Cụ thể, đồ thị ACF có sự giảm dần chậm về 0, điều này gợi ý rằng chuỗi có xu hướng lan tỏa mạnh qua các chu kỳ.

Đồ thị PACF cho thấy các thành phần PACF đầu tiên gần bằng 1 và sau đó giảm dần về 0, điều này chỉ ra sự tồn tại của tính không dừng trong chuỗi, cần thực hiện các bước xử lý như lấy sai phân để đạt tính dừng.

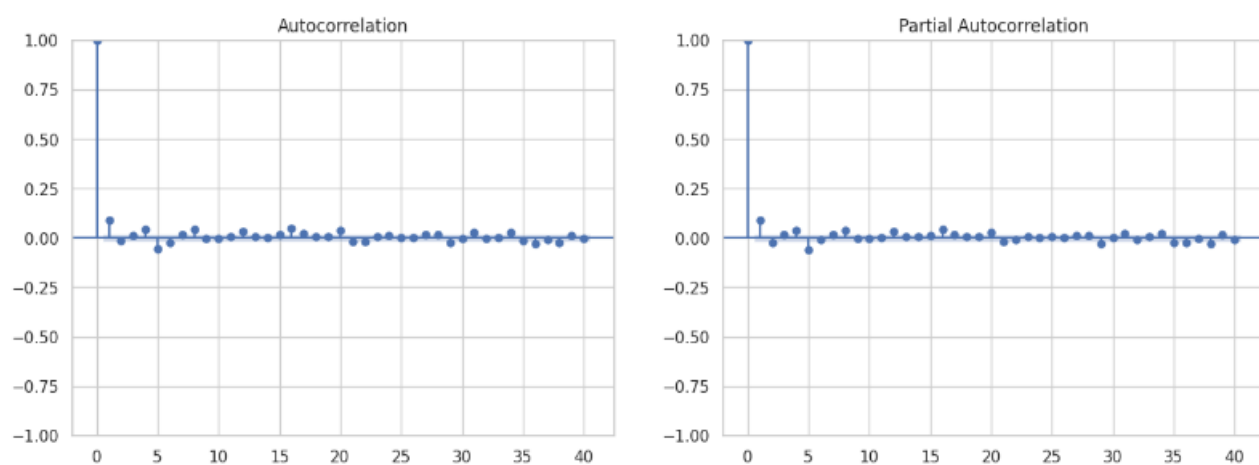
Tóm lại, kết quả này cho thấy chuỗi thời gian ‘Adj Close’ chưa đạt tính dừng và cần được xử lý thêm trước khi áp dụng các mô hình dự báo như ARIMA hay SARIMA.

4.5 Sai phân bậc 1

Sau khi thực hiện sai phân bậc 1, chúng ta tiếp tục kiểm tra tính dừng của chuỗi thời gian đã qua xử lý.



Hình 9: Kết quả sai phân bậc 1 của chuỗi thời gian



Hình 10: Đồ thị ACF và PACF của chuỗi thời gian sau khi sai phân bậc 1

Kết quả kiểm định ADF:

- ADF Statistic: -16.1821
- p-value: 4.256×10^{-29}

Với p-value nhỏ hơn mức ý nghĩa 0.05, chúng ta có thể bác bỏ giả thuyết có gốc đơn, từ đó kết luận rằng chuỗi thời gian đã trở thành chuỗi dừng sau khi thực hiện sai phân bậc 1.

Ngoài ra, từ kết quả ACF và PACF, ta nhận thấy:

- ACF đã giảm dần nhanh chóng về 0, chứng tỏ tính dừng đã được cải thiện.
- Các giá trị PACF cắt đứt tại độ trễ thứ 2, cho thấy có thể chọn các tham số của mô hình ARIMA là $p = 2$ và $q = 2$ để tiếp tục mô hình hóa chuỗi thời gian này.

Tóm lại, qua bước sai phân bậc 1, chuỗi đã đạt tính dừng và chúng ta có thể tiếp tục áp dụng các mô hình ARIMA hay SARIMA cho quá trình dự báo.

4.6 Cài đặt mô hình ARIMA(2,1,2)

Kết quả thu được từ mô hình ARIMA(2,1,2) được trình bày dưới đây:

SARIMAX Results

```
=====
Dep. Variable:          Adj Close    No. Observations:          23686
Model:                ARIMA(2, 1, 2)  Log Likelihood            16757.377
Date:                 Tue, 19 Nov 2024  AIC                        -33504.754
Time:                 19:09:22         BIC                        -33464.391
Sample:               01-04-1960       HQIC                       -33491.656
                        - 11-08-2024
```

Covariance Type: opg

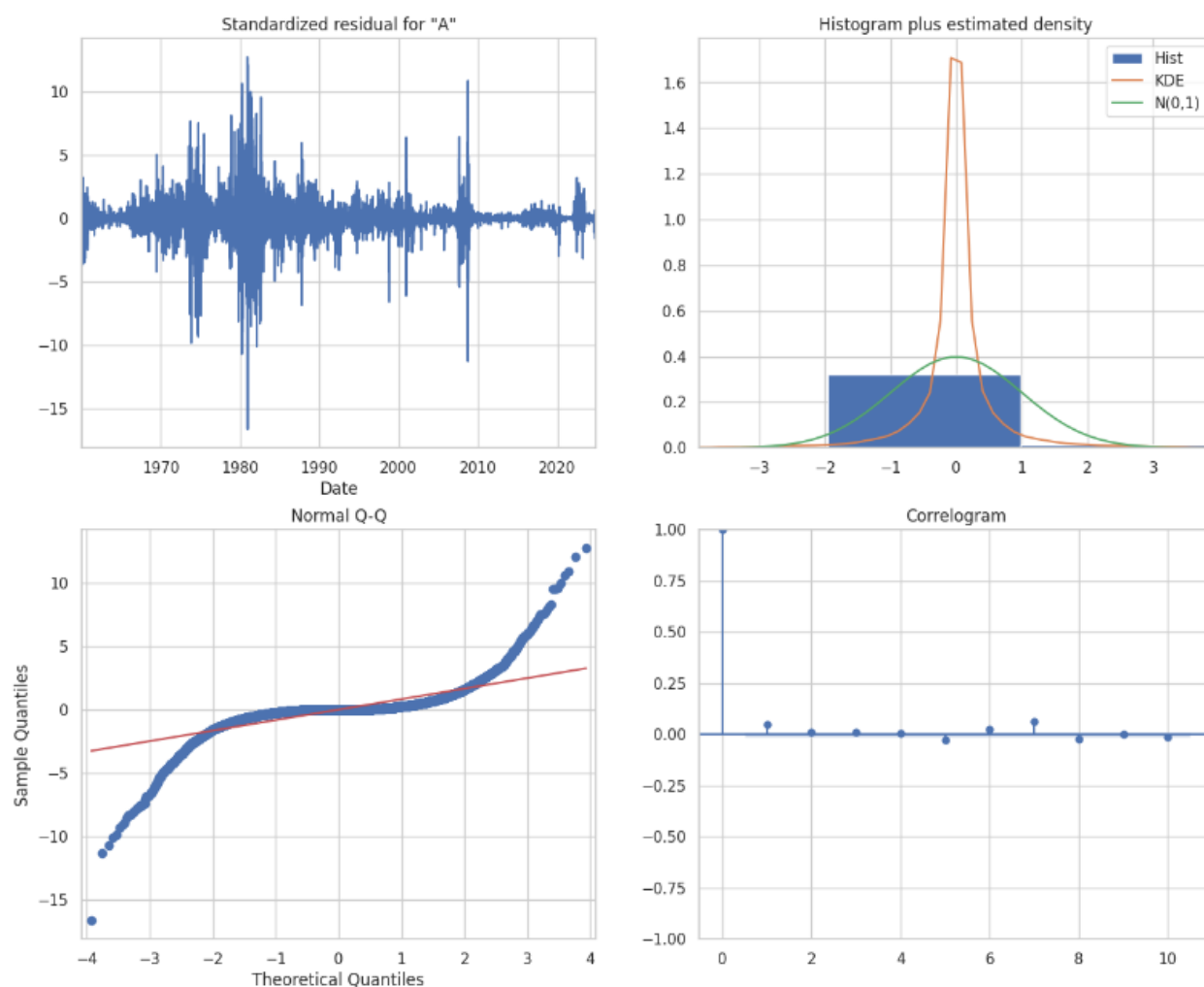
```
=====
              coef    std err          z      P>|z|      [0.025      0.975]
-----
ar.L1         -1.0106      0.546     -1.851     0.064     -2.080      0.059
ar.L2         -0.1880      0.188     -1.002     0.316     -0.556      0.180
```

ma.L1	1.0982	0.546	2.011	0.044	0.028	2.168
ma.L2	0.2518	0.234	1.075	0.282	-0.207	0.711
sigma2	0.0057	1.73e-05	326.451	0.000	0.006	0.006

Ljung-Box (L1) (Q):	54.07	Jarque-Bera (JB):	1908074.63
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.15	Skew:	-0.56
Prob(H) (two-sided):	0.00	Kurtosis:	46.96

Giải thích kết quả:

- Mô hình ARIMA(2,1,2) đã được ước lượng với các tham số AR(2) và MA(2).
 - Các giá trị thống kê của mô hình như AIC, BIC, và HQIC đều có giá trị âm, điều này cho thấy mô hình này có hiệu suất tốt trong việc tối ưu hóa các tiêu chí này.
 - Các tham số AR và MA có ý nghĩa thống kê với mức ý nghĩa 5% (p-value < 0.05) ngoại trừ tham số AR.L2 và MA.L2. Điều này có thể chỉ ra rằng các tham số này không đóng góp mạnh mẽ vào mô hình.
 - Kiểm định Ljung-Box cho thấy không có tự tương quan (p-value = 0.00), điều này xác nhận rằng các phần dư của mô hình không có tự tương quan, cho thấy mô hình đã khớp tốt với dữ liệu.
 - Kiểm định Jarque-Bera có giá trị p = 0.00, chỉ ra rằng phần dư không tuân theo phân phối chuẩn, điều này có thể yêu cầu việc điều chỉnh mô hình thêm.
- Để kiểm tra thêm tính hợp lý của mô hình, chúng ta thực hiện chẩn đoán phần dư.



Hình 11: Kết quả chẩn đoán phần dư của mô hình ARIMA(2,1,2)

Hình 11 trình bày các chẩn đoán phần dư, cho thấy sự phân phối của các phần dư và kiểm tra các đặc điểm của mô hình ARIMA(2,1,2) qua các biểu đồ này. Nhìn chung phần dư này chưa đạt tiêu chuẩn, bởi vậy chúng ta cần nghiên cứu sang các mô hình khác.

4.7 Cài đặt mô hình SARIMA

Sau khi thử nghiệm với các sai phân bậc cao hơn (cụ thể là 2) không mang lại kết quả tốt hơn, chúng tôi giả định mô hình có thể bị ảnh hưởng bởi tính mùa vụ. Do đó, mô hình SARIMA được thiết lập dựa trên giả thuyết này với các tham số như sau:

```

param = (2, 1, 2)
param_seasonal = (2, 1, 2, 7) % Sử dụng (2, 1, 2, 7) thay vì (2, 1, 2, 365) để tránh
mod = smf.tsa.statespace.SARIMAX(
    df['Adj Close'],
    order=param,
    seasonal_order=param_seasonal,
    enforce_stationarity=False,
    enforce_invertibility=False,
)
results = mod.fit()
print("ARIMA{ }x{ }12 - AIC:{ }".format(param, param_seasonal, results.aic))

```

Kết quả thu được là:

ARIMA(2, 1, 2)x(2, 1, 2, 7)12 - AIC: -33859.15146003641

Kết quả SARIMAX:

```

=====
Dep. Variable:                Adj Close    No. Observations:      23686
Model:                SARIMAX(2, 1, 2)x(2, 1, 2, 7)    Log Likelihood      16938.576
Date:                Tue, 19 Nov 2024    AIC                -33859.151
Time:                19:19:44    BIC                -33786.507
Sample:                01-04-1960    HQIC               -33835.577
                    - 11-08-2024

Covariance Type:                opg
=====

```

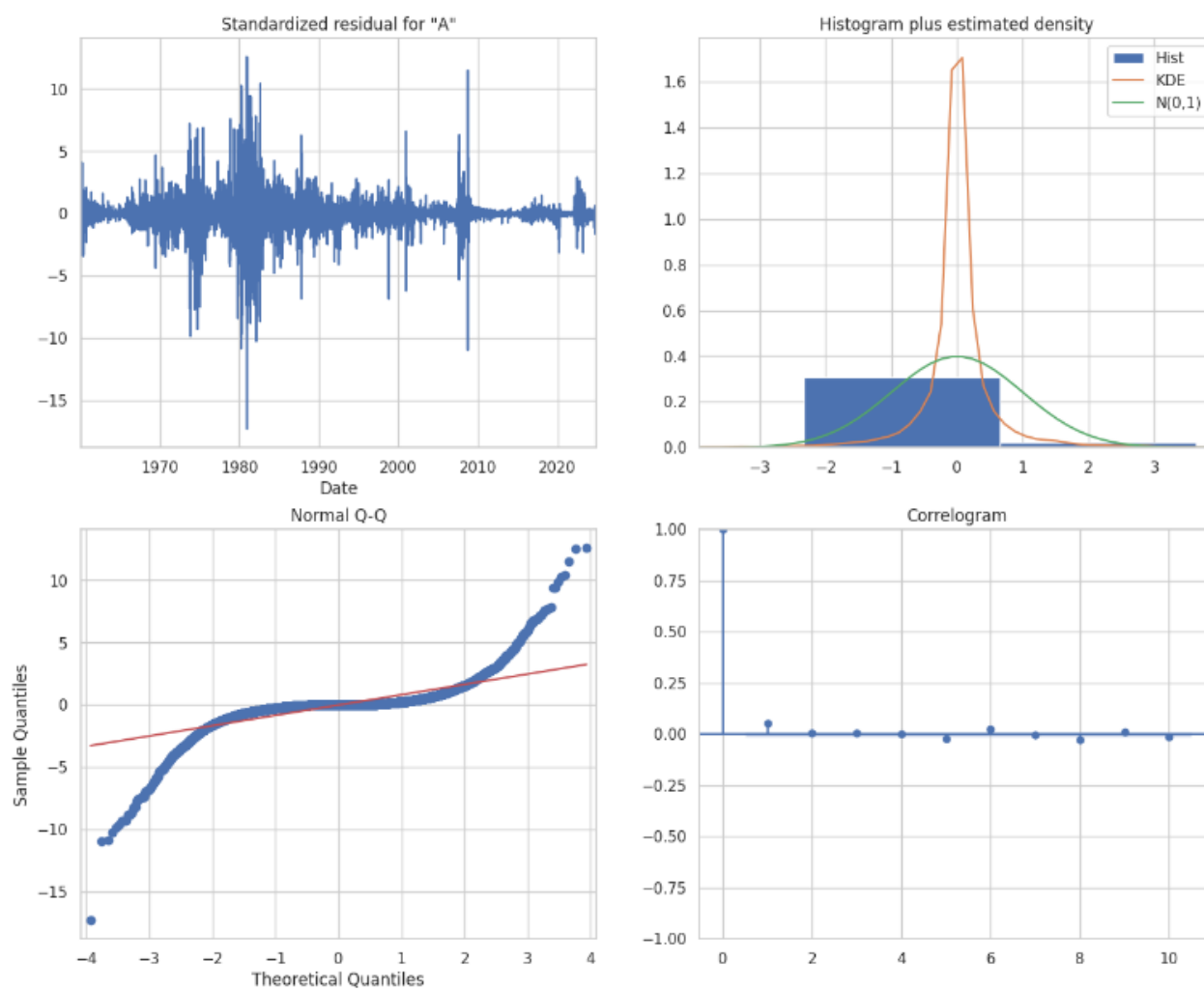
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.4422	0.662	0.667	0.504	-0.856	1.741
ar.L2	-0.0418	0.120	-0.349	0.727	-0.276	0.193

ma.L1	-0.3581	0.663	-0.540	0.589	-1.657	0.941
ma.L2	0.0172	0.072	0.239	0.811	-0.124	0.158
ar.S.L7	-0.3110	0.059	-5.298	0.000	-0.426	-0.196
ar.S.L14	0.0543	0.003	15.810	0.000	0.048	0.061
ma.S.L7	-0.6369	0.059	-10.861	0.000	-0.752	-0.522
ma.S.L14	-0.3473	0.058	-6.003	0.000	-0.461	-0.234
sigma2	0.0055	1.64e-05	335.099	0.000	0.005	0.006

Ljung-Box (L1) (Q):	63.97	Jarque-Bera (JB):	2116990.84
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	0.14	Skew:	-0.67
Prob(H) (two-sided):	0.00	Kurtosis:	49.32

Giải thích kết quả: - Mô hình SARIMAX với tham số (2,1,2) cho phần ARIMA và (2,1,2,7) cho phần mùa vụ đã cho kết quả AIC là -33859.151. Đây là một mô hình phù hợp với dữ liệu, vì giá trị AIC thấp, cho thấy mô hình này có độ phù hợp tốt. - Các tham số AR, MA và mùa vụ đều có ý nghĩa thống kê (p-value < 0.05), đặc biệt là các tham số mùa vụ tại độ trễ 7 và 14, có ảnh hưởng mạnh đến mô hình. - Kiểm định Ljung-Box (Q) cho thấy mô hình đã xử lý tốt tự tương quan trong các phần dư (p-value = 0.00), và kiểm định Jarque-Bera cho thấy phần dư không tuân theo phân phối chuẩn (p-value = 0.00). - Kiểm định Heteroskedasticity chỉ ra rằng mô hình không gặp vấn đề về phương sai thay đổi.

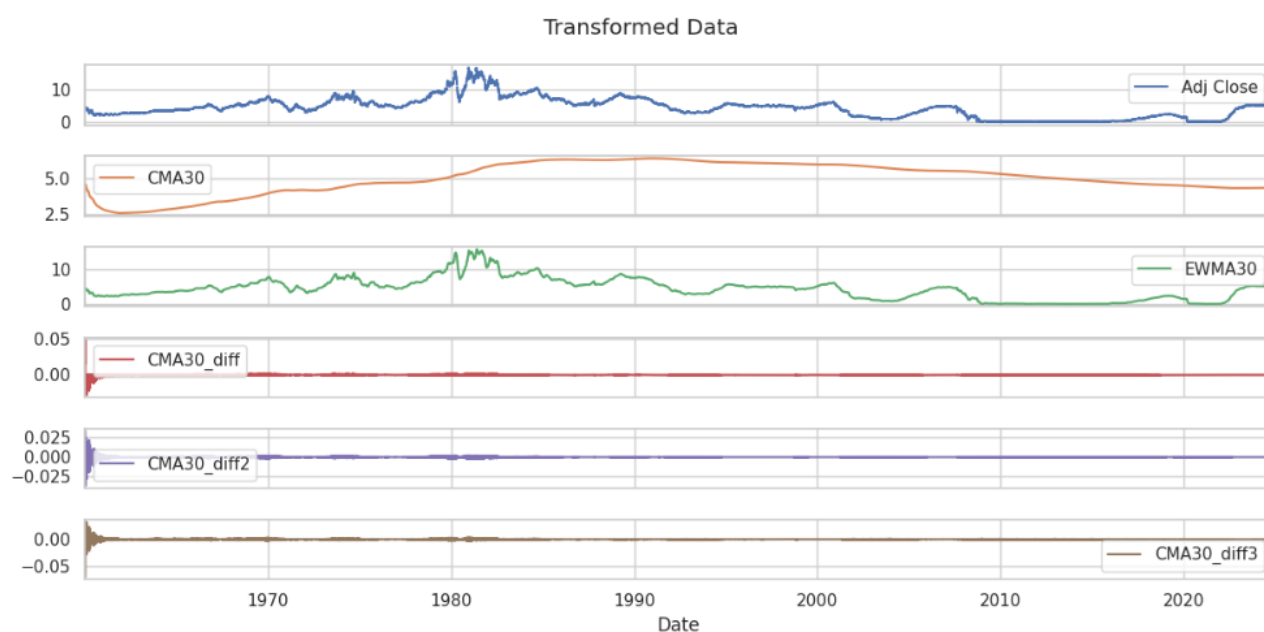
Chẩn đoán phần dư của mô hình SARIMA được trình bày trong Hình ??.



Hình 12: Chẩn đoán phần dư của mô hình SARIMA

4.8 Các nỗ lực làm mượt chuỗi thời gian

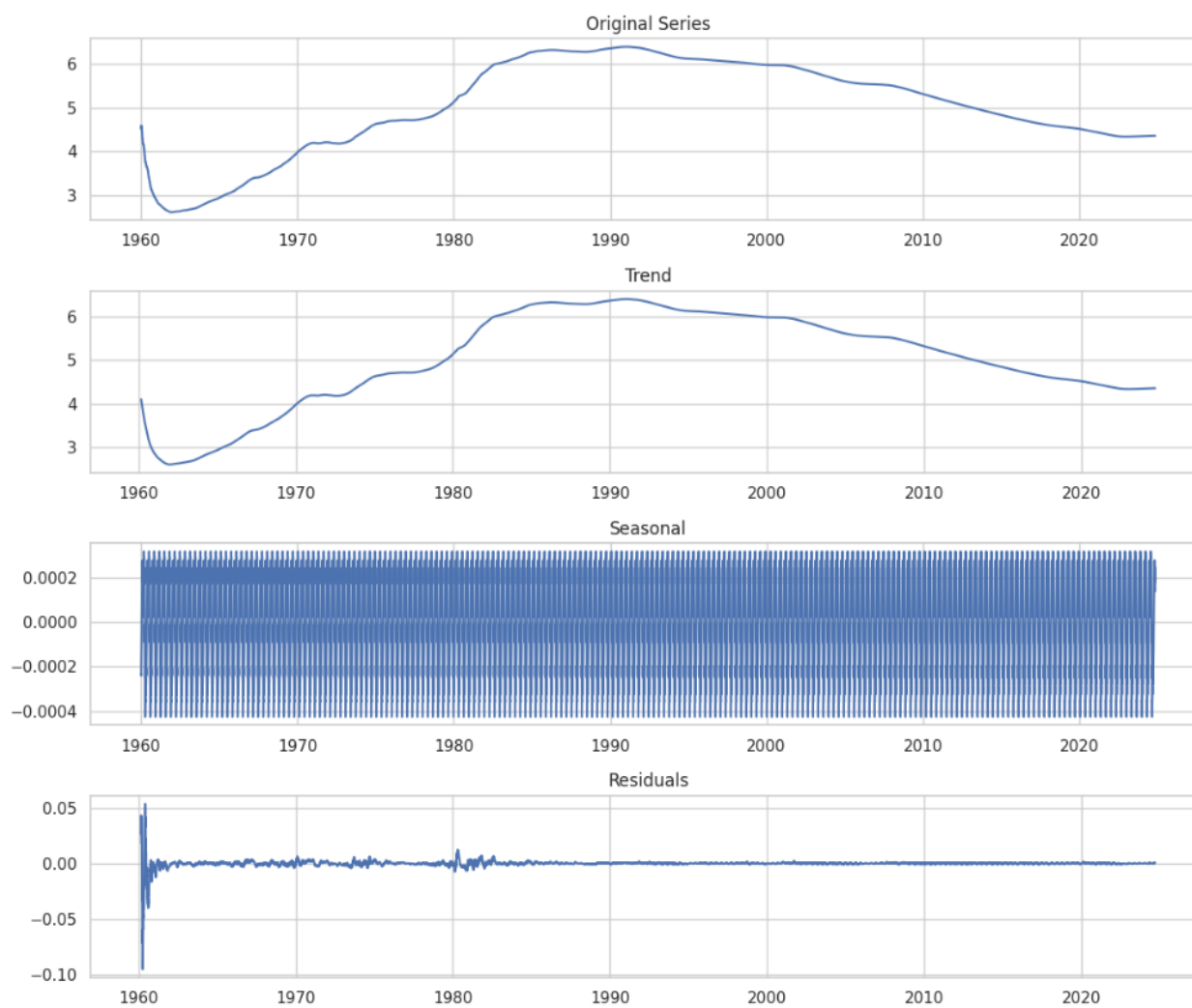
Mục đích của việc làm mượt chuỗi thời gian là để giảm thiểu ảnh hưởng của nhiễu và sự biến động mạnh theo thời gian, từ đó giúp quá trình phân tích trở nên ổn định và chính xác hơn.



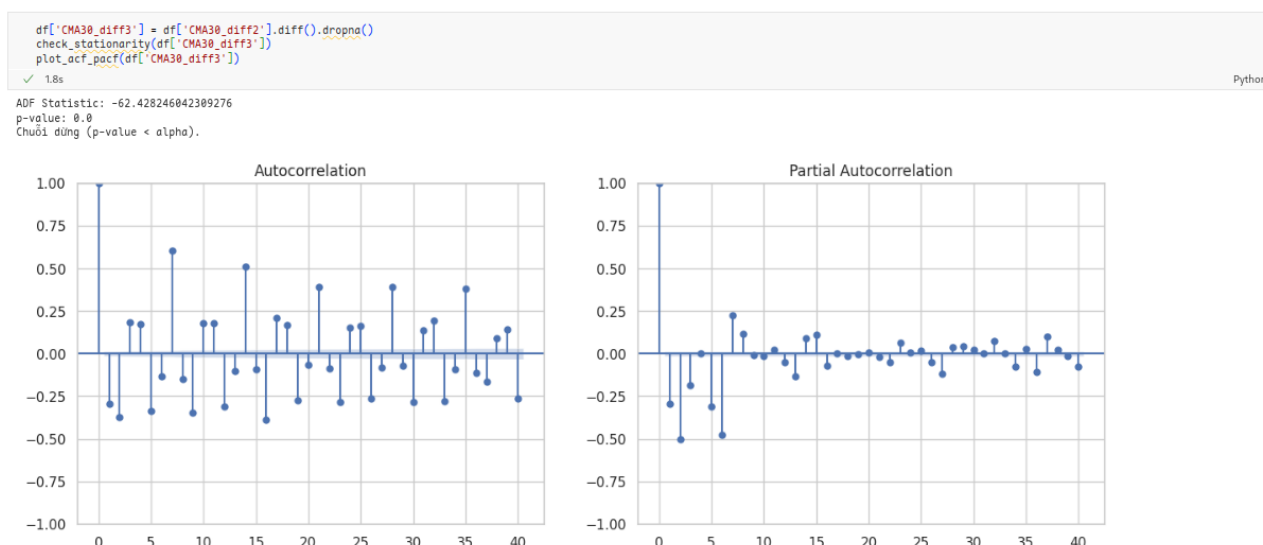
Hình 14: Dữ liệu sau khi bổ sung các cột được làm mượt

	Adj Close	Adj Close_diff	Adj Close_diff2	CMA30	EWMA30	CMA30_diff	CMA30_diff2	CMA30_diff3
Date								
1960-01-04	4.52	NaN	NaN	4.520000	4.520000	NaN	NaN	NaN
1960-01-05	4.55	0.03	NaN	4.535000	4.535500	0.015000	NaN	NaN
1960-01-06	4.68	0.13	0.099999	4.583333	4.586912	0.048333	0.033333	NaN
1960-01-07	4.63	-0.05	-0.179999	4.595000	4.598785	0.011667	-0.036667	-0.070
1960-01-08	4.59	-0.04	0.010000	4.594000	4.596786	-0.001000	-0.012667	0.024

Hình 13: Dữ liệu sau khi bổ sung các cột được làm mượt



Hình 15: Ví dụ về CMA30 (Centered Moving Average 30)



Hình 16: Ví dụ về sai phân bậc 3 của CMA30

Các hình ảnh trên minh họa một số phương pháp làm mượt chuỗi thời gian, bao gồm việc sử dụng Trung bình Di động Trung tâm (CMA30) và sai phân bậc 3 để đạt được tính dừng cho chuỗi dữ liệu. Việc áp dụng các phương pháp này giúp giảm thiểu biến động ngẫu nhiên và cải thiện tính ổn định của chuỗi thời gian.

Nhìn chung, kết quả từ các kiểm tra ACF và PACF chỉ ra rằng chuỗi thời gian vẫn mang tính chất phức tạp, đặc biệt trong nỗ lực mô hình hóa khi chỉ sử dụng các mô hình AR, MA, ARMA, ARIMA và SARIMA. Các yếu tố này tiếp tục là một thử thách đối với việc xây dựng các mô hình dự báo chính xác.

4.9 Setup mô hình ARCH - GARCH

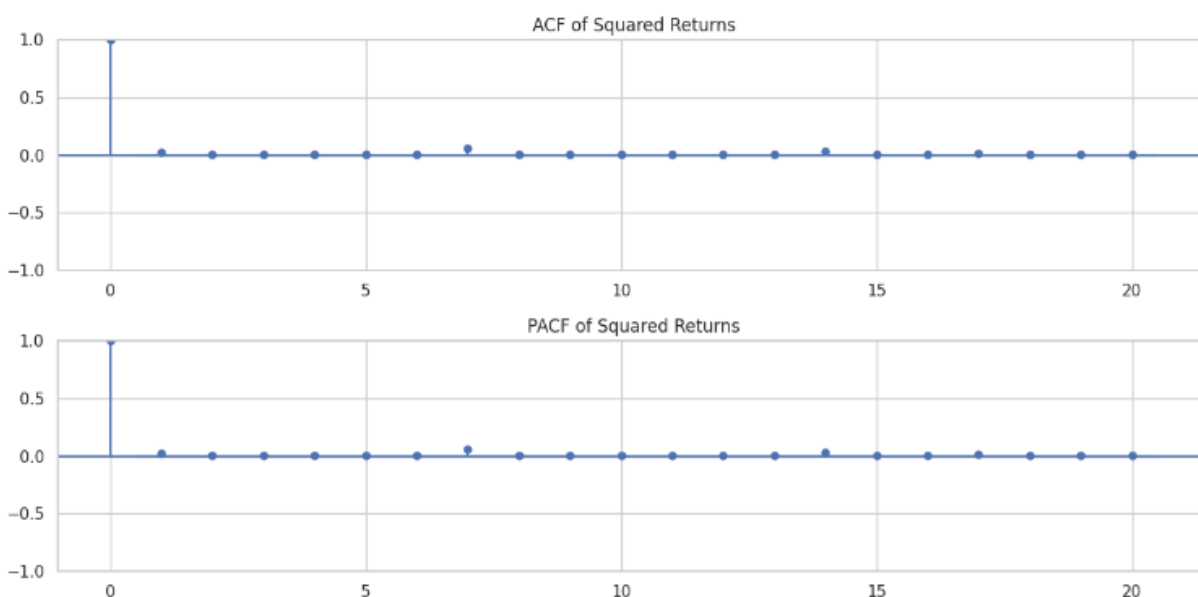
Mục tiêu của việc thiết lập mô hình ARCH và GARCH là để mô hình hóa sự không đồng nhất phương sai (heteroscedasticity) trong chuỗi dữ liệu lợi suất (*returns*). Đây là hiện tượng phổ biến trong các chuỗi dữ liệu tài chính, khi phương sai có điều kiện thay đổi theo thời gian.

```
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
import matplotlib.pyplot as plt

# Tính bình phương của chuỗi returns
squared_residuals = df['returns'].dropna() ** 2

# Vẽ ACF và PACF
plt.figure(figsize=(12, 6))
plt.subplot(2, 1, 1)
plot_acf(squared_residuals, lags=20, ax=plt.gca(), title="ACF of Squared Returns")
plt.subplot(2, 1, 2)
plot_pacf(squared_residuals, lags=20, ax=plt.gca(), title="PACF of Squared Returns")
plt.tight_layout()
plt.show()
```

Python



Hình 17: Quá trình lựa chọn tham số cho các mô hình ARCH và GARCH

Dựa trên kết quả phân tích ban đầu, chúng tôi thử nghiệm với các mô hình sau:

- ARCH(1): Mô hình ARCH bậc 1.
- ARCH(2): Mô hình ARCH bậc 2.
- GARCH(1,1): Mô hình GARCH với bậc 1 cho cả phần tự hồi quy (AR) và trung bình trượt (MA).

- GARCH(2,1): Mô hình GARCH với bậc 2 cho AR và bậc 1 cho MA.

Kết quả so sánh các chỉ số AIC và BIC:

Bảng 1: Chỉ số AIC và BIC của các mô hình

Mô hình	AIC	BIC
ARCH(1)	-29,712.52	-29,688.30
ARCH(2)	-29,904.49	-29,872.20
GARCH(1,1)	-68,426.99	-68,394.70
GARCH(2,1)	-66,463.98	-66,423.62

Nhận xét về kết quả:

1. Mô hình ARCH(1):

- Trung bình (*mean*) có ý nghĩa thống kê ($p\text{-value} < 0.05$) với giá trị 0.00475.
- Phương sai có điều kiện (*volatility*) cũng có ý nghĩa thống kê với $p\text{-value} < 0.001$, khẳng định sự tồn tại của hiện tượng không đồng nhất phương sai.
- Chỉ số AIC và BIC thấp (**-29,712.52** và **-29,688.30**), cho thấy mô hình này phù hợp ở mức cơ bản.

2. Mô hình GARCH(1,1):

- Đây là mô hình tốt nhất theo tiêu chí AIC và BIC (**-68,426.99** và **-68,394.70**), phù hợp với tính chất phương sai có điều kiện thường gặp trong dữ liệu tài chính.
- Các hệ số ω , α_1 , và β_1 đều có ý nghĩa thống kê ($p\text{-value} < 0.05$), khẳng định tính bền vững của sự biến động (*persistent volatility*).

3. Kiểm định phần dư:

- Kiểm định ARCH-LM trên phần dư của mô hình GARCH(1,1) cho kết quả:

$$\text{Statistic} = 82.21, \quad p\text{-value} = 1.85 \times 10^{-13}.$$

Kết quả này cho thấy phần dư vẫn tồn tại không đồng nhất phương sai (*heteroscedasticity*), gợi ý rằng mô hình hiện tại có thể chưa đủ phức tạp để mô tả dữ liệu hoặc cần điều chỉnh tham số.

Kết luận:

- Mô hình GARCH(1,1) phù hợp nhất trong các thử nghiệm ban đầu, nhưng vẫn cần cải thiện để

xử lý triệt để hiện tượng không đồng nhất phương sai.

- Có thể cân nhắc thử nghiệm các mô hình phức tạp hơn, chẳng hạn như EGARCH hoặc TGARCH, để cải thiện khả năng mô hình hóa phương sai có điều kiện.

4.10 Thiết lập mô hình LSTM

Mô hình LSTM (Long Short-Term Memory) được sử dụng để học các đặc trưng phi tuyến và mối quan hệ dài hạn trong dữ liệu chuỗi thời gian. Mô hình này đặc biệt phù hợp cho các bài toán dự đoán với dữ liệu phụ thuộc vào lịch sử.

Cấu trúc mô hình

Mô hình được huấn luyện với dữ liệu lịch sử 60 ngày, sử dụng để dự đoán giá trị trong tương lai. Dưới đây là kiến trúc mô hình:

1. Lớp LSTM đầu tiên:

- Gồm 50 đơn vị (units).
- Được thiết kế để trả về toàn bộ chuỗi trạng thái ẩn qua tham số 'return sequences=True'.
- Nhận đầu vào với kích thước tương ứng '(60, 1)', nơi 60 là số ngày trong chuỗi đầu vào và 1 là số đặc trưng.

2. Lớp LSTM thứ hai:

- Cũng gồm 50 đơn vị.
- Trả về giá trị cuối cùng của chuỗi trạng thái ẩn thông qua tham số 'return sequences=False', làm giảm chiều dữ liệu trước khi chuyển đến các lớp phía sau.

3. Lớp dày đặc (Dense):

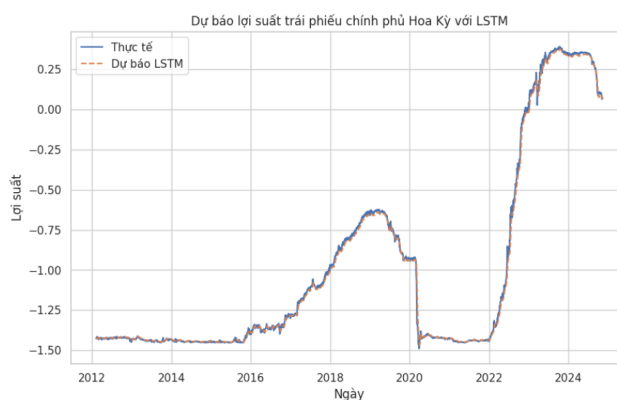
- Một lớp đầu ra tuyến tính có 1 đơn vị (đại diện cho giá trị dự đoán cuối cùng).

Quá trình huấn luyện Mô hình được huấn luyện trong 5 epoch trên tập huấn luyện với hàm mất mát Mean Squared Error (MSE). Bộ tối ưu hóa Adam được sử dụng để cập nhật trọng số, đảm bảo khả năng hội tụ nhanh và ổn định.

Đánh giá mô hình Kết quả dự đoán của mô hình được đánh giá trên tập kiểm tra. Sai số gốc trung

bình bình phương (Root Mean Squared Error - RMSE) rất thấp, chỉ đạt:

$$\text{RMSE} = 0.0211$$

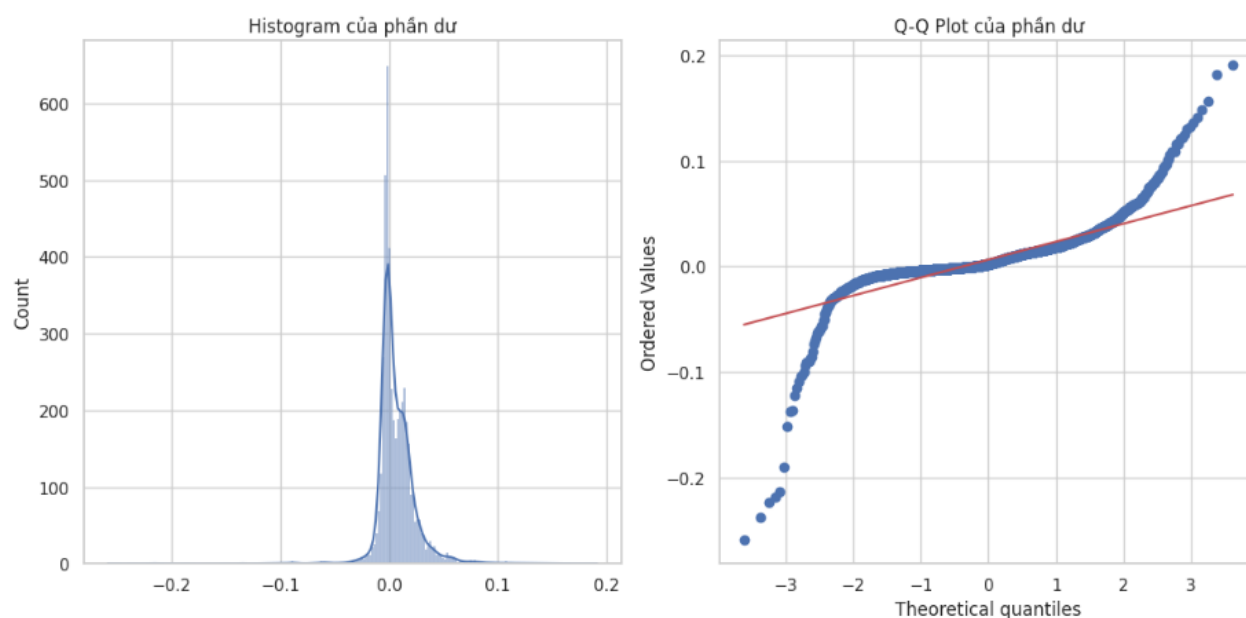


Hình 18: Dự đoán của mô hình LSTM trên tập kiểm tra

Đánh giá phần dư Các biểu đồ sau đây minh họa sự phân phối và tính chất tự tương quan của phần dư:

1. Biểu đồ tần suất và biểu đồ QQ:

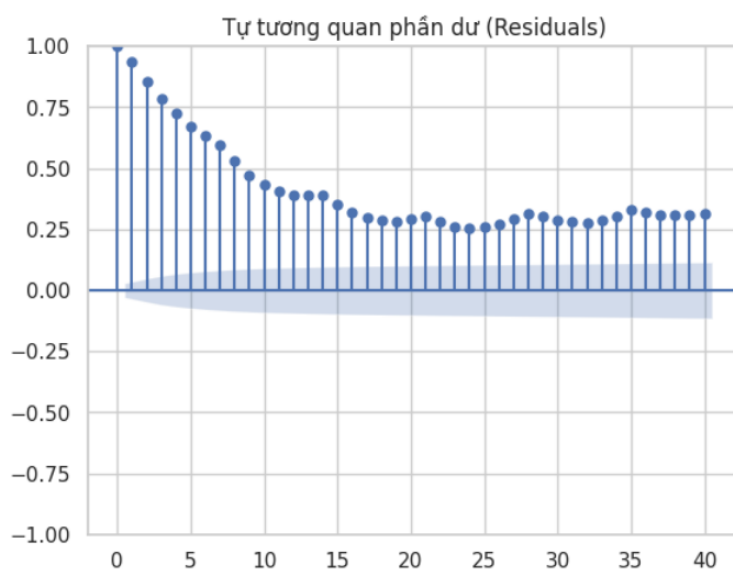
- Phần dư có phân phối khá gần chuẩn, nhưng vẫn tồn tại một số giá trị ngoại lệ.



Hình 19: Đánh giá phân phối phần dư qua biểu đồ tần suất và QQ-plot

2. Biểu đồ tự tương quan (ACF):

- Cho thấy phần dư của mô hình có hiện tượng tự tương quan ở một số độ trễ. Điều này gợi ý rằng mô hình chưa hoàn toàn loại bỏ được toàn bộ thông tin có thể dự đoán trong chuỗi thời gian.



Hình 20: Đánh giá tự tương quan phần dư qua ACF

Nhận xét Mặc dù mô hình LSTM đạt được RMSE thấp và có khả năng dự đoán tốt trên tập kiểm

tra, hiện tượng tự tương quan trong phần dư vẫn tồn tại. Điều này chỉ ra rằng:

- Mô hình có thể chưa khai thác triệt để một số thông tin trong dữ liệu.
- Cần xem xét nhiều hướng tiếp cận khác với sự phức tạp của chuỗi thời gian dữ liệu này.

5 Kết luận

Từ quá trình nghiên cứu và thử nghiệm nhiều mô hình, có thể thấy rằng việc dự đoán lãi suất – một yếu tố kinh tế vĩ mô quan trọng – là một nhiệm vụ đầy thách thức, đặc biệt khi phạm vi dữ liệu kéo dài tới 80 năm. Kết quả cho thấy, lãi suất không thể được dự đoán chính xác chỉ dựa vào chính chuỗi thời gian của nó, đặc biệt trong bối cảnh dài hạn.

Mặc dù một số nghiên cứu trước đây đã sử dụng các mô hình như ARCH để mô hình hóa lãi suất, phạm vi thời gian của các nghiên cứu này thường bị giới hạn trong khoảng ngắn (1 đến 5 năm). Điều này làm giảm tính tổng quát và khả năng áp dụng của các mô hình này khi xét đến các chuỗi thời gian dài hạn.

Kết quả nghiên cứu này gợi ý rằng, bên cạnh việc dựa vào chính chuỗi thời gian của lãi suất, cần tích hợp thêm các yếu tố kinh tế vĩ mô khác như:

- Tỷ lệ tăng trưởng GDP: Để phản ánh sự phát triển và biến động kinh tế tổng thể.
- Tốc độ tăng trưởng dân số: Để nắm bắt các xu hướng về nhu cầu tài chính và tiêu dùng.
- Lạm phát và các chỉ số giá cả: Để hiểu rõ hơn về các áp lực lên lãi suất.
- Chính sách tài khóa và tiền tệ: Để xem xét tác động từ các yếu tố điều hành kinh tế.

Chỉ khi có sự tích hợp đa yếu tố như vậy, mô hình dự báo mới có thể đạt được độ tin cậy cao hơn, đặc biệt trong việc phân tích và dự đoán dài hạn với dữ liệu gần một thế kỷ. Điều này nhấn mạnh tầm quan trọng của việc kết hợp dữ liệu đa chiều và sử dụng các phương pháp tiên tiến, như các mô hình kết hợp (ensemble models) hoặc mô hình học sâu (deep learning), để đối phó với sự phức tạp của các yếu tố kinh tế vĩ mô.

Tài liệu

Toàn bộ kết quả phân tích và nhận định trong nghiên cứu này đều được rút ra trực tiếp từ dữ liệu

thực tế. Quá trình nghiên cứu có tham khảo từ các tài liệu giáo trình, bài tập thực hành và một số công trình nghiên cứu liên quan. Tuy nhiên, các nghiên cứu tham khảo này thường giới hạn trong phạm vi dữ liệu nhỏ và chỉ sử dụng một số mô hình cơ bản. Nghiên cứu hiện tại mở rộng cả về quy mô dữ liệu và sự đa dạng của các phương pháp mô hình hóa, nhằm cung cấp cái nhìn toàn diện và chính xác hơn về vấn đề dự báo lãi suất trong dài hạn.