

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ - TÀI CHÍNH THÀNH PHỐ HỒ CHÍ MINH



ĐỒ ÁN HỌC PHẦN
“PHÂN TÍCH CHUỖI THỜI GIAN”

TÊN ĐỀ TÀI:

SỬ DỤNG PHÂN TÍCH CHUỖI THỜI GIAN CHO
VIỆC DỰ BÁO CHỈ SỐ GIÁ VÀNG

Giảng viên hướng dẫn: ThS. Ngô Thuận Dủ

Sinh viên thực hiện :

| | | |
|------------------|-----------|----------|
| Ngô Thị Thuỳ Lam | 215210030 | 21D1DA01 |
| Nguyễn Cao Bằng | 210051639 | 21D1DA01 |

TP. Hồ Chí Minh, thứ sáu, ngày 01 tháng 12, năm 2023

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC KINH TẾ - TÀI CHÍNH THÀNH PHỐ HỒ CHÍ MINH

ĐỒ ÁN HỌC PHẦN
“PHÂN TÍCH CHUỖI THỜI GIAN”

TÊN ĐỀ TÀI:
SỬ DỤNG PHÂN TÍCH CHUỖI THỜI GIAN CHO
VIỆC DỰ BÁO CHỈ SỐ GIÁ VÀNG

Giảng viên hướng dẫn: ThS. Ngô Thuận Dủ

Sinh viên thực hiện :

| | | |
|------------------|-----------|----------|
| Ngô Thị Thuỳ Lam | 215210030 | 21D1DA01 |
| Nguyễn Cao Bằng | 210051639 | 21D1DA01 |

TP. Hồ Chí Minh, thứ sáu, ngày 01 tháng 12, năm 2023

MỤC LỤC

| | |
|---|----|
| MỤC LỤC | 2 |
| DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT | 4 |
| DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ | 5 |
| CHƯƠNG 1: TỔNG QUAN | 7 |
| 1.1. Tóm tắt nghiên cứu..... | 7 |
| 1.2. Đồ án..... | 8 |
| 1.2.1. Nhiệm vụ đồ án..... | 8 |
| 1.2.2. Cấu trúc đồ án..... | 9 |
| CHƯƠNG 2: CƠ SỞ LÝ THUYẾT..... | 11 |
| 2.1. Khái niệm vàng và chỉ số giá vàng..... | 11 |
| 2.1.1. Vàng..... | 11 |
| 2.1.2. Chỉ số giá vàng | 11 |
| 2.2. Vai trò của vàng..... | 11 |
| 2.3. Giới thiệu về dữ liệu | 12 |
| CHƯƠNG 3: PHÂN TÍCH THIẾT KẾ..... | 13 |
| 3.1. Ngôn ngữ lập trình..... | 13 |
| 3.1.1. Ngôn ngữ R | 13 |
| 3.2. Mô hình tự hồi quy (AR), mô hình trung bình trượt (MA), mô hình trung bình trượt kết hợp tự hồi quy (ARMA) và mô hình ARIMA..... | 13 |
| 3.2.1. Mô hình tự hồi quy (AR - p) | 13 |
| 3.2.2. Mô hình trung bình trượt (MA - q)..... | 14 |
| 3.2.3. Quá trình trung bình trượt kết hợp tự hồi quy (ARMA) | 16 |
| 3.2.4. Mô hình ARIMA (p, d, q) | 17 |
| CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM | 19 |

| | |
|---|----|
| 4.1. Bước 1: Làm mịn dữ liệu, kiểm tra tính mùa vụ của dữ liệu và khử mùa | 19 |
| 4.1.1. Làm mịn dữ liệu | 20 |
| 4.1.2. Kiểm tra tính mùa vụ..... | 21 |
| 4.1.3. Khử yếu tố mùa vụ | 23 |
| 4.2. Bước 2: Sai phân, kiểm tra tính dừng, mô hình ACF – PACF và chọn các giá trị P, D, Q (dữ liệu chưa khử mùa)..... | 25 |
| 4.2.1. Sai phân – kiểm tra tính dừng (dữ liệu chưa khử mùa)..... | 25 |
| 4.2.2. Mô hình ACF – PACF (dữ liệu chưa khử mùa)..... | 26 |
| 4.3. Bước 3: Sai phân, kiểm tra tính dừng, mô hình ACF – PACF và chọn các giá trị p, d, q (dữ liệu đã khử mùa)..... | 29 |
| 4.3.1. Sai phân – kiểm tra tính dừng (dữ liệu đã khử mùa)..... | 29 |
| 4.3.2. Mô hình ACF – PACF (dữ liệu đã khử mùa)..... | 30 |
| 4.4. Bước 4: Lựa chọn mô hình ARIMA..... | 32 |
| 4.5. Bước 5: Kiểm định mô hình đã chọn | 35 |
| 4.6. Bước 6: Dự báo sử dụng mô hình ARIMA | 38 |
| CHƯƠNG 5: KẾT LUẬN VÀ KIẾN NGHỊ | 41 |
| TÀI LIỆU THAM KHẢO | 43 |

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

| STT | Ký hiệu/Viết tắt | Viết đầy đủ | Nội dung/Ý nghĩa |
|-----|------------------|---|--|
| 1 | ARIMA | Autoregressive Integrated Moving Average | Trung bình động tích hợp tự hồi quy |
| 2 | COMEX | CME Group's Commodity Exchange | Sàn giao dịch hàng hóa của CME Group |
| 3 | AR | AutoRegressive | Tự động hồi quy |
| 4 | MA | Moving Average | Đường trung bình động |
| 5 | ARMA | AutoRegressive Moving Average | Đường trung bình động tự hồi quy |
| 6 | ARCH | Autoregressive Conditional Heteroskedasticity | Mô hình tự hồi quy với phương sai có điều kiện khác nhau |
| 7 | ACF | AutoCorrelation Function | Hàm tự tương quan |
| 8 | PACF | Partial AutoCorrelation Function | Hàm tự tương quan từng phần |
| 9 | AIC | Akaike's Information Criterion | Tiêu chí thông tin của Akaike |
| 10 | BIC | Bayesian Information Criterion | Tiêu chí thông tin Bayesian |
| 11 | MSE | Mean Squared Error | Sai số trung bình bình phương |

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

| | |
|---|----|
| Hình 2.3.1: Chỉ số giá vàng các tháng trong năm từ tháng 1/2013 đến tháng 9/2022 | 12 |
| Hình 4.1.1: Dữ liệu chỉ số giá vàng..... | 19 |
| Hình 4.1.2: Mô hình xu thế chỉ số giá vàng | 20 |
| Hình 4.1.3: Mô hình chỉ số vàng sau khi đã làm mịn..... | 21 |
| Hình 4.1.4: Giá trị ước tính mùa vụ của chỉ số vàng..... | 22 |
| Hình 4.1.5: Đồ thị xu hướng ước tính, theo mùa và các thành phần bất thường của chỉ số vàng | 23 |
| Hình 4.1.6: Đồ thị chỉ số vàng đã được điều chỉnh mùa vụ | 24 |
| Hình 4.1.7: Giá trị ước tính của thành phần mùa vụ của chỉ số giá vàng | 24 |
| Hình 4.2.1: Mô hình chỉ số vàng sau khi sai phân lần 1 (chưa khử mùa)..... | 26 |
| Hình 4.2.2: Giá trị tương quan từng phần ở 60 độ trễ của chỉ số vàng | 27 |
| Hình 4.2.3: Mô hình tương quan từng phần ở 60 độ trễ của chỉ số vàng..... | 27 |
| Hình 4.2.4: Giá trị tương quan ở 60 độ trễ của chỉ số vàng | 28 |
| Hình 4.2.5: Mô hình tương quan ở 60 độ trễ của chỉ số vàng | 28 |
| Hình 4.3.1: Mô hình chỉ số vàng sau khi sai phân lần 1 (chưa khử mùa)..... | 29 |
| Hình 4.3.2; Giá trị tương quan từng phần ở 60 độ trễ của chỉ số vàng (đã khử mùa)..... | 30 |
| Hình 4.3.3: Mô hình tương quan từng phần ở 60 độ trễ của chỉ số vàng (đã khử mùa) ... | 31 |
| Hình 4.3.4: Giá trị tương quan ở 60 độ trễ của chỉ vàng (đã khử mùa) | 31 |
| Hình 4.3.5: Mô hình tương quan ở 60 độ trễ của chỉ số vàng (đã khử mùa) | 32 |
| Hình 4.4.1: Kết quả đề cử của auto.arima (chỉ số vàng) | 33 |
| Hình 4.4.2: Giá trị AIC của 4 mô hình thuộc chỉ số vàng..... | 34 |
| Hình 4.4.3: Giá trị BIC của 4 mô hình thuộc chỉ số vàng | 34 |

| | |
|--|----|
| Hình 4.5.1: Mô hình các nghiệm nghịch đảo của chỉ số vàng | 35 |
| Hình 4.5.2: Mô hình kiểm định ARCH của chỉ số giá vàng | 36 |
| Hình 4.5.3: Giá trị chi tiết từ mô hình kiểm định ARCH chỉ số giá vàng..... | 36 |
| Hình 4.5.4: Đồ thị histogram của phần dư dữ liệu chỉ số vàng..... | 37 |
| Hình 4.6.1: Kết quả dự đoán của mô hình m1 thuộc dữ liệu chỉ số vàng | 38 |
| Hình 4.6.2: Đồ thị dự đoán của mô hình m1 thuộc dữ | 38 |
| Hình 4.6.3: Kết quả dự đoán của mô hình auto.arima thuộc dữ liệu chỉ số vàng | 39 |
| Hình 4.6.4: Đồ thị dự đoán của mô hình auto.arima thuộc dữ liệu chỉ số vàng..... | 39 |
| Hình 4.6.5: Giá trị chỉ số vàng thực tế ba tháng cuối năm 2022 | 39 |
| Hình 4.6.6: Giá trị MSE của hai mô hình ARIMA thuộc chỉ số vàng | 39 |

CHƯƠNG 1: TỔNG QUAN

1.1. Tóm tắt nghiên cứu

Việc nghiên cứu và dự đoán chỉ số giá vàng là một lĩnh vực quan trọng trong tài chính và đầu tư. Chỉ số giá vàng thường được sử dụng để đánh giá giá trị tài sản và là một trong những chỉ số quan trọng trong thị trường vàng toàn cầu. Hiểu rõ các yếu tố ảnh hưởng đến giá vàng có thể giúp các nhà đầu tư và các chuyên gia tài chính đưa ra các quyết định thông minh về đầu tư và quản lý rủi ro. Tuy vậy vẫn có rất nhiều các yếu tố ảnh hưởng đến chỉ số giá vàng có thể kể đến như:

- Cung và cầu: Giá vàng phụ thuộc vào sự cân đối giữa cung và cầu trên thị trường. Nếu cung vàng giảm hoặc nhu cầu tăng, giá vàng thường tăng lên và ngược lại.

- Tình hình kinh tế toàn cầu: Chỉ số giá vàng thường ảnh hưởng bởi tình hình kinh tế toàn cầu. Khi kinh tế suy thoái hoặc không ổn định, nhiều nhà đầu tư tìm đến vàng như một cách để bảo vệ giá trị tài sản của họ, dẫn đến tăng giá vàng.

- Chính sách tiền tệ: Chính sách tiền tệ của các ngân hàng trung ương và quốc gia cũng ảnh hưởng đáng kể đến giá vàng. Ví dụ, việc giảm lãi suất có thể làm giảm giá trị đồng tiền và tăng giá vàng.

- Biến động tỷ giá hối đoái: Tỷ giá hối đoái giữa các đồng tiền cũng có thể ảnh hưởng đến giá vàng. Khi đồng tiền mất giá, giá vàng thường tăng.

- Tình hình chính trị và địa phương: Những biến động chính trị, xung đột quốc gia hoặc các sự kiện địa phương có thể tạo ra sự không ổn định và làm tăng giá vàng.

- Sự biến động trên thị trường tài chính: Chỉ số giá vàng cũng có thể chịu ảnh hưởng từ biến động trên thị trường tài chính, bao gồm cổ phiếu, trái phiếu và thị trường hàng hóa khác.

- Trữ lượng vàng của các quốc gia và tổ chức: Sự mua bán và sở hữu vàng của các quốc gia và tổ chức có thể ảnh hưởng đến giá vàng.

Việc dự đoán chỉ số giá vàng là một vấn đề quan trọng trong lĩnh vực tài chính và đầu tư. Giá vàng có sức ảnh hưởng đáng kể đến rất nhiều ngành công nghiệp và thị trường tài chính trên toàn cầu. Do đó, có một sự quan tâm lớn đối với việc dự đoán và dự báo giá vàng để đưa ra quyết định đầu tư và quản lý rủi ro.

Đó là lý do mà chúng tôi đã chọn ***“Sử dụng phân tích chuỗi thời gian cho việc dự báo chỉ số vàng”*** làm đề tài cho môn học này với mong muốn có thể dự báo được chỉ số của giá vàng trong khoảng thời gian vài tháng tới từ các số liệu chi tiết đã có trong 10 năm qua (2013 – 2022) để một phần giúp cho các nhà đầu tư có thể tham khảo và đưa ra được những lựa chọn đúng đắn để đầu tư cho vàng.

1.2. Đồ án

1.2.1. Nhiệm vụ đồ án

1. Tính cấp thiết và lý do hình thành đề tài

Tính cấp thiết: Dự đoán chỉ số giá vàng là một đề tài quan trọng và cấp thiết trong lĩnh vực tài chính và đầu tư. Các nhà đầu tư, các ngân hàng, và các tổ chức tài chính sẽ rất quan tâm đến việc dự đoán trước những sự biến đổi của chỉ số giá vàng để có thể giúp họ đưa ra quyết định đầu tư thông minh và hiệu quả nhất. Mặc dù chắc chắn sẽ có những sai số tương đối nhưng nó cũng sẽ giúp được cho các nhà đầu tư giảm thiểu được các rủi ro không mong muốn.

Lý do hình thành đề tài: Chỉ số giá vàng có những tác động và ảnh hưởng lớn đến nền kinh tế toàn cầu và trong thị trường tài chính. Việc có thể dự đoán chính xác xu hướng và biến động của của chỉ số này có tác dụng lớn trong việc giúp các nhà đầu tư, doanh nghiệp và chính phủ đưa ra những quyết định chiến lược và phòng ngừa các rủi ro lớn liên quan đến kinh tế và tài chính.

2. Ý nghĩa khoa học và thực tiễn

Ý nghĩa khoa học: Nghiên cứu và dự đoán chỉ số giá vàng bằng cách sử dụng mô hình ARIMA có thể mang lại những kiến thức mới về sự biến đổi và tương tác giữa các

chỉ số này với các yếu tố kinh tế, chính trị và xã hội. Nó cung cấp cho chúng ta cái nhìn sâu sắc hơn về các yếu tố đã đang và sẽ ảnh hưởng đến chỉ số giá vàng.

Ý nghĩa thực tiễn: Áp dụng mô hình ARIMA để dự đoán giá vàng mang lại lợi ích thực tiễn cho nhiều nhóm đối tượng khác nhau, bao gồm:

- + Nhà đầu tư: Dự đoán chính xác giúp nhà đầu tư đưa ra quyết định mua bán và định hình chiến lược đầu tư trên các sản chứng khoán cũng như tích trữ tài sản.

- + Ngân hàng và các tổ chức tài chính: Dự đoán giúp họ quản lý rủi ro và đưa ra quyết định về việc bảo vệ tài sản và quản lý vốn.

- + Doanh nghiệp: Dự đoán giúp họ định hình chiến lược về nhập khẩu, xuất khẩu và quản lý giá thành.

- + Chính phủ: Dự đoán giúp chính phủ đưa ra các biện pháp kinh tế và chính sách tiền tệ hợp lý để ổn định thị trường và tăng trưởng kinh tế.

3. Mục tiêu của đề tài

Mục tiêu chính của đề tài là xây dựng và chọn lựa một mô hình ARIMA có khả năng dự đoán chỉ số giá vàng một cách chính xác và tin cậy trong khoảng thời gian ngắn. Mô hình sẽ được sử dụng để dự đoán các xu hướng và những sự biến động trong tương lai, đồng thời cung cấp thông tin hữu ích cho các nhà đầu tư, các ngân hàng, tổ chức tài chính, doanh nghiệp và chính phủ để họ có thể đưa ra quyết định thông minh và hiệu quả trong lĩnh vực tài chính và đầu tư.

1.2.2. Cấu trúc đồ án

✓ CHƯƠNG 1: TỔNG QUAN

Bao gồm tóm tắt nghiên cứu là sự trình bày ngắn gọn về những nghiên cứu đã có và trình bày nên những nhiệm vụ của đồ án và cấu trúc của đồ án một cách ngắn gọn.

✓ CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

Chương tập trung chủ yếu vào việc giải thích các cụm từ, nếu những khái niệm khoa học về những từ ngữ mà chúng ta sẽ nghiên cứu trong đề tài này cũng như giới thiệu sơ lược về bộ dữ liệu được sử dụng trong đề tài.

✓ CHƯƠNG 3: PHÂN TÍCH THIẾT KẾ

Đây là chương sẽ giới thiệu về ngôn ngữ lập trình đã sử dụng, mô hình phân tích chuỗi thời gian sử dụng trong đề tài này và các lý thuyết cơ bản về mô hình đó.

✓ CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM

Chương này sẽ cho thấy từng bước thực hiện mô hình đã lý giải trong chương 3 và khởi chạy chi tiết từng bước rõ ràng, cho thấy kết quả, giải thích và nhận xét chung về những kết quả đó.

✓ CHƯƠNG 5: KẾT LUẬN VÀ KIẾN NGHỊ

Chương cuối cùng trong bài báo cáo sẽ rút ra các kết luận cuối cùng sau khi chạy mô hình và dự đoán ở chương 4. Sau đó sẽ đưa ra những kiến nghị về kết quả dự báo cũng như sẽ kiến nghị về các chiến lược đầu tư khả thi nhất dựa theo kết quả dự đoán.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT

2.1. Khái niệm vàng và chỉ số giá vàng

2.1.1. Vàng

Vàng là nguyên tố hóa học có ký hiệu “Au” và số nguyên tử 79, là một nguyên tố quý, làm cho nó trở thành một trong những nguyên tố có số nguyên tử cao tồn tại ngoài tự nhiên. Vàng là kim loại quý đã được sử dụng làm chất phản xạ neutron trong vũ khí hạt nhân (w71), và để đúc tiền, đồ trang sức và nghệ thuật khác trong suốt lịch sử được ghi lại.

Trước đây, Vàng thường được thực hiện như một chính sách tiền tệ, nhưng tiền vàng đã không còn được coi là một loại tiền tệ lưu hành trong những năm 1930, và tiêu chuẩn vàng thế giới đã bị thay thế bằng một hệ thống tiền tệ định danh sau năm 1971.

2.1.2. Chỉ số giá vàng

Chỉ số giá vàng là chỉ tiêu tương đối (tính bằng %) phản ánh xu hướng và mức độ biến động giá theo thời gian của mặt hàng vàng trên thị trường. Chỉ số này đều cung cấp thông tin quan trọng về giá trị của vàng trên thị trường tài chính.

Chỉ số giá vàng thường được sử dụng để đo lường giá trị của vàng trên thị trường. Nó thường được tính toán dựa trên giá vàng trong một thời gian nhất định và so sánh với một điểm tham chiếu. Chỉ số giá vàng cho phép người ta theo dõi xu hướng chung của giá vàng và phân tích sự biến động của nó trong thời gian.

Chỉ số giá vàng có thể bao gồm nhiều phương pháp tính toán khác nhau. Một trong những phương pháp phổ biến nhất là chỉ số giá vàng COMEX (CME Group's Commodity Exchange). Chỉ số này được tính toán dựa trên giá vàng trên thị trường giao ngay COMEX và được cập nhật liên tục theo giờ.

2.2. Vai trò của vàng

Vàng đã có nhiều vai trò quan trọng trong lịch sử tài chính và tiền tệ. Hiện nay, vai trò chủ yếu của vàng là một tài sản trữ giá và một công cụ đầu tư. Dưới đây là những vai trò cụ thể của vàng:

- Trữ giá: Vàng thường được xem là một cách để bảo vệ giá trị tài sản trong thời gian dài. Trong những thời kỳ không chắc chắn hoặc khủng hoảng kinh tế, người ta thường tìm đến vàng như một giá trị ổn định và an toàn.
- Đầu tư: Vàng cũng là một cách để đầu tư và đa dạng hóa danh mục đầu tư. Giá vàng có thể tăng lên khi có sự không ổn định kinh tế, lạm phát hoặc giảm giá trị đồng tiền.

2.3. Giới thiệu về dữ liệu

Dữ liệu về các chỉ số vàng mà bài nghiên cứu này đã sử dụng được cập nhật từ trang web “<https://www.gso.gov.vn/>”. Đây chính là trang web chính thống của Tổng cục thống kê Việt Nam. Tổng Cục Thống Kê là một cơ quan thuộc Chính phủ Việt Nam, có trách nhiệm chủ đạo về nghiên cứu, thu thập, xử lý và công bố các số liệu thống kê chính thức về kinh tế, dân số và xã hội của đất nước. Nhiệm vụ chính của Tổng Cục là cung cấp các thông tin thống kê đáng tin cậy nhằm trong việc hỗ trợ về các quyết định chính sách, quản lý và nghiên cứu trong nhiều lĩnh vực khác nhau như kinh tế, xã hội, dân số, lao động, môi trường và các chỉ số xã hội khác. Chúng ta có thể tìm thấy được nhiều báo cáo, thống kê, dữ liệu và tài liệu nghiên cứu khác nhau liên quan đến Việt Nam trên trang web này.

Đây là lượng dữ liệu được lấy từ trang web trên và nó chứa dữ liệu về các chỉ số vàng trong khoảng thời gian gần 10 năm (từ 1/2013 đến 6/2022) để sử dụng trong việc dự đoán trong đề tài này:

| Chỉ số chỉ số giá vàng các tháng trong năm so với tháng trước chia theo Nhóm hàng, Năm và Tháng | | | | | | | | | | | | | |
|---|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Năm | Nhóm hàng | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2013 | Chỉ số giá vàng | 98.27 | 99.67 | 97.27 | 97.44 | 95.38 | 95.89 | 93.72 | 100.32 | 101.97 | 97.13 | 98.96 | 96.67 |
| 2014 | Chỉ số giá vàng | 98.18 | 101.87 | 103.31 | 98.96 | 99.15 | 99.88 | 101.38 | 99.66 | 98.34 | 97.18 | 98.51 | 99.95 |
| 2015 | Chỉ số giá vàng | 100.55 | 102.7 | 98.37 | 98.96 | 99.73 | 99.92 | 98.76 | 96.08 | 103.54 | 99.95 | 99.39 | 97.2 |
| 2016 | Chỉ số giá vàng | 99.77 | 103.02 | 104.87 | 100.3 | 101.45 | 99.99 | 105.36 | 101.72 | 99.64 | 98.31 | 98.88 | 97.48 |
| 2017 | Chỉ số giá vàng | 99.82 | 102.45 | 99.72 | 101.01 | 99.9 | 100.25 | 99.12 | 101.11 | 102.61 | 99.12 | 99.7 | 99.88 |
| 2018 | Chỉ số giá vàng | 101.69 | 101.83 | 99.65 | 100.47 | 99.25 | 99.21 | 98.43 | 98.59 | 99.03 | 100.12 | 100.98 | 100.41 |
| 2019 | Chỉ số giá vàng | 102.25 | 101.53 | 99.51 | 99.5 | 99.5 | 101.98 | 104.79 | 104.61 | 103.25 | 99.43 | 99.37 | 99.64 |
| 2020 | Chỉ số giá vàng | 104.37 | 102.74 | 103.87 | 100.69 | 102.41 | 101.71 | 103.49 | 109.86 | 99.67 | 98.9 | 100.87 | 99.17 |
| 2021 | Chỉ số giá vàng | 102.17 | 100.25 | 97.03 | 98.1 | 101.68 | 101.12 | 98.61 | 99.51 | 100 | 99.79 | 102.65 | 100.25 |
| 2022 | Chỉ số giá vàng | 101.08 | 101.85 | 104.51 | 100.73 | 99.48 | 98.86 | 97.61 | 99.1 | 98.29 | | | |

Hình 2.3.1: Chỉ số giá vàng các tháng trong năm từ tháng 1/2013 đến tháng 9/2022

CHƯƠNG 3: PHÂN TÍCH THIẾT KẾ

3.1. Ngôn ngữ lập trình

Trong đồ án lần này, chúng tôi sử dụng “ngôn ngữ R” làm ngôn ngữ lập trình chính trong việc thực hiện các dự đoán về chỉ số vàng trong tương lai gần.

3.1.1. Ngôn ngữ R

Ngôn ngữ R là một ngôn ngữ lập trình và môi trường có năng lực tính toán mạnh mẽ, rất phổ biến và cực kì linh hoạt được sử dụng rộng rãi trong các lĩnh vực thống kê và phân tích dữ liệu. R cung cấp một loạt các gói thư viện, phần mềm phong phú và có chức năng thống kê mạnh mẽ, khiến nó trở thành một trong những công cụ yêu thích của nhiều nhà khoa học dữ liệu và những nhà nghiên cứu.

R có một cú pháp dễ hiểu và tương đối linh hoạt, cho phép người dùng có thể tạo, kiểm tra và triển khai các mô hình phân tích mang khuynh hướng phức tạp. Với cú pháp dễ đọc, R giúp người dùng dễ dàng thực hiện các phân tích thống kê phức tạp và tạo ra những đồ thị để trực quan hóa kết quả.

3.2. Mô hình tự hồi quy (AR), mô hình trung bình trượt (MA), mô hình trung bình trượt kết hợp tự hồi quy (ARMA) và mô hình ARIMA

Trong đồ án, chúng tôi sẽ sử dụng mô hình ARIMA (Autoregressive Integrated Moving Average) làm mô hình chính để dự báo chỉ số vàng trong tương lai. Mô hình ARIMA là một công cụ phân tích chuỗi thời gian rất mạnh mẽ, kết hợp nhiều thành phần quan trọng như mô hình tự hồi quy (AR), mô hình trung bình trượt (MA) và tiêu chuẩn hóa sai số (I).

3.2.1. Mô hình tự hồi quy (AR - p)

Mô hình tự hồi quy (AR), còn được gọi là mô hình tự hồi quy tự động (AutoRegressive), là một mô hình thống kê được sử dụng để mô hình hóa và dự đoán các chuỗi dữ liệu thay đổi theo thời gian. Mô hình tự hồi quy là một dạng đặc biệt của mô hình

dự báo chuỗi thời gian, trong đó giá trị của biến phụ thuộc tại một thời điểm được dự đoán dựa trên các giá trị của biến phụ thuộc trong quá khứ.

Với mô hình tự hồi quy, mục tiêu chính là tìm ra mối quan hệ tự hồi quy (autoregressive relationship) giữa các giá trị trong chuỗi dữ liệu. Điều này có nghĩa là giá trị của biến phụ thuộc tại một thời điểm hiện tại được ước lượng thông qua việc kết hợp các giá trị của biến phụ thuộc trong quá khứ.

Mô hình tự hồi quy có thể được sử dụng để dự đoán các chuỗi dữ liệu thay đổi theo thời gian trong nhiều lĩnh vực, bao gồm tài chính, kinh tế, khoa học xã hội và nhiều lĩnh vực khác. Nó có thể cung cấp thông tin quan trọng về xu hướng và mô hình hoá các quy luật tự nhiên của dữ liệu thay đổi theo thời gian.

Dạng tổng quát của mô hình AR(p):

$$X_t = c + \sum(\varphi_i * X_{t-i}) + \varepsilon_t$$

Trong đó:

X_t : giá trị biến phụ thuộc tại thời điểm t.

c : hằng số.

φ_i : hệ số AR, đại diện cho sự tương quan giữa giá trị quá khứ - hiện tại.

X_{t-i} : các giá trị quá khứ của biến phụ thuộc.

ε_t : nhiễu ngẫu nhiên (sai số), đại diện cho các yếu tố không thể dự đoán.

Trong mô hình AR(p), p là một số nguyên dương cho biết số lượng giá trị quá khứ được sử dụng để dự đoán giá trị hiện tại.

3.2.2. Mô hình trung bình trượt (MA - q)

Mô hình trung bình trượt (MA), còn được gọi là mô hình trung bình động (moving average model), là một mô hình thống kê được sử dụng để mô hình hóa và dự đoán các chuỗi dữ liệu thay đổi theo thời gian. Mô hình MA dựa trên ý tưởng rằng giá trị của biến

phụ thuộc tại một thời điểm được ước lượng dựa trên một tổng trọng số của các thành phần sai số ngẫu nhiên trong quá khứ.

Với mô hình MA, mục tiêu chính là tìm ra mối quan hệ trung bình trượt giữa các thành phần sai số ngẫu nhiên trong chuỗi dữ liệu. Điều này có nghĩa là giá trị của biến phụ thuộc tại một thời điểm hiện tại được ước lượng thông qua việc kết hợp các thành phần sai số ngẫu nhiên trong quá khứ.

Mô hình trung bình trượt có thể được sử dụng để dự đoán và mô hình hóa các chuỗi dữ liệu thay đổi theo thời gian trong nhiều lĩnh vực, bao gồm tài chính, kinh tế, khoa học xã hội và nhiều lĩnh vực khác. Nó có thể cung cấp thông tin về sự biến động ngẫu nhiên và xu hướng trung bình của chuỗi dữ liệu.

Dạng tổng quát của mô hình MA(q):

$$X_t = \mu + \varepsilon_t + \sum(\theta_i * \varepsilon_{t-i})$$

Trong đó:

X_t : giá trị của biến phụ thuộc tại thời điểm t.

μ : giá trị trung bình của biến phụ thuộc.

ε_t : nhiễu ngẫu nhiên (sai số) tại thời điểm t, đại diện cho yếu tố không thể dự đoán và không được giải thích bởi mô hình MA.

θ_i : hệ số MA, đại diện cho sự tương quan giữa các giá trị nhiễu ngẫu nhiên quá khứ và giá trị hiện tại.

ε_{t-i} : giá trị nhiễu ngẫu nhiên trong quá khứ.

Trong mô hình MA(q), q là một số nguyên dương cho biết số lượng giá trị nhiễu ngẫu nhiên quá khứ được sử dụng để dự đoán giá trị hiện tại.

3.2.3. Quá trình trung bình trượt kết hợp tự hồi quy (ARMA)

Mô hình ARMA là một mô hình thống kê sử dụng để mô phỏng và dự đoán dữ liệu chuỗi thời gian. Nó kết hợp hai thành phần chính: quá trình trung bình trượt tự hồi quy (AR) và quá trình trung bình trượt diễn biến (MA).

+ Quá trình tự hồi quy (AR): Trong thành phần AR của mô hình, giá trị của chuỗi thời gian tại thời điểm hiện tại được dự đoán dựa trên các giá trị trước đó của chuỗi thời gian. Cụ thể, mô hình AR sử dụng các hệ số tự hồi quy để xác định mối quan hệ giữa giá trị hiện tại và các giá trị trước đó. Số lượng giá trị trước đó được sử dụng để dự đoán giá trị hiện tại được xác định bởi một tham số là "p".

+ Quá trình trung bình trượt (MA): Trong thành phần MA của mô hình, giá trị của chuỗi thời gian tại thời điểm hiện tại được dự đoán dựa trên các giá trị lỗi trước đó của chuỗi thời gian. Điều này có nghĩa là mô hình MA sử dụng các hệ số diễn biến để xác định mối quan hệ giữa giá trị hiện tại và các giá trị lỗi trước đó. Số lượng giá trị lỗi trước đó được sử dụng để dự đoán giá trị hiện tại được xác định bởi một tham số là "q".

Kết hợp cả AR và MA, mô hình ARMA có thể mô phỏng và dự đoán chuỗi thời gian. Mô hình ARMA được xác định bởi hai tham số là "p" và "q", tương ứng với số lượng giá trị trước đó và số lượng giá trị lỗi trước đó được sử dụng trong mô hình.

Dạng tổng quát của một mô hình ARMA(p, q) là:

$$X_t = c + \sum(\phi_i * X_{t-i}) + \varepsilon_t + \sum(\theta_j * \varepsilon_{t-j})$$

Trong đó:

X_t : giá trị biến phụ thuộc tại thời điểm t.

c : hằng số.

ϕ_i : hệ số AR, đại diện cho sự tương quan giữa những giá trị quá khứ và giá trị hiện tại.

X_{t-i} : giá trị quá khứ của biến phụ thuộc.

ε_t : nhiễu ngẫu nhiên (sai số) tại thời điểm t , đại diện cho yếu tố không thể dự đoán.

θ_j : hệ số MA, đại diện cho tương quan giữa các giá trị nhiễu ngẫu nhiên quá khứ và giá trị hiện tại.

$\varepsilon_{\{t-j\}}$: các giá trị nhiễu ngẫu nhiên trong quá khứ.

Trong mô hình ARMA(p, q), p là số lượng hệ số AR và q là số lượng hệ số MA. Các hệ số φ_i và θ_j sẽ được ước lượng để phù hợp với mô hình và dự đoán giá trị của chuỗi thời gian.

3.2.4. Mô hình ARIMA (p, d, q)

Mô hình ARIMA, còn được gọi là mô hình tự hồi quy trung bình trượt tự động tích hợp (AutoRegressive Integrated Moving Average), là một mô hình thống kê phổ biến được sử dụng để mô hình hóa và dự đoán các chuỗi dữ liệu thay đổi theo thời gian. Khác với mô hình trung bình trượt tự hồi quy ARMA kết hợp các thành phần tự hồi quy (AR), thành phần trung bình trượt (MA), thì ARIMA còn kết hợp thêm cả một thành phần đặc biệt gọi là thành phần tích hợp (I) để mô phỏng các mô hình mang tính chất không tuần hoàn và không tuyến tính.

Mô hình ARIMA thường được sử dụng để mô hình hóa các chuỗi dữ liệu không tuần hoàn và có xu hướng trong việc dự đoán giá trị tương lai. Mô hình này phụ thuộc vào ba thông số chính: p , d và q .

- + Tham số p đại diện cho số lượng thành phần tự hồi quy trong mô hình (AR).
- + Tham số d đại diện cho số lần tích hợp để biến đổi chuỗi dữ liệu thành một dạng tuần hoàn (số lần sai phân).
- + Tham số q đại diện cho số lượng thành phần trung bình trượt trong mô hình (MA).

Sau khi các tham số đã được xác định thì mô hình ARIMA có thể được sử dụng để dự đoán giá trị tương lai của chuỗi dữ liệu dựa trên thông tin trong quá khứ.

Mô hình ARIMA có thể được áp dụng trong nhiều lĩnh vực, bao gồm tài chính, kinh tế, thống kê, dự báo thời tiết và nhiều lĩnh vực khác. Điều quan trọng là ARIMA có khả năng mô hình hóa và dự đoán các chuỗi dữ liệu không tuần hoàn, có xu hướng và có thành phần ngẫu nhiên.

Dạng tổng quát của mô hình ARIMA(p, d, q):

$$(1 - \Phi_1 * B - \Phi_2 * B^2 - \dots - \Phi_p * B^p) * (1 - B)^d * X_t = c + (1 + \theta_1 * B + \theta_2 * B^2 + \dots + \theta_q * B^q) * \varepsilon_t$$

Trong đó:

X_t : giá trị chuỗi thời gian tại thời điểm t.

c : hằng số.

B : toán tử trễ (lag operator), B^d biểu thị việc tích phân (sai phân) “d” lần.

Φ_i : hệ số AR, đại diện cho tương quan giữa các giá trị quá khứ và giá trị hiện tại sau khi đã loại bỏ xu hướng.

θ_i : hệ số MA, đại diện cho tương quan giữa các giá trị nhiễu ngẫu nhiên quá khứ và giá trị hiện tại sau khi đã loại bỏ xu hướng.

ε_t : nhiễu ngẫu nhiên (sai số).

Trong mô hình ARIMA(p, d, q), p là số lượng hệ số AR, d là mức độ tích phân (sai phân) và q là số lượng hệ số MA. Các hệ số “ Φ_i ” và “ θ_i ” được ước lượng để phù hợp với mô hình và dự đoán giá trị của chuỗi thời gian sau khi loại bỏ xu hướng.

CHƯƠNG 4: KẾT QUẢ THỰC NGHIỆM

4.1. Bước 1: Làm mịn dữ liệu, kiểm tra tính mùa vụ của dữ liệu và khử mùa

Đầu tiên, chúng ta gán bộ dữ liệu về chỉ số giá vàng vào biến “gold” .

```
gold <- read_excel('/content/Chi_so_gia_vang(2013-9.2022).xlsx', col_names = TRUE)
```

Sau đó chuyển hóa dữ liệu đã đưa vào trong biến thành dữ liệu “time series” và gán nó vào biến mới là “gold1”:

```
gold1 <- ts(gold$value, frequency = 12, start = c(2013, 1))
```

Ở đây gọi biến đã được gán dữ liệu chuyển hóa “time series” ra và ta sẽ quan sát được bộ dữ liệu đã rất gọn gàng:

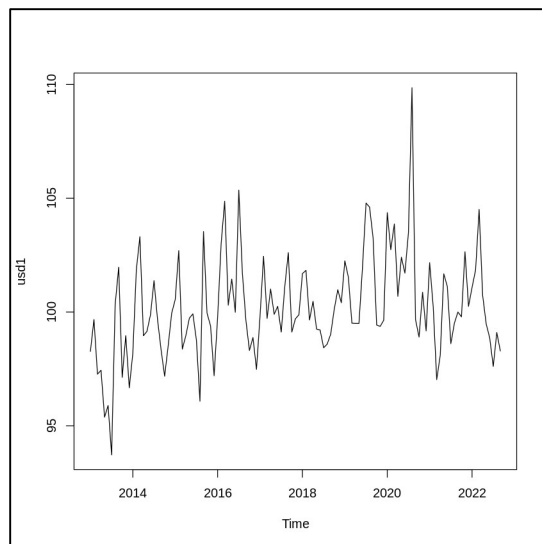
gold1

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2013 | 98.27 | 99.67 | 97.27 | 97.44 | 95.38 | 95.89 | 93.72 | 100.32 | 101.97 | 97.13 | 98.96 | 96.67 |
| 2014 | 98.18 | 101.87 | 103.31 | 98.96 | 99.15 | 99.88 | 101.38 | 99.66 | 98.34 | 97.18 | 98.51 | 99.95 |
| 2015 | 100.55 | 102.70 | 98.37 | 98.96 | 99.73 | 99.92 | 98.76 | 96.08 | 103.54 | 99.95 | 99.39 | 97.20 |
| 2016 | 99.77 | 103.02 | 104.87 | 100.30 | 101.45 | 99.99 | 105.36 | 101.72 | 99.64 | 98.31 | 98.88 | 97.48 |
| 2017 | 99.82 | 102.45 | 99.72 | 101.01 | 99.90 | 100.25 | 99.12 | 101.11 | 102.61 | 99.12 | 99.70 | 99.88 |
| 2018 | 101.69 | 101.83 | 99.65 | 100.47 | 99.25 | 99.21 | 98.43 | 98.59 | 99.03 | 100.12 | 100.98 | 100.41 |
| 2019 | 102.25 | 101.53 | 99.51 | 99.50 | 99.50 | 101.98 | 104.79 | 104.61 | 103.25 | 99.43 | 99.37 | 99.64 |
| 2020 | 104.37 | 102.74 | 103.87 | 100.69 | 102.41 | 101.71 | 103.49 | 109.86 | 99.67 | 98.90 | 100.87 | 99.17 |
| 2021 | 102.17 | 100.25 | 97.03 | 98.10 | 101.68 | 101.12 | 98.61 | 99.51 | 100.00 | 99.79 | 102.65 | 100.25 |
| 2022 | 101.08 | 101.85 | 104.51 | 100.73 | 99.48 | 98.86 | 97.61 | 99.10 | 98.29 | | | |

Hình 4.1.1: Dữ liệu chỉ số giá vàng

Tiếp đến, chúng ta sẽ vẽ dữ liệu đồ thị bằng hàm “plot()” để có thể quan sát được xu thế của dữ liệu:

```
plot(gold1)
```



Hình 4.1.2: Mô hình xu thế chỉ số giá vàng

Mô hình cho thấy sự biến động liên tục của các chỉ số, tuy nhiên nếu sử dụng xu thế này sẽ rất khó để có thể nhìn thấy rõ ràng những biến động. Do đó chúng ta sẽ tiến hành một bước trung gian để có thể dễ dàng quan sát dữ liệu hơn, đó chính là làm mịn dữ liệu.

4.1.1. Làm mịn dữ liệu

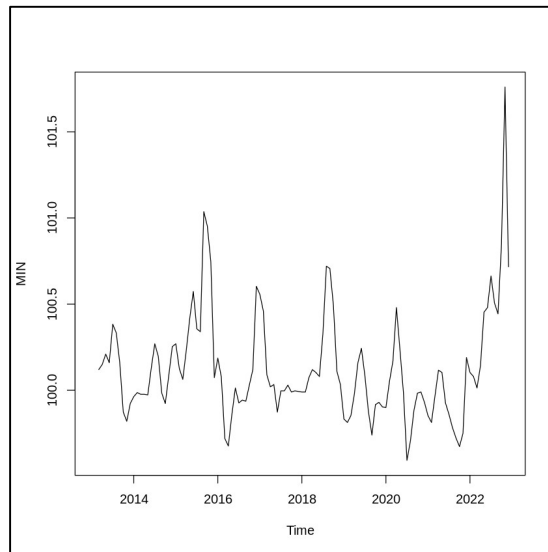
Phân tách chuỗi thời gian sẽ liên quan đến việc cố gắng phân tách chuỗi thời gian ra thành các thành phần, nghĩa là ước tính thành phần xu hướng và thành phần bất thường. Để ước tính thành phần xu hướng của chuỗi thời gian, ta thường dùng phương pháp làm mịn (smoothing method), ví dụ như là tính toán trung bình động đơn giản (simple moving average) của chuỗi.

Để thuận lợi hơn trong việc phân tách, ta sẽ sử dụng hàm “**SMA()**” trong gói thư viện “**TTR**”, hàm này có chức năng làm mượt chuỗi bằng cách sử dụng kỹ thuật trung bình trượt đơn giản. Để có thể sử dụng hàm “**SMA()**” để làm mịn dữ liệu chuỗi thời gian. Đầu tiên là chỉ định thứ tự (span) của đường trung bình động đơn giản, sử dụng tham số “**n**”.

Trở lại với đề tài của chúng ta chứa dữ liệu về chỉ số giá vàng biến động trong 10 năm. Ta có thể ước tính thành phần xu hướng của chuỗi bằng phương pháp làm mịn chuỗi với kỹ thuật trung bình trượt đơn giản. Thiết lập chỉ số “**n = 3**” để tính toán trung bình 3 kỳ và hiển thị chuỗi đã được làm mịn như sau:

```
MINgold <- SMA(gold1, n=3)
```

Sau khi làm mịn:



Hình 4.1.3: Mô hình chỉ số vàng sau khi đã làm mịn

Nhìn vào biểu đồ, ta có thể nhận ra được chu kì tăng giảm của chỉ số vàng khá rõ rệt, ví dụ như các chỉ số thường có xu hướng tăng vào cuối và đầu các năm trong khi đó chúng lại giảm và giữ mức trung bình ở các tháng giữa các năm.

=> Do đó chúng ta có thể phần nào xác định được trong dữ liệu của chúng ta có yếu tố mùa vụ. Và để có thể xác định chính xác rằng liệu bản thân của dữ liệu có bị ảnh hưởng bởi yếu tố mùa vụ không thì ta sẽ tiến hành kiểm tra và phân rã dữ liệu.

4.1.2. Kiểm tra tính mùa vụ

Đối với một chuỗi thời gian có yếu tố mùa vụ (**seasonal time series**) sẽ bao gồm 3 thành phần chính: thành phần xu hướng, thành phần theo mùa và thành phần bất thường. Vì vậy, để có thể phân tách một chuỗi thời có yếu tố mùa vụ, chúng ta sẽ tách chuỗi thời gian thành ba thành phần trên.

Để dễ dàng hơn trong việc ước tính các thành phần xu hướng và thành phần theo mùa của chuỗi thời gian có yếu tố mùa vụ dạng mô hình cộng, chúng ta có thể sử dụng hàm **“decompose()”** trong R. Hàm này có chức năng ước tính thành phần xu hướng, thành phần mùa vụ và thành phần bất thường của chuỗi.

Kết quả trả về của hàm “**decompose()**” sẽ trả về cho chúng ta một đối tượng danh sách với kết quả ước tính của các thành phần với các tên gọi tương ứng của chúng là “**Seasonal**”, “**Trend**” và “**Random**”.

Như với dữ liệu chuỗi thời gian của số của đề tài về chỉ số vàng, mùa vụ trong chuỗi dữ liệu biến động cao về cuối năm và đầu các năm.

Sử dụng hàm “**decompose()**” để phân rã các chuỗi thời gian có yếu tố mùa vụ và gán nó vào biến “**MUA**”.

```
MUAgold <- decompose(gold1)
```

Các giá trị ước tính của thành phần mùa vụ, xu hướng và bất thường được lưu trong các biến như “**MUA\$seasonal**”, “**MUA\$trend**” và “**MUA\$random**”.

Vì chúng ta đang kiểm tra về tính mùa vụ nên sẽ in ra các giá trị ước tính của thành phần mùa vụ:

```
MUAgold$seasonal
```

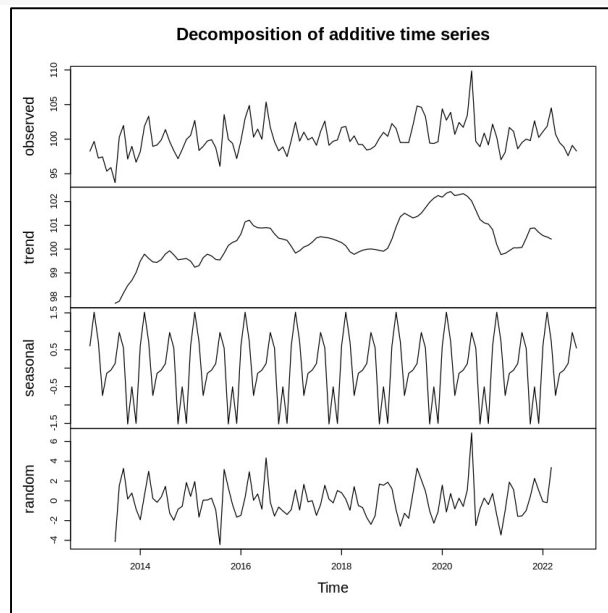
| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|------------|------------|------------|-------------|-------------|-------------|------------|------------|------------|-------------|-------------|-------------|
| 2013 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2014 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2015 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2016 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2017 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2018 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2019 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2020 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2021 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | -1.51729408 | -0.50928482 | -1.50423852 |
| 2022 | 0.60066889 | 1.51719666 | 0.71765963 | -0.74341676 | -0.14148968 | -0.05560426 | 0.12372444 | 0.96728926 | 0.54478926 | | | |

Hình 4.1.4: Giá trị ước tính mùa vụ của chỉ số vàng

Các yếu tố mùa vụ trên cho thấy được sự ước tính từ tháng 1 đến tháng 12 và cũng giống nhau cho mỗi năm. Tác động của yếu tố mùa vụ lớn nhất là vào tháng 2 (khoảng 1,517) và thấp nhất là vào tháng 10 (khoảng -1,517). Điều này cho thấy dường như chỉ số giá vàng có xu thế tăng cao nhất vào tháng 2 và giảm sâu nhất về chỉ số vào tháng 10 hàng năm.

Chúng ta có thể vẽ đồ thị xu hướng ước tính, theo mùa và các thành phần bất thường của chuỗi thời gian bằng cách sử dụng hàm “**plot()**” như sau:

```
plot(MUAgold
```



Hình 4.1.5: Đồ thị xu hướng ước tính, theo mùa và các thành phần bất thường của chỉ số vàng

Biểu đồ trên hiển thị:

- + Chuỗi thời gian gốc (trên cùng).
- + Thành phần xu hướng ước tính (thứ hai từ trên xuống).
- + Thành phần mùa vụ ước tính (thứ ba từ trên xuống).
- + Thành phần bất thường ước tính (dưới cùng).

Chúng ta thấy rằng thành phần xu hướng ước tính của chỉ số có xu thế tăng đều qua các năm 2013 và lên cao nhất vào đầu năm 2020 nhưng đến giữa năm 2020 lại có xu thế giảm đến cuối năm 2021 và giữ mức ổn định ở năm 2022.

4.1.3. Khử yếu tố mùa vụ

Để khử mùa cũng như là điều chỉnh chuỗi thời gian theo mùa này bằng cách ước tính thành phần mùa vụ và trừ đi thành phần này ước tính này khỏi chuỗi ban đầu.

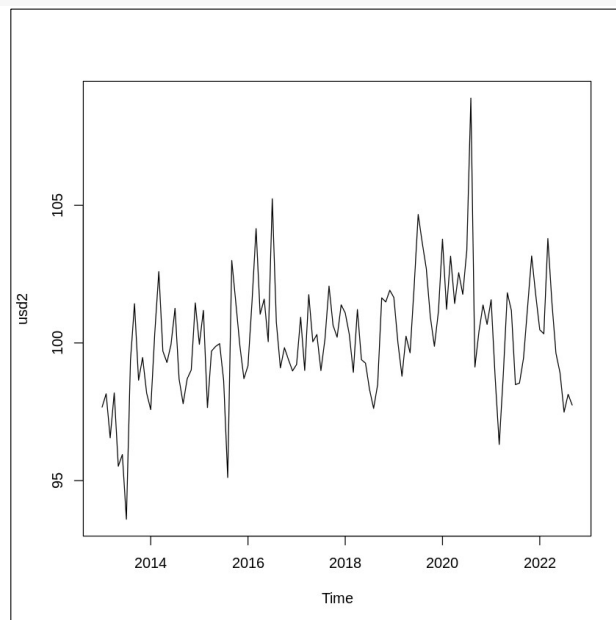
Theo dữ liệu trên đề tài, ta sẽ điều chỉnh theo yếu tố mùa vụ ở chuỗi thời gian của chỉ số vàng mỗi tháng trong 10 năm, chúng ta có thể ước tính thành phần mùa vụ này bằng

cách sử dụng hàm “**decompose()**” như đã sử dụng và gán vào biến “MUA” ở trên, sau đó trừ thành phần ước tính này khỏi chuỗi thời gian ban đầu và gán nó vào biến mới “gold2”:

```
gold2 <- gold1 - MUAgold$seasonal
```

Tiếp tục dùng hàm “**plot()**” hiển thị mô hình dữ liệu và gọi biến “gold2” để hiển thị chuỗi đã điều chỉnh yếu tố mùa vụ:

```
plot(gold2)
gold2
```



Hình 4.1.6: Đồ thị chỉ số vàng đã được điều chỉnh mùa vụ

| | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 2013 | 97.66933 | 98.15280 | 96.55234 | 98.18342 | 95.52149 | 95.94560 | 93.59628 | 99.35271 | 101.42521 | 98.64729 | 99.46928 | 98.17424 |
| 2014 | 97.57933 | 100.35280 | 102.59234 | 99.70342 | 99.29149 | 99.93560 | 101.25628 | 98.69271 | 97.79521 | 98.69729 | 99.01928 | 101.45424 |
| 2015 | 99.94933 | 101.18280 | 97.65234 | 99.70342 | 99.87149 | 99.97560 | 98.63628 | 95.11271 | 102.99521 | 101.46729 | 99.89928 | 98.70424 |
| 2016 | 99.16933 | 101.50280 | 104.15234 | 101.04342 | 101.59149 | 100.04560 | 105.23628 | 100.75271 | 99.09521 | 99.82729 | 99.38928 | 98.98424 |
| 2017 | 99.21933 | 100.93280 | 99.00234 | 101.75342 | 100.04149 | 100.30560 | 98.99628 | 100.14271 | 102.06521 | 100.63729 | 100.20928 | 101.38424 |
| 2018 | 101.08933 | 100.31280 | 98.93234 | 101.21342 | 99.39149 | 99.26560 | 98.30628 | 97.62271 | 98.48521 | 101.63729 | 101.48928 | 101.91424 |
| 2019 | 101.64933 | 100.01280 | 98.79234 | 100.24342 | 99.64149 | 102.03560 | 104.66628 | 103.64271 | 102.70521 | 100.94729 | 99.87928 | 101.14424 |
| 2020 | 103.76933 | 101.22280 | 103.15234 | 101.43342 | 102.55149 | 101.76560 | 103.36628 | 108.89271 | 99.12521 | 100.41729 | 101.37928 | 100.67424 |
| 2021 | 101.56933 | 98.73280 | 96.31234 | 98.84342 | 101.82149 | 101.17560 | 98.48628 | 98.54271 | 99.45521 | 101.30729 | 103.15928 | 101.75424 |
| 2022 | 100.47933 | 100.33280 | 103.79234 | 101.47342 | 99.62149 | 98.91560 | 97.48628 | 98.13271 | 97.74521 | | | |

Hình 4.1.7: Giá trị ước tính của thành phần mùa vụ của chỉ số giá vàng

Chúng ta có thể thấy rằng sự biến động theo mùa đã bị loại bỏ khỏi chuỗi thời gian được điều chỉnh theo mùa. Chuỗi thời gian được điều chỉnh theo mùa như trên hiện chỉ chứa 2 thành phần là thành phần xu hướng và thành phần bất thường.

***Lưu ý*:**

Đối với các bộ dữ liệu mang yếu tố mùa vụ như bộ dữ liệu về chỉ số vàng này thì mô hình ARIMA sẽ có dạng $ARIMA(p,d,q)(P,D,Q)[S]$, trong đó:

- + (p,d,q) là ba giá trị ta sẽ xác định được thông qua phân tích bộ dữ liệu đã khử mùa.
- + (P,D,Q) là ba giá trị ta sẽ xác định thông qua phân tích bộ dữ liệu gốc.
- + $[S]$ là biến thể hiện chu kì thời gian theo bộ dữ liệu.

4.2. Bước 2: Sai phân, kiểm tra tính dừng, mô hình ACF – PACF và chọn các giá trị P, D, Q (dữ liệu chưa khử mùa)

4.2.1. Sai phân – kiểm tra tính dừng (dữ liệu chưa khử mùa)

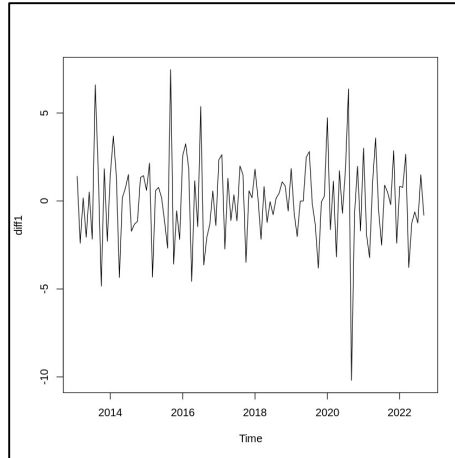
Mô hình ARIMA được xác định cho các chuỗi thời gian dừng, do đó, nếu bắt đầu với một chuỗi không dừng (non-stationary) thì trước tiên cần phải lấy sai phân chuỗi cho đến khi chuỗi dừng. Nếu phải lấy sai phân chuỗi “D (d)” lần để có được chuỗi dừng, thì mô hình ARIMA sẽ có dạng tổng quát là $ARIMA(p,d,q)(P,D,Q)[S]$. Trong đó, D (d) được gọi là bậc tích hợp.

Trên R, Ta có thể sử dụng hàm “**diff()**” để lấy sai phân của một chuỗi.

Trong trường hợp của đề tài này sẽ sai phân 1 lần để xem kết quả.

```
diff1gold <- diff(gold1, differences = 1)
plot.ts(diff1gold)
```

Trong đó “gold1” là biến chứa dữ liệu chưa khử mùa mà ta đã gán ở đầu mục 4.1.



Hình 4.2.1: Mô hình chỉ số vàng sau khi sai phân lần 1 (chưa khử mùa)

Sau khi sai phân lần 1, chúng ta có thể thấy rằng mô hình đã có tính dừng, vì các dữ liệu chỉ biến thiên xoay quanh một giá trị cố định đó là 0.

=> Giá trị **D = 1** cho chuỗi dữ liệu.

4.2.2. Mô hình ACF – PACF (dữ liệu chưa khử mùa)

Nếu chuỗi thời gian là chuỗi dừng (bậc gốc hoặc sai phân), thì bước tiếp theo là chọn mô hình ARIMA thích hợp. Điều này nghĩa là tìm giá trị của các giá trị phù hợp nhất của p và q cho mô hình ARIMA(p,d,q). Ở đề tài hiện tại là sẽ tìm P và Q cho mô hình gốc chưa khử mùa. Thông thường, để làm điều này thì phải xem xét biểu đồ tương quan (correlogram) và biểu đồ tương quan từng phần (partial correlogram) của các chuỗi dừng.

Để có thể vẽ được biểu đồ tương quan và biểu đồ tương quan từng phần, chúng ta lần lượt sử dụng các hàm “**acf()**” và hàm “**pacf()**” trong R.

Trở lại ví dụ về các chỉ số, để vẽ biểu đồ tương quan cho các độ trễ đầu tiên của 2 chuỗi sai phân bậc 1, ta thực hiện như sau:

4.2.2.1 Chuỗi sai phân chỉ số vàng (dữ liệu chưa khử mùa):

+ Mô hình PACF

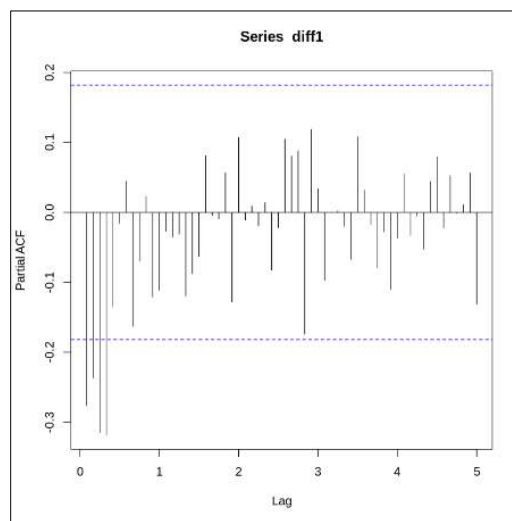
```
pacf(diff1gold, lag.max = 20, na.action = na.pass, plot = FALSE)
pacf(diff1gold, lag.max = 20, na.action = na.pass)
```

```

0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333 0.9167
-0.376 -0.479 -0.141 -0.191 -0.055 -0.086 -0.105 -0.101 -0.024 -0.168 -0.053
1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833 1.6667 1.7500 1.8333
0.025 -0.108 -0.247 0.119 -0.070 -0.036 -0.033 -0.155 0.063 0.021 0.117
1.9167 2.0000 2.0833 2.1667 2.2500 2.3333 2.4167 2.5000 2.5833 2.6667 2.7500
0.039 -0.073 -0.085 -0.027 0.046 0.078 -0.091 -0.008 0.027 0.044 -0.046
2.8333 2.9167 3.0000 3.0833 3.1667 3.2500 3.3333 3.4167 3.5000 3.5833 3.6667
0.007 -0.026 -0.012 0.051 -0.030 -0.045 0.010 0.040 -0.040 -0.117 -0.011
3.7500 3.8333 3.9167 4.0000 4.0833 4.1667 4.2500 4.3333 4.4167 4.5000 4.5833
0.073 0.017 0.048 -0.056 0.014 -0.060 -0.056 -0.032 0.046 -0.071 0.191
4.6667 4.7500 4.8333 4.9167 5.0000
-0.034 0.008 -0.047 -0.102 0.036

```

Hình 4.2.2: Giá trị tương quan từng phần ở 60 độ trễ của chỉ số vàng



Hình 4.2.3: Mô hình tương quan từng phần ở 60 độ trễ của chỉ số vàng

Biểu đồ tương quan từng phần PACF cho thấy có rất nhiều độ trễ vượt quá giới hạn của ý nghĩa thống kê nhưng đa số lại thuộc những độ trễ có giá trị thập phân, sẽ không thể mang tính chất khách quan và không thể sử dụng trong bộ dữ liệu này vì bộ dữ liệu sử dụng đơn vị tháng nên chúng ta sẽ chỉ xem xét những độ trễ mang giá trị nguyên dương. Do đó, khi chúng ta nhìn vào biểu đồ cho thấy ở độ trễ 1 và 5 là hai độ trễ có giá trị tương quan lớn nhất (1: 0,025; 2: - 0,073).

=> Giá trị **P = 1 và 2**.

+ Mô hình ACF

```

acf(diff1gold, lag.max = 20, na.action = na.pass, plot = FALSE)
acf(diff1gold, lag.max = 20, na.action = na.pass)

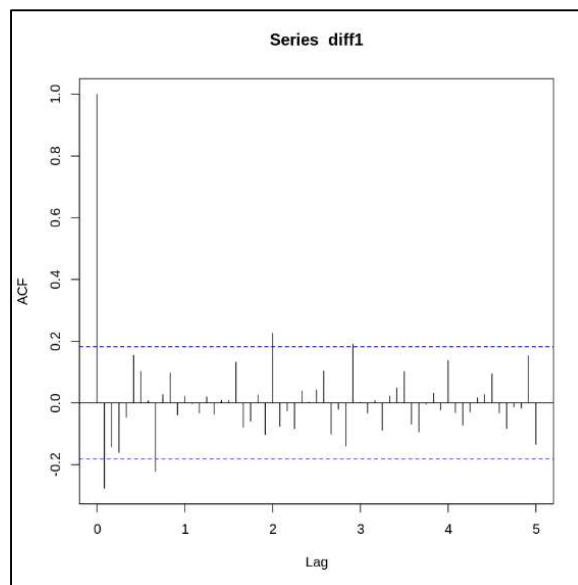
```

```

Autocorrelations of series 'diff1', by lag
0.0000 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333
1.000 -0.276 -0.142 -0.161 -0.047 0.154 0.102 0.006 -0.222 0.027 0.097
0.9167 1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833 1.6667 1.7500
-0.039 0.022 -0.002 -0.033 0.019 -0.037 0.008 0.008 0.132 -0.079 -0.060
1.8333 1.9167 2.0000 2.0833 2.1667 2.2500 2.3333 2.4167 2.5000 2.5833 2.6667
0.026 -0.103 0.226 -0.076 -0.026 -0.083 0.039 0.002 0.042 0.103 -0.101
2.7500 2.8333 2.9167 3.0000 3.0833 3.1667 3.2500 3.3333 3.4167 3.5000 3.5833
-0.021 -0.140 0.189 0.002 -0.033 0.007 -0.089 0.022 0.048 0.101 -0.069
3.6667 3.7500 3.8333 3.9167 4.0000 4.0833 4.1667 4.2500 4.3333 4.4167 4.5000
-0.094 -0.004 0.031 -0.022 0.138 -0.031 -0.073 -0.029 0.017 0.027 0.093
4.5833 4.6667 4.7500 4.8333 4.9167 5.0000
-0.033 -0.083 -0.013 -0.017 0.153 -0.135

```

Hình 4.2.4: Giá trị tương quan ở 60 độ trễ của chỉ số vàng



Hình 4.2.5: Mô hình tương quan ở 60 độ trễ của chỉ số vàng

=> Giá trị **Q = 2**. (vì có giá trị tương quan lớn nhất là 0,226).

+ Lựa chọn mô hình ARIMA

Từ các giá trị đã tìm được **D = 1**; **P = 1 và 2**; **Q = 2**. Chúng ta có thể rút ra được tổng $P*Q = 2$ tổ hợp (P,D,Q) thuộc bộ dữ liệu vàng chưa khử mùa:

(1,1,2);

(2,1,2)

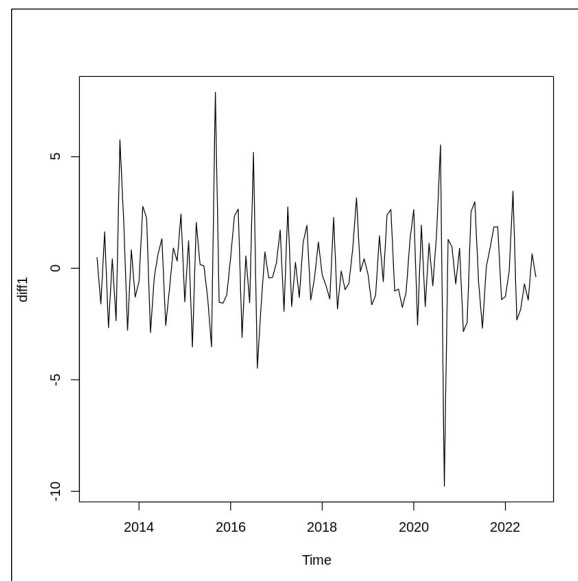
4.3. Bước 3: Sai phân, kiểm tra tính dừng, mô hình ACF – PACF và chọn các giá trị p , d , q (dữ liệu đã khử mùa)

4.3.1. Sai phân – kiểm tra tính dừng (dữ liệu đã khử mùa)

Tương tự với bộ dữ liệu chưa khử mùa, với dữ liệu khử mùa này vẫn sẽ lấy sai phân bằng hàm “**diff()**” để kiểm tra tính dừng của bộ dữ liệu.

```
diff1gold2 <- diff(gold2, differences = 1)
plot.ts(diff1)
```

Trong đoạn code trên, “gold2” là biến chứa dữ liệu đã khử mùa mà chúng ta đã gán trong mục 4.1.3.



Hình 4.3.1: Mô hình chỉ số vàng sau khi sai phân lần 1 (chưa khử mùa)

Sau khi sai phân lần 1, tương tự như bộ dữ liệu chưa khử mùa thì chúng ta có thể thấy rằng mô hình đã có tính dừng, vì các dữ liệu chỉ biến thiên xoay quanh một giá trị cố định đó là 0.

=> Giá trị **$d = 1$** cho chuỗi dữ liệu.

4.3.2. Mô hình ACF – PACF (dữ liệu đã khử mùa)

Nếu chuỗi thời gian đã là chuỗi dừng thì bước tiếp theo là chọn mô hình ARIMA thích hợp. Điều này nghĩa là tìm giá trị của các giá trị phù hợp nhất của p và q cho mô hình ARIMA(p,d,q). Ở mục hiện tại, ta sẽ tìm p và q cho dữ liệu đã khử mùa.

Để có thể vẽ được biểu đồ tương quan và biểu đồ tương quan từng phần, chúng ta lần lượt sử dụng các hàm “**acf()**” và hàm “**pacf()**” trong R.

Trở lại ví dụ về các chỉ số, để vẽ biểu đồ tương quan các độ trễ đầu tiên của chuỗi sai phân bậc 1, ta thực hiện như sau.

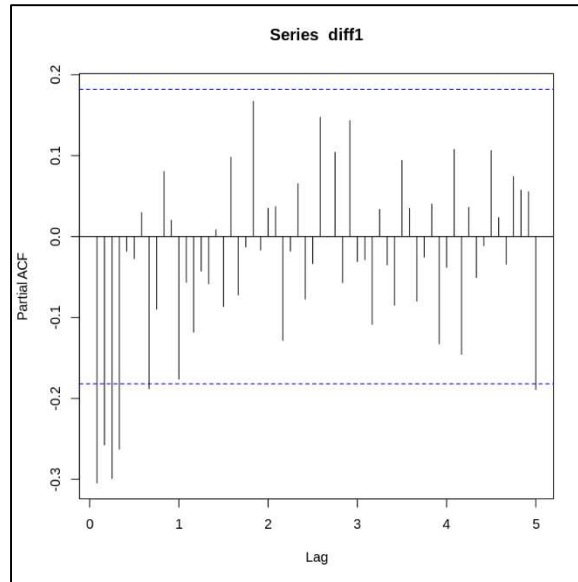
4.3.2.1. Chuỗi sai phân chỉ số vàng (dữ liệu đã khử mùa):

+ Mô hình PACF

```
pacf(diff1, lag.max = 60, na.action = na.pass, plot = FALSE)
pacf(diff1, lag.max = 60, na.action = na.pass)
```

```
Partial autocorrelations of series 'diff1', by lag
0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333 0.9167
-0.305 -0.258 -0.299 -0.263 -0.018 -0.027 0.030 -0.188 -0.090 0.081 0.020
1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833 1.6667 1.7500 1.8333
-0.176 -0.056 -0.118 -0.043 -0.059 0.008 -0.087 0.098 -0.072 -0.013 0.167
1.9167 2.0000 2.0833 2.1667 2.2500 2.3333 2.4167 2.5000 2.5833 2.6667 2.7500
-0.017 0.035 0.037 -0.129 -0.018 0.065 -0.077 -0.033 0.147 0.000 0.104
2.8333 2.9167 3.0000 3.0833 3.1667 3.2500 3.3333 3.4167 3.5000 3.5833 3.6667
-0.057 0.143 -0.031 -0.029 -0.109 0.034 -0.035 -0.085 0.094 0.035 -0.080
3.7500 3.8333 3.9167 4.0000 4.0833 4.1667 4.2500 4.3333 4.4167 4.5000 4.5833
-0.025 0.040 -0.133 -0.038 0.108 -0.146 0.036 -0.051 -0.011 0.106 0.024
4.6667 4.7500 4.8333 4.9167 5.0000
-0.034 0.074 0.057 0.055 -0.189
```

Hình 4.3.2; Giá trị tương quan từng phần ở 60 độ trễ của chỉ số vàng (đã khử mùa)



Hình 4.3.3: Mô hình tương quan từng phần ở 60 độ trễ của chỉ số vàng (đã khử mùa)

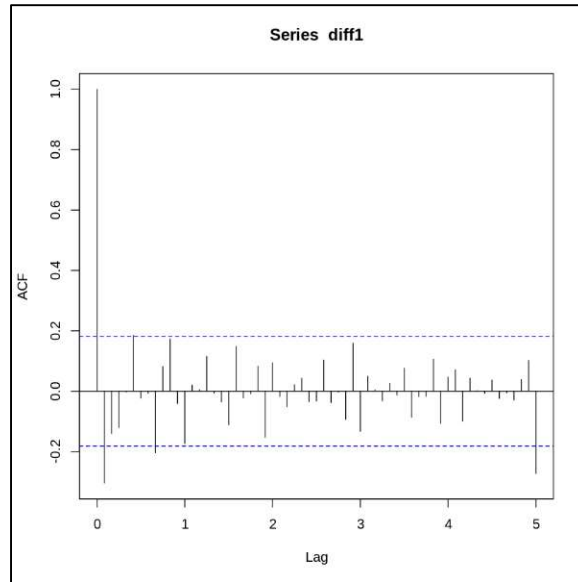
=> Giá trị **p = 1**. (vì có giá trị tương quan lớn nhất là - 0,176).

+ Mô hình ACF

```
acf(diff1, lag.max = 60, na.action = na.pass, plot = FALSE)
acf(diff1, lag.max = 60, na.action = na.pass)
```

```
Autocorrelations of series 'diff1', by lag
0.0000 0.0833 0.1667 0.2500 0.3333 0.4167 0.5000 0.5833 0.6667 0.7500 0.8333
1.000 -0.305 -0.141 -0.121 -0.003 0.185 -0.023 -0.007 -0.204 0.081 0.173
0.9167 1.0000 1.0833 1.1667 1.2500 1.3333 1.4167 1.5000 1.5833 1.6667 1.7500
-0.041 -0.173 0.020 0.004 0.115 -0.006 -0.036 -0.111 0.148 -0.022 -0.009
1.8333 1.9167 2.0000 2.0833 2.1667 2.2500 2.3333 2.4167 2.5000 2.5833 2.6667
0.082 -0.153 0.094 -0.018 -0.052 0.022 0.043 -0.035 -0.032 0.103 -0.038
2.7500 2.8333 2.9167 3.0000 3.0833 3.1667 3.2500 3.3333 3.4167 3.5000 3.5833
-0.003 -0.093 0.159 -0.133 0.049 0.004 -0.032 0.026 -0.013 0.077 -0.086
3.6667 3.7500 3.8333 3.9167 4.0000 4.0833 4.1667 4.2500 4.3333 4.4167 4.5000
-0.018 -0.017 0.106 -0.107 0.046 0.071 -0.100 0.043 0.001 -0.006 0.037
4.5833 4.6667 4.7500 4.8333 4.9167 5.0000
-0.024 -0.005 -0.030 0.039 0.102 -0.273
```

Hình 4.3.4: Giá trị tương quan ở 60 độ trễ của chỉ vàng (đã khử mùa)



Hình 4.3.5: Mô hình tương quan ở 60 độ trễ của chỉ số vàng (đã khử mùa)

=> Giá trị **q = 1 và 5**. (vì có giá trị tương quan lớn nhất là - 0,173 và - 0,273).

+ Lựa chọn mô hình ARIMA

Từ các giá trị đã tìm được $d = 1$; $p = 1$; $q = 1$ và 5 . Chúng ta có thể rút ra được tổng $p * q = 2$ mô hình thuộc dữ liệu đã khử mùa:

(1,1,1);

(1,1,5)

4.4. Bước 4: Lựa chọn mô hình ARIMA

- **Mô hình của auto.arima**

Trước khi chúng ta kết hợp các mô hình theo hướng thủ công thì ta có thể khởi chạy hàm “**auto.arima()**” trong gói thư viện “**forecast**” vì nó có thể được sử dụng để tìm mô hình ARIMA thích hợp một cách máy móc.

```
forecast::auto.arima(diff1)
```

***Lưu ý*:** ta sẽ sử dụng bộ dữ liệu đã khử mùa.

```

Series: diff1
ARIMA(1,0,1)(1,0,0)[12] with zero mean

Coefficients:
          ar1          ma1          sar1
      0.2785   -0.8818   -0.2009
s.e.  0.1164    0.0656    0.0981

sigma^2 = 3.588:  log likelihood = -237.98
AIC=483.97   AICc=484.33   BIC=494.98

```

Hình 4.4.1: Kết quả đề cử của auto.arima (chỉ số vàng)

Kết quả cho thấy hàm auto.arima() đã đề cử mô hình “ARIMA(1,0,1)(1,0,0)[12]” có sai phân “d = 1” là mô hình thích hợp nhất nhưng ta vẫn cần xác nhận lại thông qua những mô hình khác.

- **Mô hình chọn thủ công**

Kết hợp từ các mô hình ARIMA đã tìm được từ mục 4.2, chúng ta sẽ có được các mô hình ARIMA tổng quát như sau:

(I) 2 mô hình thuộc dữ liệu chưa khử mùa:

(1,1,2);

(2,1,2).

(II) 2 mô hình thuộc dữ liệu đã khử mùa:

(1,1,1);

(1,1,5).

Từ (I) và (II) ta có tổng (I) * (II) = 4 mô hình tổng quát với chu kỳ [S] = 12 tháng:

ARIMA(1,1,1)(1,1,2)[12];

ARIMA(1,1,5)(1,1,2)[12];

ARIMA(1,1,1)(2,1,2)[12];

ARIMA(1,1,5)(2,1,1)[12].

Gán lần lượt 4 mô hình vào các biến đánh số thứ tự từ m1 đến m5 và chạy:

```

m1 <- arima(diff1, order = c(1,1,1), seas = list(order = c(1,1,2), 12))
m2 <- arima(diff1, order = c(1,1,1), seas = list(order = c(2,1,2), 12))
m3 <- arima(diff1, order = c(1,1,5), seas = list(order = c(1,1,2), 12))
m4 <- arima(diff1, order = c(1,1,5), seas = list(order = c(2,1,2), 12))

```

Vì cả 4 mô hình cơ bản đều không có lỗi nên chúng ta sẽ sử dụng AIC và BIC làm công cụ để tìm ra mô hình tối ưu nhất.

- **Mô hình AIC, BIC**

Trong phân tích dữ liệu chuỗi thời gian, AIC (Akaike's Information Criterion) và BIC (Bayesian Information Criterion) là hai phương pháp chính dùng để đánh giá mô hình dự báo. Cả hai đều được sử dụng để đo lường sự tối ưu của mô hình dự báo dựa trên dữ liệu quan sát được.

Cả hai mô hình AIC và BIC đều sử dụng nguyên lý đơn giản là cố gắng cân bằng giữa khả năng phù hợp của mô hình và độ phức tạp của mô hình. Tuy nhiên, cách tiếp cận và công thức tính toán của chúng có một số khác biệt.

AIC có mục tiêu là tìm mô hình có AIC score thấp nhất, tức là mô hình có sự phù hợp tốt nhất với dữ liệu với độ phức tạp thấp nhất.

BIC có mục tiêu là tìm mô hình có BIC score thấp nhất, tức là mô hình có sự phù hợp tốt nhất với dữ liệu với độ phức tạp thấp nhất, đồng thời cân nhắc số lượng quan sát.

Khi áp dụng ARIMA cho một chuỗi dữ liệu, chúng ta thường cần xác định các siêu tham số như p , d và q . Để đảm bảo độ tin cậy và chất lượng của mô hình ARIMA, chúng ta có thể sử dụng AIC và BIC để kiểm tra sự tối ưu của các mô hình khác nhau, sau đó chọn mô hình có AIC hoặc BIC score thấp nhất. Mô hình có AIC hoặc BIC score thấp nhất được coi là mô hình tốt nhất hoặc tối ưu nhất cho dữ liệu đang xem xét.

AIC (m1, m2, m3, m4)

BIC (m1, m2, m3, m4)

| Mô hình | BIC |
|-------------------------|-----------|
| arima(1,1,1)(1,1,2)[12] | 4,889,717 |
| arima(1,1,1)(2,1,2)[12] | 4,930,668 |
| arima(1,1,5)(1,1,2)[12] | 5,043,047 |
| arima(1,1,5)(2,1,2)[12] | 5,079,938 |

Hình 4.4.2: Giá trị AIC của 4 mô hình thuộc chỉ số vàng

| Mô hình | AIC |
|-------------------------|-----------|
| arima(1,1,1)(1,1,2)[12] | 4,731,053 |
| arima(1,1,1)(2,1,2)[12] | 4,745,561 |
| arima(1,1,5)(1,1,2)[12] | 4,778,607 |
| arima(1,1,5)(2,1,2)[12] | 4,789,055 |

Hình 4.4.3: Giá trị BIC của 4 mô hình thuộc chỉ số vàng

Dựa vào hai mô hình AIC và BIC trên ta có thể thấy được rằng mô hình m1 có giá trị AIC, BIC thấp nhất.

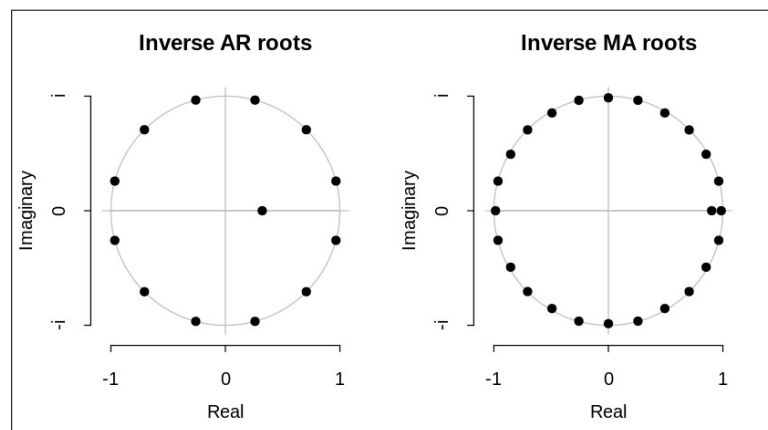
=> Chọn mô hình “arima(1,1,1)(1,1,2)[12]” để dự báo chỉ số giá vàng.

4.5. Bước 5: Kiểm định mô hình đã chọn

- **Kiểm định nghiệm nghịch đảo AR và MA**

Một mô hình ARIMA ổn định khi giá trị tuyệt đối của các chỉ số AR và MA đều phải nhỏ hơn 1 và các nghiệm nghịch đảo này buộc phải nằm bên trong hoặc trên đường tròn đơn vị. Do đó chúng ta sẽ sử dụng hàm “plot()” để vẽ ra mô hình các nghiệm nghịch đảo của thành phần AR, MA và sau đó quan sát.

```
plot(m2)
```



Hình 4.5.1: Mô hình các nghiệm nghịch đảo của chỉ số vàng

Như chúng ta có thể thấy trong mô hình rằng tất cả các nghiệm nghịch đảo AR của mô hình đều nằm ở bên trong đường tròn đơn vị có bán kính bằng 1 nhưng với các nghiệm của MA thì lại nằm ngay trên đường tròn tuy nhiên đây không phải là điều đáng quan ngại.

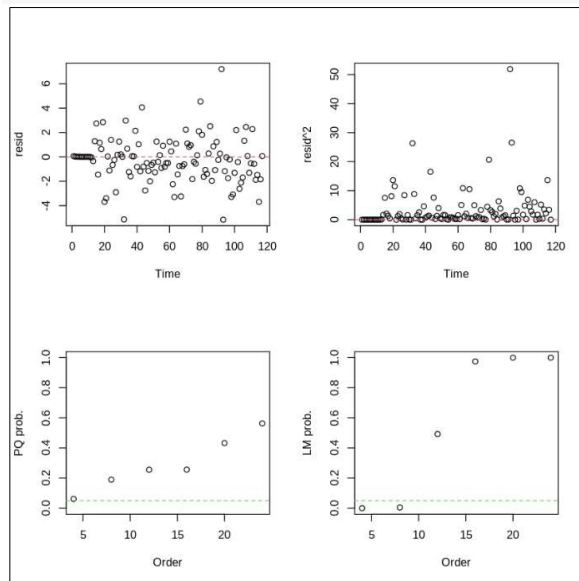
=> Mô hình đã thỏa mãn điều kiện về tính ổn định.

Và tiếp đến sẽ kiểm tra những kiểm định liên quan đến sai số để có thể quan sát được các sai số có phụ thuộc vào quá khứ hay không.

- **Kiểm định phương sai thay đổi**

Mô hình ARIMA được xây dựng dựa trên giả định rằng phương sai của sai số (hay nhiễu) là không đổi theo thời gian. Tuy nhiên, trong thực tế, có thể xảy ra hiện tượng biến động phương sai, tức phương sai của sai số thay đổi theo thời gian. Điều này vi phạm giả định cơ bản của mô hình ARIMA, dẫn đến việc ước lượng tham số của mô hình có thể không chính xác. Do đó chúng tôi đã thực hiện kiểm định hiện tượng phương sai thay đổi của sai số theo kiểm định ARCH để kiểm tra phương sai thay đổi có phụ thuộc vào quá khứ hay không bằng hàm “arch.test()”.

```
arch.test(m2)
```



Hình 4.5.2: Mô hình kiểm định ARCH của chỉ số giá vàng

| Portmanteau-Q test: | | | |
|---------------------------|-------|-------|----------|
| | order | PQ | p.value |
| [1,] | 4 | 9.0 | 0.0611 |
| [2,] | 8 | 11.2 | 0.1893 |
| [3,] | 12 | 14.7 | 0.2558 |
| [4,] | 16 | 19.2 | 0.2563 |
| [5,] | 20 | 20.4 | 0.4326 |
| [6,] | 24 | 22.3 | 0.5633 |
| Lagrange-Multiplier test: | | | |
| | order | LM | p.value |
| [1,] | 4 | 48.96 | 1.33e-10 |
| [2,] | 8 | 20.42 | 4.72e-03 |
| [3,] | 12 | 10.43 | 4.92e-01 |
| [4,] | 16 | 6.28 | 9.75e-01 |
| [5,] | 20 | 4.45 | 1.00e+00 |
| [6,] | 24 | 3.41 | 1.00e+00 |

Hình 4.5.3: Giá trị chi tiết từ mô hình kiểm định ARCH chỉ số giá vàng

Theo kiểm định PQ (kiểm định P – value) cho thấy các độ trễ (bình phương sai số) đều cho giá trị p-value lớn hơn 0.05.

=> Chấp nhận H_0 => các độ trễ đều là nhiễu trắng (resid^2).

Đối với kiểm định LM, cho thấy hai độ trễ 4 và 8 nhỏ hơn 0.05.

=> Bác bỏ H_0 => hai độ trễ này xuất hiện hiện tượng phương sai thay đổi.

Việc xuất hiện hiện tượng phương sai thay đổi có thể làm ảnh hưởng đến kết quả dự báo, khiến giảm độ tin cậy của dự báo. Tuy nhiên, chỉ 1 trong 2 kiểm định cho kết quả có

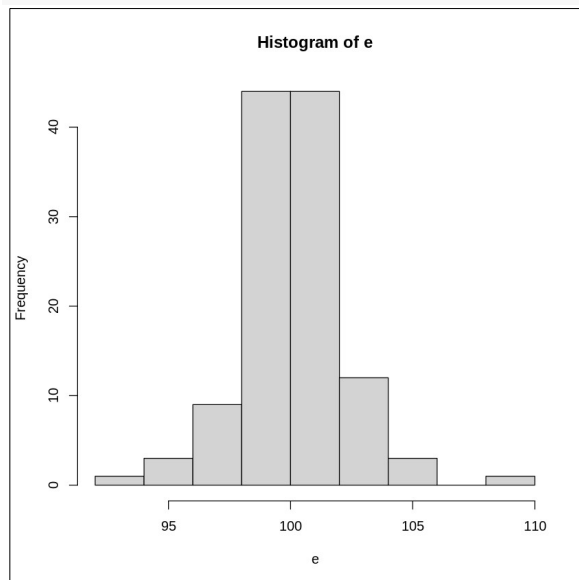
tồn tại hiệu ứng ARCH, vì vậy chúng tôi sẽ tiếp tục cho mô hình thực hiện dự báo chỉ số đô la Mỹ trong ba tháng cuối năm 2022 để kiểm định kết quả dự báo sau đó mới có thể kết luận về độ tin cậy của dự báo.

- **Kiểm định tính phân phối chuẩn của chuỗi phần dư**

Mô hình ARIMA được xây dựng dựa trên giả định rằng sai số (hay nhiễu) có phân phối chuẩn. Chuỗi phần dư, là sự khác biệt giữa giá trị thực tế và giá trị dự đoán bởi mô hình, được coi như là ước lượng của sai số. Nếu chuỗi phần dư không tuân theo phân phối chuẩn, điều này có thể chỉ ra rằng mô hình không phù hợp hoặc có vấn đề. Do đó, chúng tôi đã thực hiện kiểm định phân phối của chuỗi phần dư để đánh giá về độ phù hợp của mô hình đối với bộ dữ liệu.

Trước hết ta sẽ lưu các giá trị phần dư (sai số) của mô hình vào biến “e” và sau đó sẽ kiểm định tính chất phân phối chuẩn trên “e”.

```
e <- arima.errors(m2)
shapiro.test(e)
hist(e)
```



Hình 4.5.4: Đồ thị histogram của phần dư dữ liệu chỉ số vàng

Kiểm định Shapiro và vẽ mô hình ra cho thấy được đồ thị mang hình dáng chuông, các giá trị tập trung vào ở giữa và phân tán ít ra ngoài.

=> Phần dư (sai số) có phân phối chuẩn.

Vì các phần dư đều phân phối chuẩn do đó các kết luận và giả thiết ở trên đều đáng tin cậy.

=> Mô hình phù hợp với bộ dữ liệu.

4.6. Bước 6: Dự báo sử dụng mô hình ARIMA

Chúng ta có thể sử dụng mô hình ARIMA để thực hiện dự báo các giá trị tương lai của các chuỗi bằng cách sử dụng hàm “**forecast.Arima()**” với thiết lập số kỳ dự báo ở tùy chọn “h” (ở hai bộ dữ liệu này mang ý nghĩa tháng) như sau.

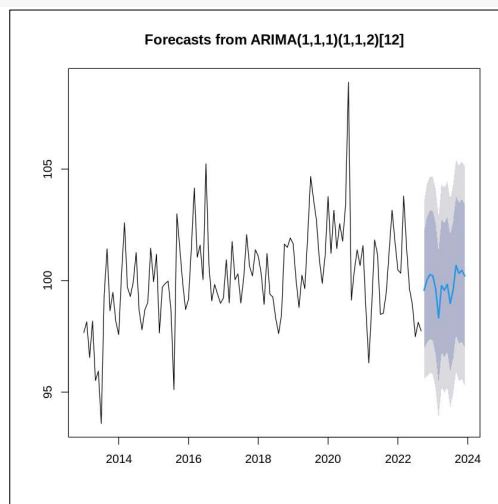
Dự đoán chỉ số vàng sẽ thay đổi trong vòng 15 tháng “h = 15” bằng 2 mô hình m1 và mauto đã được kiểm định thuộc dữ liệu chỉ số vàng:

Mô hình m1

```
DOANm1 <- forecast::forecast.Arima(m1, h = 15)
DOANm1
```

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|--------|-------------------|----------|----------|----------|----------|
| Oct-22 | 99.56329 | 96.97036 | 102.1562 | 95.59775 | 103.5288 |
| Nov-22 | 100.0281 | 97.21611 | 102.8401 | 95.72752 | 104.3287 |
| Dec-22 | 100.2667 | 97.39135 | 103.142 | 95.86925 | 104.6641 |
| Jan-23 | 100.2219 | 97.32386 | 103.1199 | 95.78973 | 104.6541 |
| Feb-23 | 99.62782 | 96.70302 | 102.5526 | 95.15472 | 104.1009 |
| Mar-23 | 98.33087 | 95.38167 | 101.2801 | 93.82045 | 102.8413 |
| Apr-23 | 99.77513 | 96.80237 | 102.7479 | 95.22868 | 104.3216 |
| May-23 | 99.57198 | 96.57606 | 102.5679 | 94.99012 | 104.1539 |
| Jun-23 | 99.83666 | 96.81782 | 102.8555 | 95.21975 | 104.4536 |
| Jul-23 | 98.9789 | 95.93734 | 102.0205 | 94.32723 | 103.6306 |
| Aug-23 | 99.58752 | 96.52341 | 102.6516 | 94.90137 | 104.2737 |
| Sep-23 | 100.6811 | 97.59459 | 103.7676 | 95.9607 | 105.4015 |
| Oct-23 | 100.3271 | 97.17536 | 103.4788 | 95.50693 | 105.1473 |
| Nov-23 | 100.4614 | 97.27277 | 103.65 | 95.58483 | 105.3379 |
| Dec-23 | 100.212 | 96.99398 | 103.43 | 95.29048 | 105.1335 |

Hình 4.6.1: Kết quả dự đoán của mô hình m1 thuộc dữ liệu chỉ số vàng



Hình 4.6.2: Đồ thị dự đoán của mô hình m1 thuộc dữ liệu chỉ số vàng

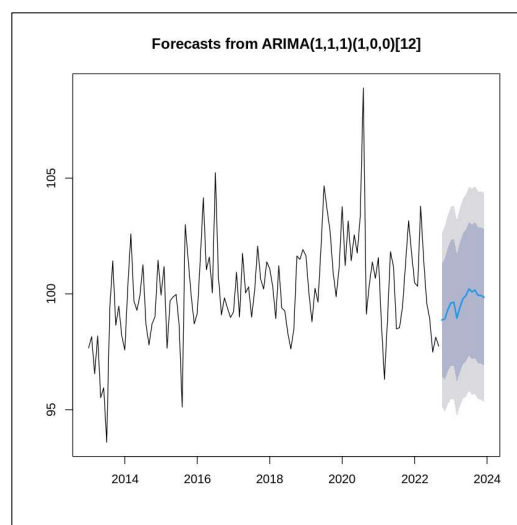
Ngoài việc lựa chọn thủ công các thành phần AR, MA và sai phân để xây dựng mô hình ARIMA, ngôn ngữ R còn cung cấp một hàm có chức năng tự lựa chọn mô hình ARIMA tối ưu nhất dựa trên bộ dữ liệu được cung cấp được gọi là hàm “**auto.arima()**” trong gói thư viện “**forecast**”. Chúng tôi sẽ sử dụng hàm này để xây dựng dự báo và đưa ra so sánh so với dự báo của mô hình được lựa chọn thủ công.

Mô hình m4 (auto arima)

```
DOANmauto <- forecast::forecast.Arima(m4, h = 15)
DOANmauto
```

| | Point Forecast | Lo 80 | Hi 80 | Lo 95 | Hi 95 |
|--------|---------------------------|--------------|--------------|--------------|--------------|
| Oct 22 | 98.87235 | 96.4447 | 101.3 | 95.1596 | 102.585 |
| Nov 22 | 98.91793 | 96.3062 | 101.53 | 94.9236 | 102.912 |
| Dec 22 | 99.31648 | 96.6464 | 101.987 | 95.2329 | 103.4 |
| Jan 23 | 99.60498 | 96.8986 | 102.311 | 95.4659 | 103.744 |
| Feb 23 | 99.64344 | 96.9062 | 102.381 | 95.4572 | 103.83 |
| Mar 23 | 98.95102 | 96.1846 | 101.718 | 94.7201 | 103.182 |
| Apr 23 | 99.41753 | 96.6225 | 102.213 | 95.1429 | 103.692 |
| May 23 | 99.78973 | 96.9665 | 102.613 | 95.472 | 104.108 |
| Jun 23 | 99.93158 | 97.0805 | 102.783 | 95.5712 | 104.292 |
| Jul 23 | 100.2187 | 97.34 | 103.097 | 95.8161 | 104.621 |
| Aug 23 | 100.0889 | 97.1828 | 102.995 | 95.6444 | 104.533 |
| Sep 23 | 100.1667 | 97.2335 | 103.1 | 95.6808 | 104.653 |
| Oct 23 | 99.94029 | 97.0058 | 102.875 | 95.4523 | 104.428 |
| Nov 23 | 99.93113 | 96.9895 | 102.873 | 95.4323 | 104.43 |
| Dec 23 | 99.85107 | 96.8956 | 102.807 | 95.331 | 104.371 |

Hình 4.6.3: Kết quả dự đoán của mô hình auto.arima thuộc dữ liệu chỉ số vàng



Hình 4.6.4: Đồ thị dự đoán của mô hình auto.arima thuộc dữ liệu chỉ số vàng

Hàm “**forecast.Arima()**” cho ta dự báo chỉ số vàng trong 10 tháng tiếp theo, cũng như tính toán khoảng dự đoán 80% và 95% cho những dự đoán đó.

So sánh kết quả dự đoán với thực tế 3 tháng cuối năm 2022

| Thời gian | Thực tế | arima(1,1,1)(1,1,2)[12] | auto arima |
|-----------|---------|-------------------------|------------|
| 10/2022 | 100,5 | 99.56329 | 98,87 |
| 11/2022 | 101,82 | 100.02813 | 98,92 |
| 12/2022 | 100,45 | 100.26666 | 99,32 |

Hình 4.6.5: Giá trị chỉ số vàng thực tế ba tháng cuối năm 2022

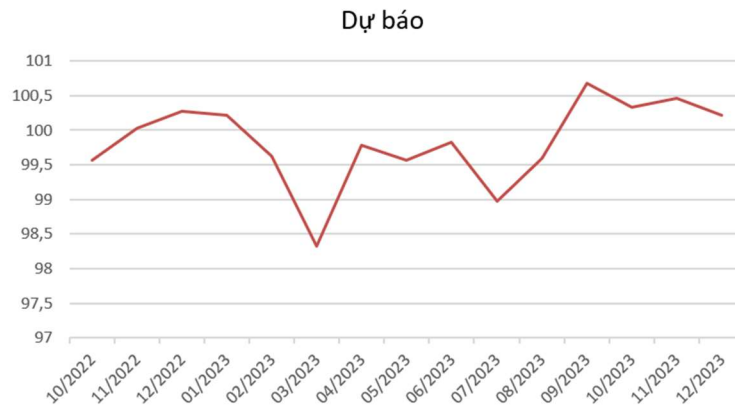
Chúng ta có thể thấy rằng sự dự báo từ đến từ cả 2 mô hình đều đúng ở khoảng tin cậy 95%. Vì vậy, để có thể chọn được mô hình mang khả năng dự báo tin cậy nhất thì chúng ta cần tính sai số bình phương trung bình (MSE - mean squared error) để đo lường độ chính xác của kết quả dự báo.

| arima(1,1,1)(1,1,2)[12] | auto arima |
|-------------------------|------------|
| 1,373367 | 4,1146 |

Hình 4.6.6: Giá trị MSE của hai mô hình ARIMA thuộc chỉ số vàng

Mô hình arima được chọn thủ công có MSE thấp hơn rất nhiều so với mô hình được hàm auto arima lựa chọn.

=> Chọn mô hình ARIMA(1,1,1)(1,1,2)[12] là mô hình ổn định và phù hợp nhất đối với bộ dữ liệu chỉ số giá vàng.



Hình 4.6.7: Chi tiết dự báo 15 tháng tiếp theo bằng mô hình arima(1,1,1)(1,1,2)[12]

CHƯƠNG 5: KẾT LUẬN VÀ KIẾN NGHỊ

Dựa vào dữ liệu của chỉ số giá vàng trong khoảng 10 năm từ tháng 1/2013 đến tháng 9/2023, chúng tôi đã sử dụng mô hình ARIMA để dự báo 15 tháng tiếp theo từ tháng 10/2022 đến tháng 12/2023, trong đó 3 tháng cuối năm 2022 được sử dụng để kiểm định độ tin cậy dự báo của các mô hình, 12 tháng tiếp theo của năm 2023 là phần dự báo cho tương lai từ những mô hình tối ưu nhất. Dưới đây là những kết luận và kiến nghị có thể đưa ra cho các nhà đầu tư trong năm 2023:

- Vào năm 2023 sẽ có 2 đợt giảm mạnh vào tháng 3 (giảm 1,3 điểm) và tháng 7 (giảm 0,85 điểm).
- Giá vàng vào 3 tháng đầu năm 2023 có xu hướng giảm liên tục và chạm đáy vào tháng 3 (98,95).

=> Đây là thời điểm nên đầu tư mua vàng vì đây sẽ là thời điểm giá vàng có giá trị thấp nhất trong năm.

- Sau khi đã chạm đáy ở tháng 3, chỉ số giá vàng sẽ có xu hướng tăng mạnh trở lại vào tháng 4 (tăng 1,45 điểm) và duy trì tương đối ổn định đến đầu tháng 7 (quanh mức 99,7 điểm).
- Dự báo vào tháng 7 cho thấy chỉ số giá vàng sẽ giảm gần 1 điểm so với tháng 6, tuy nhiên tình trạng giảm này chỉ xảy ra trong tháng này, sau đó sẽ phục hồi và trên đà tăng trưởng mạnh cho 5 tháng cuối trong năm 2023.
- Trong 5 tháng còn lại, chỉ số tăng cao hơn so với thời điểm đầu năm và đạt được đỉnh vào tháng 9 (100,16).

=> Đây là thời điểm các nhà đầu tư có thể bán ra.

- Trong 3 tháng cuối năm, chỉ số giá vàng không có quá nhiều biến động và có xu hướng giảm nhẹ ở tháng 12.

Nhìn chung ta có thể thấy chỉ số giá vàng trong năm 2023 có xu hướng tăng giảm với biên độ dao động tương đối lớn. Điều này cho thấy vàng có tiềm năng mang lại biên độ lợi nhuận tương đối cao cho các nhà đầu tư ngắn hạn trong năm 2023.

Kết luận và kiến nghị: Dựa trên dự báo này, chúng tôi dự đoán chỉ số giá vàng sẽ biến động tương đối lớn trong năm 2023. Tuy nhiên, các nhà đầu tư vẫn nên hết sức thận trọng khi đưa ra các quyết định đầu tư, đặc biệt là khi đánh giá rủi ro và xác định chiến lược ứng phó phù hợp bao gồm những rủi ro tiềm ẩn. Ngoài ra, việc theo dõi và phân tích các yếu tố kinh tế và chính trị toàn cầu có thể cung cấp thông tin hữu ích để đưa ra quyết định đầu tư thông minh và linh hoạt trong tương lai gần.

TÀI LIỆU THAM KHẢO

(1) Hương, T. s. . *TỔNG CỤC THỐNG KÊ*.

Đã truy lục 12/21, 2023, từ TỔNG CỤC THỐNG KÊ: <https://www.gso.gov.vn/px-web-2/?pxid=V0835&theme=Th%C6%B0%C6%A1ng%20m%E1%BA%A1i%2C%20gi%C3%A1%20c%E1%BA%A3>

(2) *Investing*. (2023, 12/21). Retrieved from Investing.com: <https://vn.investing.com/currencies/xau-usd>

(3) Shumway, R. H., & Stoffer, D. S. (2011). Time Series Analysis and Its Applications With R Examples. In R. H. Shumway, & D. S. Stoffer, *Time Series Analysis and Its Applications With R Examples* (p. 596). New York: Springer.

(4) *xauusd*. (2023, 12/21). Retrieved from xauusd.vn: <https://xauusd.vn/>